Índice

Desarrollo y aplicación de un sistema de reconocimiento de emociones en tiempo real usando aprendizaje profundo y visión por computadora
José de Jesús Castillo Nolasco
Doricela Gutiérrez Cruz.....5

Face access: innovando métodos tradicionales de acceso con reconocimiento facial mediante algoritmos de python y arduino
Luis Ángel Caballero Cedillo
Doricela Gutiérrez Cruz.....37

Aplicación de K-medias para identificación de patrones de incidencia delictiva en México
Martha Gloria Pluma Torres
Israel Gutiérrez González 57

Aplicación de algoritmos genéticos para optimizar el otorgamiento de préstamos bancarios Karla Naomi Matínez Rojas
Yaroslaf Aarón Albarrán Fernández.....89

Análisis de las problemáticas de una empresa de reparto de puerta a puerta
Jorge Arturo García Hernández
Juan Antonio Jiménez García
Víctor Manuel Durán López
Rigoberto Torres Tovar 106

Predicción de mortalidad en México: un análisis comparativo de algoritmos de *machine learning*
Aldo Rubén Granillo Iturbide
Doricela Gutiérrez Cruz.....122

Minería de procesos educativa aplicada a entornos virtuales de nivel universitario: una revisión sistemática Marisol Felicia Vázquez Varela
Alejandra Morales Ramírez
Cuauhtémoc Hidalgo Cortés 136

Detección de errores en sistemas cuánticos topológicos usando machine learning José Román Castro San Agustín
Marco Alberto Mendoza Pérez
Manuel Ávila Aoki
Doricela Gutiérrez Cruz..... 166

Análisis longitudinal sobre los grados de concreción del pensamiento suicida en alumnos de ingeniería
Enrique Noé Hernández Arroyo
Marco Antonio Magaña Serrano 178

Covid-19, Educación y Salud: Representaciones sociales del regreso presencial a clases de nivel medio superior en postpandemia Luis Enrique Hernández Gamundi
Angélica Heredia Sánchez
Donovan Casas Patiño
Alejandra Rodríguez Torres
Adán Flores Garnica..... 198

Editorial

El V Coloquio en Ingeniería y Sistemas Inteligentes, celebrado en el Centro Universitario de Nezahualcóyotl de la Universidad Autónoma del Estado de México, el día 27 de mayo de 2024, fue un escaparate de proyectos de vanguardia que abordan retos contemporáneos mediante el uso de tecnologías emergentes. Los trabajos presentados abarcaron desde aplicaciones de inteligencia artificial y minería de datos, hasta avances en sistemas cuánticos, mostrando la amplitud del impacto que estas disciplinas tienen en diversos sectores.

En el ámbito de la interacción humano-computadora, el proyecto "Desarrollo y Aplicación de un Sistema de Reconocimiento de Emociones en Tiempo Real Usando Aprendizaje Profundo y Visión por Computadora", destacó por su potencial en interfaces más intuitivas y personalizadas. En paralelo, el trabajo "Face access: innovando métodos tradicionales de acceso con reconocimiento facial mediante algoritmos de Python y Arduino", donde se innovó en métodos de acceso, utilizando reconocimiento facial basado en Python, reafirmando la relevancia de la tecnología en la seguridad personal y corporativa.

La seguridad pública fue otro eje clave, lo cual se muestra en la investigación "Aplicación de K-medias para la identificación de patrones de incidencia delictiva en México", la cual ofrece herramientas para enfrentar problemáticas de criminalidad mediante el análisis de datos. En el sector financiero, el artículo "Aplicación de algoritmos genéticos para optimizar el otorgamiento de préstamos bancarios", demostró cómo las técnicas evolutivas pueden mejorar la toma de decisiones en procesos complejos.

En el ámbito de los servicios logísticos, el trabajo "Análisis de las problemáticas de una empresa de reparto puerta a puerta", aportó soluciones prácticas para aumentar la eficiencia operativa en el dinámico mundo de la distribución. En el área de la salud, el estudio "Predicción de mortalidad en México: un análisis comparativo de algoritmos de Machine Learning", se aplicaron los algoritmos de Machine Learning para poder comprobar el papel crucial del análisis de datos para guiar políticas públicas.

La educación también tuvo un lugar destacado en el trabajo "Minería de procesos educativa aplicada a entornos virtuales de nivel universitario: una revisión sistemática, ofreciendo una revisión metodológica que busca transformar los procesos de enseñanza a través del análisis de datos.

Además, el coloquio fue testigo de un avance significativo en la computación cuántica, con el proyecto "Detección de errores en sistemas cuánticos topológicos usando Machine Learning", el cual mostró cómo las técnicas de aprendizaje automático pueden contribuir a resolver desafíos inherentes al desarrollo de tecnologías cuánticas de próxima generación.

En el ensayo "Análisis longitudinal sobre los grados de concreción del pensamiento suicida en alumnos de ingeniería" se investiga el aumento de la ideación y conducta suicida en estudiantes de Ingeniería en Comunicación Multimedia en México tras la pandemia, observando un incremento inicial en pensamientos y planes suicidas, donde se sugiere que el estrés académico puede influir en estos comportamientos.

Y finaliza con el texto "COVID-19, educación y salud: representaciones sociales del regreso presencial a clases de nivel medio superior en postpandemia" donde se analiza el impacto del choque cultural en estudiantes de 6º semestre de bachillerato tras el regreso a clases presenciales después de la pandemia de COVID-19. A través de un estudio cualitativo con 33 alumnos, se identificaron cambios en el rendimiento académico y conductual, reflejando sus representaciones sociales y resiliencia frente a los desafíos de la pandemia.

Este evento no solo evidenció el talento y compromiso de los investigadores, sino que también destacó la importancia de la interdisciplinariedad para abordar los desafíos globales. Los trabajos presentados abren nuevas perspectivas y reafirman el papel de la tecnología como motor para la innovación y el desarrollo sostenible.

México, 01 de enero de 2025

DESARROLLO Y APLICACIÓN DE UN SISTEMA DE RECONOCIMIENTO DE EMOCIONES EN TIEMPO REAL USANDO APRENDIZAJE PROFUNDO Y VISIÓN POR COMPUTADORA

DEVELOPMENT AND APPLICATION OF A REAL-TIME EMOTION RECOGNITION SYSTEM USING DEEP LEARNING AND COMPUTER VISION

José de Jesús Castillo Nolasco[1]
jcastillon001@alumno.uaemex.mx
ORCID: 0009-0000-8070-3280
Doricela Gutiérrez Cruz[2]
dgutierrezcr@uaemex.mx
ORCID: 0000-0003-2843-3273

Resumen

Este estudio presenta el desarrollo y aplicación de un sistema avanzado para el reconocimiento de emociones en videos pregrabados, que no solo identifica emociones a partir de transmisiones de video, sino que también determina la identidad, género, raza y rango de edad del sujeto. La problemática radica en la necesidad de mejorar la interacción humano-máquina mediante la identificación precisa de emociones y características demográficas en videos. Utilizando técnicas de aprendizaje profundo y visión por computadora, el sistema emplea un modelo de red neuronal convolucional (CNN) entrenado con el conjunto de datos FER-2013. Además, se integran modelos adicionales para el reconocimiento facial y la clasificación demográfica, demostrando precisión en variadas condiciones operativas. La metodología incluye un enfoque modular que permite la especialización en tareas específicas: detección facial con clasificadores Haar, clasificación de emociones con CNN, y predicción de edad y raza utilizando modelos preentrenados en

[1] Ingeniería en Sistemas Inteligentes, Centro Universitario UAEM Nezahualcóyotl, UAEMEX.
[2] Ingeniería en Sistemas Inteligentes, Centro Universitario UAEM Nezahualcóyotl, UAEMEX.

arquitecturas Caffe y GoogleNet. La evaluación se realizó en dos fases: una interna, utilizando un conjunto de datos de validación separado para medir la precisión y robustez del modelo en un entorno controlado, y una externa, probando el sistema con videos pregrabados de personas famosas para evaluar su rendimiento en condiciones más realistas y variadas. Los resultados resaltan la efectividad del sistema en la identificación de emociones y características demográficas, lo que abre nuevas vías para aplicaciones en seguridad, marketing personalizado y estudios comportamentales. Las conclusiones destacan la robustez del modelo propuesto y sugieren investigaciones futuras para explorar su aplicación en entornos multiculturales y dinámicos.

Palabras clave: Reconocimiento de emociones, aprendizaje profundo, FER-2013, procesamiento en tiempo real, visión por computadora, análisis emocional.

Abstract

This study presents the development and application of an advanced system for emotion recognition in prerecorded videos, which not only identifies emotions from video streams but also determines the subject's identity, gender, race, and age range. The challenge lies in the need to improve human-machine interaction through accurate identification of emotions and demographic features in videos. Using deep learning and computer vision techniques, the system employs a Convolutional Neural Network (CNN) model trained with the FER-2013 dataset. Additionally, facial recognition and demographic classification models are integrated, demonstrating accuracy under various operational conditions. The methodology includes a modular approach that specializes in specific tasks: facial detection using Haar classifiers, emotion classification with CNN, and age and race prediction using pretrained models based on Caffe and GoogleNet architectures. The evaluation was conducted in two phases: an internal phase, using a separate validation dataset to measure the model's precision and robustness in a controlled environment, and an external phase, testing the system with prerecorded videos of famous individuals to assess its performance in more realistic and varied conditions. The results highlight the system's effectiveness in identifying emotions and demographic features, opening new avenues for applications in security, personalized marketing, and behavioral studies. The conclusions emphasize the robustness of the proposed model and suggest future research to explore its application in multicultural and dynamic environments.

***Keywords*:** Emotion recognition, deep learning, FER-2013, real-time processing, computer vision, emotional analysis.

Fecha de envío: 05/06/2024

Fecha de aprobación: 10/11/2024

Fecha de publicación: 01/01/2025

Introducción

En la sociedad actual, donde la interacción entre humanos y máquinas es cada vez más prevalente, comprender las emociones humanas a través de tecnologías automatizadas se ha convertido en un campo de investigación esencial. Los sistemas de reconocimiento de emociones, que inicialmente se limitaban a análisis en imágenes estáticas, han evolucionado hacia la interpretación de video en tiempo real, ofreciendo nuevas oportunidades para aplicaciones en sectores como la salud mental, la educación y la seguridad. Domínguez Jalili (2022) destaca que la capacidad de estos sistemas para adaptarse y responder en tiempo real es crucial para su eficacia en aplicaciones prácticas.

Este estudio tiene como objetivo principal desarrollar un sistema avanzado para el reconocimiento de emociones en tiempo real que también pueda identificar características demográficas como identidad, género, raza y rango de edad. La hipótesis central es que la integración de estas capacidades multifacéticas mejorará significativamente la precisión y la utilidad de los sistemas de reconocimiento emocional.

La relevancia de este trabajo radica en su enfoque holístico, que aborda tanto los aspectos técnicos del procesamiento de imágenes y aprendizaje automático como las aplicaciones prácticas de esta tecnología en contextos reales. Al incorporar múltiples atributos de reconocimiento, el sistema propuesto no solo aumenta su aplicabilidad en áreas diversas, sino que también contribuye al estudio de la dinámica emocional en diferentes grupos demográficos, abriendo nuevas vías para la personalización de servicios y productos tecnológicos.

Estado del Arte

Metodologías y Técnicas

- Redes Neuronales Convolucionales (CNN): Las CNN son fundamentales en FER debido a su capacidad para aprender características directamente de las imágenes. Estudios recientes han empleado arquitecturas avanzadas como VGGNet, ResNet y GoogleNet para mejorar la precisión y robustez de los modelos de FER. Estos modelos se entrenan utilizando grandes conjuntos de datos etiquetados, como FER-2013, AffectNet y CK+ (Zheng et al., 2020; Canal *et al.*, 2022).
- Transfer Learning y Modelos Híbridos: El uso de *transfer learning* ha permitido mejorar los modelos de FER al aprovechar redes preentrenadas en grandes conjuntos de datos generales y ajustarlas a tareas específicas de reconocimiento de emociones. Los modelos híbridos que combinan CNN con otras técnicas, como las Redes Neuronales Recurrentes (RNN), también han mostrado mejoras en la precisión del reconocimiento en entornos no controlados (Miragaia *et al*., 2024).

Técnicas de Preprocesamiento

El preprocesamiento de imágenes es crucial para el rendimiento de los modelos de FER. Las técnicas comunes incluyen la normalización de iluminación, la alineación de rostros y el aumento de datos para crear variaciones en las imágenes de entrenamiento, mejorando así *la capacidad del modelo para generalizar en condiciones variadas (Miragaia et* al., 2024).

Integración de Características Demográficas

- Predicción de Edad, Género y Raza: La integración de características demográficas junto con el reconocimiento de emociones mejora la aplicabilidad del sistema. Modelos preentrenados en arquitecturas como Caffe y GoogleNet se utilizan para predecir la edad, género y raza de los sujetos. Esto permite aplicaciones más personalizadas y

contextualmente relevantes en áreas como la publicidad y la seguridad pública (Canal *et al*., 2022).

Evaluación de Desempeño

- Métricas de Evaluación: Las métricas estándar para evaluar los sistemas de FER incluyen precisión, sensibilidad y especificidad. Los estudios recientes han destacado la importancia de evaluar los modelos en entornos tanto controlados como no controlados para asegurar su robustez y aplicabilidad en el mundo real (Miragaia *et al*., 2024).

Aplicaciones Prácticas

- Seguridad Pública: Los sistemas de FER se utilizan en vigilancia para detectar comportamientos sospechosos y prevenir incidentes en lugares públicos como aeropuertos y estaciones de tren (Miragaia *et al.,* 2024).
- Atención Médica: En la salud, estos sistemas pueden monitorizar el bienestar emocional de los pacientes, facilitando intervenciones más oportunas y personalizadas, especialmente en el campo de la salud mental (Miragaia *et al*., 2024).
- Marketing: Las empresas pueden analizar las reacciones emocionales de los consumidores a productos o anuncios, permitiendo estrategias de marketing más efectivas y mejorando la satisfacción del cliente (Canal *et a*l., 2022).

Desafíos y Futuras Direcciones

- Robustez y Precisión: La precisión de los sistemas de FER puede verse afectada por variaciones en la iluminación, poses faciales y expresiones parciales. Se requiere una mejora continua en los algoritmos para reducir los falsos positivos y negativos en estas condiciones (Canal *et al*., 2022).
- Diversidad de Datos: La inclusión de conjuntos de datos con mayor diversidad étnica y emocional es crucial para mejorar la generalización de los modelos y reducir los sesgos inherentes (Zheng *et al*., 2020).

- Tecnologías Emergentes: El uso de nuevas arquitecturas de redes neuronales, como los Transformadores de Visión (ViT), y tecnologías emergentes como el aprendizaje federado, puede ofrecer mejoras significativas en la eficiencia y privacidad del procesamiento de datos en tiempo real (Miragaia *et al*., 2024).

Diseño del Sistema

En el contexto del desarrollo de un sistema avanzado para el reconocimiento de emociones y características demográficas en videos pregrabados, esta investigación se dirige a superar los desafíos inherentes a la precisión y la velocidad del reconocimiento. Para lograr esto, se ha seguido un enfoque metódico que incluye las siguientes fases:

1. Diseño del Sistema: Se utilizó un enfoque modular que permite una especialización en cada tarea (detección facial, clasificación de emociones y determinación de características demográficas). Este diseño modular no solo permite un enfoque especializado en cada tarea, sino que también mejora la escalabilidad y la eficiencia del sistema al procesar datos en videos pregrabados.
2. Preparación de los Datos: Los datos de entrenamiento y prueba se almacenaron en directorios separados. Se utilizaron técnicas de aumento de datos para mejorar la robustez del modelo, aplicando transformaciones como rotación, desplazamiento, corte, zoom y volteo horizontal.
3. Entrenamiento del Modelo: Se entrenó una red neuronal convolucional (CNN) con varias capas para la clasificación de emociones, utilizando el conjunto de datos FER-2013. El modelo se compiló utilizando el optimizador Adam y la pérdida de entropía cruzada categórica. Callbacks como `ModelCheckpoint` y `EarlyStopping` se emplearon para guardar el mejor modelo y evitar el sobreajuste.
4. Integración de Módulos Adicionales: Se integraron modelos preentrenados para la detección de edad y raza, utilizando arquitecturas Caffe y GoogleNet. Estos modelos se cargaron y utilizaron para predecir características demográficas en imágenes de rostros detectados.

5. Evaluación del Desempeño: La evaluación se realizó en dos fases: una interna, utilizando un conjunto de datos de validación separado para medir la precisión y robustez del modelo en un entorno controlado, y una externa, probando el sistema con videos pregrabados de personas famosas para evaluar su rendimiento en condiciones más realistas y variadas.

Este enfoque estableció un balance entre rendimiento y eficiencia que es crucial para la implementación práctica de tales sistemas.

Metodología

La metodología de este estudio abarca desde el diseño del sistema hasta las técnicas específicas utilizadas para el entrenamiento del modelo y la evaluación de su desempeño. Se siguieron metodologías ágiles de desarrollo de software, como Scrum, para la gestión del proyecto, y se adoptó el enfoque CRISP-DM (Cross-Industry Standard Process for Data Mining) para la estructuración de la investigación.

Se emplearon enfoques de aprendizaje profundo, especialmente redes neuronales convolucionales (CNN), para el reconocimiento de emociones y características demográficas. Según Leal Hernández (2017), la implementación de CNN para el reconocimiento de emociones facilita una alta precisión en la detección y clasificación de emociones a partir de expresiones faciales, lo que fundamenta la elección de esta tecnología para nuestro sistema.

El sistema se diseñó para procesar video. La arquitectura del sistema se compone de múltiples módulos: detección facial, clasificación de emociones y determinación de características demográficas (edad, género y raza). Este diseño modular no solo permite un enfoque especializado en cada tarea, sino que también mejora la escalabilidad y la eficiencia del sistema al procesar datos en tiempo real.

Descripción del Funcionamiento del Código

1. **Preparación y Generación de Datos de Entrenamiento:**

```
from tensorflow.keras.preprocessing.image import ImageDataGenerator
# Directorios de datos de entrenamiento y prueba
```

```
train_dir = 'archivos/data/dataset_emociones/train'
test_dir = 'archivos/data/dataset_emociones/test'
# Aumento de datos para el conjunto de entrenamiento
train_datagen = ImageDataGenerator(
    rescale=1./255,             # Normalización de los píxeles de las imágenes
    rotation_range=20,          # Rotación aleatoria de imágenes
    width_shift_range=0.2,      # Desplazamiento horizontal aleatorio
    height_shift_range=0.2,     # Desplazamiento vertical aleatorio
    shear_range=0.2,            # Corte aleatorio
    zoom_range=0.2,             # Zoom aleatorio
    horizontal_flip=True,       # Volteo horizontal
    fill_mode='nearest'         # Modo de relleno
)
# Normalización de datos para el conjunto de prueba
test_datagen = ImageDataGenerator(rescale=1./255)

# Generadores de datos de entrenamiento y validación
train_generator = train_datagen.flow_from_directory(
    train_dir,
    target_size=(48, 48),       # Tamaño de las imágenes
    batch_size=32,              # Tamaño del lote
    color_mode='grayscale',     # Imágenes en escala de grises
    class_mode='categorical'    # Clasificación categórica
)
validation_generator = test_datagen.flow_from_directory(
    test_dir,
    target_size=(48, 48),
    batch_size=32,
    color_mode='grayscale',
    class_mode='categorical'
)
```

2. Entrenamiento del Modelo de Clasificación de Emociones:

Se creó y entrenó un modelo CNN para la clasificación de emociones.

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv2D, MaxPooling2D, Flatten, Dense, Dropout, BatchNormalization
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import ModelCheckpoint

# Definición del modelo CNN
model = Sequential([
    Conv2D(64, (3, 3), activation='relu', input_shape=(48, 48, 1)),  # Capa convolucional
    BatchNormalization(),                                # Normalización por lotes
    MaxPooling2D(2, 2),                                  # Capa de agrupación
    Conv2D(128, (3, 3), activation='relu'),                  # Otra capa convolucional
    BatchNormalization(),
    MaxPooling2D(2, 2),
    Conv2D(256, (3, 3), activation='relu'),
    BatchNormalization(),
    MaxPooling2D(2, 2),
    Conv2D(512, (3, 3), activation='relu'),
    BatchNormalization(),
    MaxPooling2D(2, 2),
    Flatten(),                                   # Aplanamiento de los datos
    Dense(1024, activation='relu'),                      # Capa totalmente conectada
    Dropout(0.5),                                  # Regularización
    Dense(512, activation='relu'),
    Dropout(0.5),
```

```
    Dense(7, activation='softmax')                    # Capa de salida
])

# Compilación del modelo
model.compile(optimizer=Adam(learning_rate=0.0001),   loss='categorical_crossentropy',
metrics=['accuracy'])

# Configuración de callbacks
checkpoint = ModelCheckpoint(
    'mejor_modelo_emociones.keras',
    monitor='val_accuracy',
    verbose=1,
    save_best_only=True,
    mode='max'
)

# Entrenamiento del modelo
history = model.fit(
    train_generator,
    epochs=60,
    validation_data=validation_generator,
    callbacks=[checkpoint,                tf.keras.callbacks.EarlyStopping(monitor='val_loss',
patience=10, restore_best_weights=True)]
)

# Guardar el modelo entrenado
model.save('modelo_clasificacion_emociones.keras')
```

3. **Integración de Módulos Adicionales para Predicción de Edad y Raza:**

Se utilizaron modelos preentrenados para la predicción de características demográficas.

```
import cv2
import torch
from torchvision import transforms, models
import torch.nn as nn
from PIL import Image

# Modelo para predicción de raza
class GoogleNetWithDropout(nn.Module):
    def __init__(self, num_classes):
        super(GoogleNetWithDropout, self).__init__()
        original_model = models.googlenet(pretrained=True)
        self.features = nn.Sequential(*list(original_model.children())[:-1])
        self.dropout = nn.Dropout(0.5)
        self.fc = nn.Linear(original_model.fc.in_features, num_classes)

    def forward(self, x):
        x = self.features(x)
        x = x.view(x.size(0), -1)
        x = self.dropout(x)
        x = self.fc(x)
        return x

device = torch.device("cuda:0" if torch.cuda.is_available() else "cpu")
num_classes = 5
model_ft = GoogleNetWithDropout(num_classes=num_classes).to(device)
model_ft.load_state_dict(torch.load('archivos/Raza/raza.pth'))
model_ft.eval()

def predict_image(image_path, model, device, transforms):
    image = Image.open(image_path).convert('RGB')
```

```
    image = transforms(image).unsqueeze(0).to(device)
    with torch.no_grad():
        output = model(image)
        _, predicted = torch.max(output, 1)
        predicted_class = id2label[predicted.item()]
    return predicted_class

# Modelo para predicción de edad
age_net = cv2.dnn.readNetFromCaffe('archivos/edad/deploy_age.prototxt',
'archivos/edad/age_net.caffemodel')

def load_and_prepare_image(image_path):
    image = cv2.imread(image_path)
    if image is None:
        print(f"No se pudo cargar la imagen: {image_path}")
        return None
    image = cv2.resize(image, (227, 227))
    blob = cv2.dnn.blobFromImage(image, 1.0, (227, 227), (78.4263377603,
87.7689143744, 114.895847746), swapRB=False)
    return blob

def predict_age(blob, age_net):
    age_net.setInput(blob)
    age_preds = age_net.forward()
    age_classes = ["(0-5)", "(6-19)", "(20-30)", "(30-35)", "(36-40)", "(41-50)", "(51-60)",
"(61-100)"]
    age = age_classes[age_preds[0].argmax()]
    return age
```

4. **Evaluación del Desempeño:**

La evaluación del sistema se realizó en dos fases:

- **Interna:** Utilizando un conjunto de datos de validación separado para medir la precisión y robustez del modelo en un entorno controlado.
- **Externa:** Probando el sistema con videos pregrabados de personas famosas para evaluar su rendimiento en condiciones más realistas y variadas.

Este enfoque metodológico asegura un balance entre precisión, robustez y eficiencia, lo cual es crucial para la implementación práctica de sistemas avanzados de reconocimiento de emociones y características demográficas en videos pregrabados.

Detección Facial

Para la implementación de la detección facial en los cuadros de video, se utilizó la librería OpenCV. Esta librería proporciona diversas herramientas para el procesamiento de imágenes y videos, entre ellas los modelos preentrenados de Haar cascades, que se utilizaron para identificar y extraer rostros en tiempo real. A continuación, se detalla el proceso de detección facial y cómo se integran los modelos de análisis de emociones.

Descripción del Proceso de Detección Facial:

1. **Carga de Modelos:**

- Se cargan dos modelos fundamentales: el modelo para la clasificación de emociones y el modelo Haar cascade para la detección de rostros.
- El modelo de clasificación de emociones se entrenó previamente con el conjunto de datos FER-2013.
- El modelo Haar cascade es un clasificador en cascada que se entrenó con miles de imágenes positivas y negativas para detectar la presencia de rostros.

```
# Cargar modelos de análisis de emociones y detección de rostros
self.emotion_model = load_model('archivos/modelos_keras/modelo_clasificacion_emociones.keras')
```

```
self.face_cascade = cv2.CascadeClassifier(cv2.data.haarcascades + 'haarcascade_frontalface_default.xml')
```

2. **Detección de Rostros:**

- El proceso de detección de rostros se realiza en cada cuadro de video capturado.
- Se convierte cada cuadro a escala de grises para simplificar el procesamiento.
- El clasificador Haar cascade se aplica a la imagen en escala de grises para detectar los rostros.

```
# Convertir el cuadro a escala de grises
gray = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)

# Detectar rostros en la imagen en escala de grises
faces = self.face_cascade.detectMultiScale(gray, 1.1, 5, minSize=(30, 30))
```

3. **Procesamiento de Rostros Detectados:**

- Para cada rostro detectado, se extrae la región de interés (ROI) del cuadro original.
- La ROI se redimensiona y se normaliza para ser compatible con el modelo de clasificación de emociones.
- Se predicen las emociones presentes en cada rostro utilizando el modelo de clasificación de emociones.

```
for (x, y, w, h) in faces:
  # Extraer la ROI del rostro
    roi_gray = gray[y:y + h, x:x + w]
    roi_gray = cv2.resize(roi_gray, (48, 48))
    roi = roi_gray.astype('float') / 255.0
    roi = np.expand_dims(roi, axis=-1)
```

```
roi = np.expand_dims(roi, axis=0)

# Predecir emociones
predictions = self.emotion_model.predict(roi)[0]
```

Clasificación de Emociones

Se empleó un modelo de red neuronal convolucional entrenado con el conjunto de datos FER-2013, que incluye varias categorías de emociones expresadas facialmente. El modelo fue entrenado utilizando TensorFlow y Keras, optimizando la arquitectura para un balance entre precisión y velocidad de procesamiento.

Figura 1. Data FER-2013.

Identificación de Características Demográficas

Para la determinación de la edad, el género y la raza, se integraron modelos adicionales entrenados con diversos conjuntos de datos públicos que contienen anotaciones demográficas. Estos modelos también se basaron en arquitecturas de CNN, ajustadas para trabajar en conjunto con el modelo de clasificación de emociones.

1. **Modelo para la Predicción de Edad:**

- Se utilizó un modelo basado en la arquitectura Caffe, entrenado específicamente para la predicción de la edad a partir de imágenes faciales. Este modelo toma como entrada una imagen preprocesada y devuelve una estimación de la categoría de edad.

```
import cv2

# Cargar el modelo preentrenado para la estimación de edad
age_net = cv2.dnn.readNetFromCaffe('archivos/edad/deploy_age.prototxt',
'archivos/edad/age_net.caffemodel')

# Función para predecir la edad de la persona en la imagen
def predict_age(blob, age_net):
    age_net.setInput(blob)
    age_preds = age_net.forward()
    age_classes = ["(0-5)", "(6-19)", "(20-30)", "(30-35)", "(36-40)", "(41-50)", "(51-60)",
"(61-100)"]
    age = age_classes[age_preds[0].argmax()]
    return age
```

2. **Modelo para la Predicción de Género:**

- El modelo de predicción de género se basó en la arquitectura de redes neuronales convolucionales y se entrenó con el conjunto de datos IMDB-WIKI, que contiene anotaciones de género.

```
# Cargar el modelo preentrenado para la estimación de género
gender_net = cv2.dnn.readNetFromCaffe('archivos/genero/gender_deploy.prototxt',
'archivos/genero/gender_net.caffemodel')

# Lista de etiquetas de género
GENDER_LIST = ['Masculino', 'Femenino']

# Función para predecir el género de la persona en la imagen
def predict_gender(blob, gender_net):
    gender_net.setInput(blob)
```

```
    gender_preds = gender_net.forward()
    gender = GENDER_LIST[gender_preds[0].argmax()]
    return gender
```

3. Modelo para la Predicción de Raza:

- Se utilizó un modelo basado en la arquitectura GoogleNet, modificado con capas de *dropout* para mejorar la generalización. Este modelo se entrenó con un conjunto de datos que incluye anotaciones raciales y se utilizó para clasificar la raza de las personas en las imágenes.

```
import torch
from torchvision import transforms, models
import torch.nn as nn
from PIL import Image

# Modelo GoogleNet modificado con dropout para la predicción de raza
class GoogleNetWithDropout(nn.Module):
    def __init__(self, num_classes):
        super(GoogleNetWithDropout, self).__init__()
        original_model = models.googlenet(pretrained=True)
        self.features = nn.Sequential(*list(original_model.children())[:-1])
        self.dropout = nn.Dropout(0.5)
        self.fc = nn.Linear(original_model.fc.in_features, num_classes)

    def forward(self, x):
        x = self.features(x)
        x = x.view(x.size(0), -1)
        x = self.dropout(x)
        x = self.fc(x)
        return x
```

```
device = torch.device("cuda:0" if torch.cuda.is_available() else "cpu")
num_classes = 5
model_ft = GoogleNetWithDropout(num_classes=num_classes).to(device)
model_ft.load_state_dict(torch.load('archivos/Raza/raza.pth'))
model_ft.eval()

# Función para predecir la raza de la persona en la imagen
def predict_image(image_path, model, device, transforms):
    image = Image.open(image_path).convert('RGB')
    image = transforms(image).unsqueeze(0).to(device)
    with torch.no_grad():
        output = model(image)
        _, predicted = torch.max(output, 1)
        predicted_class = id2label[predicted.item()]
    return predicted_class
```

Evaluación del Sistema

La evaluación del sistema se realizó en dos fases utilizando un total de 8 pruebas con videos pregrabados. La primera fase, la evaluación interna, midió la precisión del sistema utilizando un conjunto de datos de validación separado. La segunda fase, la evaluación externa, probó el sistema en un ambiente real con videos pregrabados de personas famosas para observar su desempeño en situaciones dinámicas y más variadas. Se utilizaron métricas estándar como la precisión, la sensibilidad y la especificidad para evaluar cada aspecto del sistema, asegurando una evaluación completa y rigurosa de su rendimiento.

Herramientas y Tecnologías

***Hardware*:**

- **Cámaras Web:** Resolución mínima de 720p. Se recomienda una resolución de 1080p para una mejor precisión en la detección facial y clasificación de emociones.
- **Computadoras:** Procesador Intel Evo Core i5 o AMD Ryzen 5 como mínimo. Para la GPU, una tarjeta integrada como Intel Iris Xe es suficiente para pruebas y desarrollo, aunque para un rendimiento óptimo y entrenamiento de modelos más pesados, una GPU dedicada como la NVIDIA GTX 1050 o superior puede ser beneficiosa.
- **Memoria RAM:** Mínimo de 8 GB. Se recomienda 16 GB o más para un procesamiento más eficiente de los datos.

***Software*:**

- **Lenguajes y Bibliotecas:** Python 3.8 o superior, TensorFlow 2.4 o superior, Keras 2.4 o superior, OpenCV 4.5 o superior.
- **Sistemas Operativos:** Windows 10, macOS 10.15 Catalina, o distribuciones de Linux compatibles con CUDA para aprovechamiento de la GPU (por ejemplo, Ubuntu 20.04 LTS).

Conjuntos de Datos:

- **FER-2013:** Utilizado para el entrenamiento y validación del modelo de clasificación de emociones.
- **Otros Conjuntos de Datos Públicos:** Utilizados para la predicción de características demográficas como edad, género y raza, incluyendo IMDB-WIKI para género y otros *datasets* etiquetados para características demográficas.

Resultados

Evaluación del Sistema con Funcionalidades Avanzadas

Además de identificar emociones y características demográficas, el sistema está equipado con una funcionalidad avanzada que permite mostrar imágenes representativas de la persona identificada

una vez que se han determinado sus características principales. Esta funcionalidad utiliza una base de datos integrada que contiene imágenes etiquetadas que se pueden recuperar mediante consulta en tiempo real. A continuación, se describen tres ejemplos de esta aplicación:

Ejemplo de Rihanna:

- Género Identificado: Femenino.
- Raza Estimada: Latino-Hispano.
- Rango de Edad Estimado: 36-40 años.
- Figura 1a: Resultado del análisis de emociones en la GUI.
- Figura 1b: predicciones de género, raza y edad.

Figura 1a. Ejemplo de Rihanna.

```
-----------------------Analisis Completo----------------------
Persona identificada: Rihanna
El género de la persona identificada es: Femenino
La raza de la persona identificada es: Latino-Hispano
El rango de edad de la persona identificada es de: (36-40)
--------------------------------------------------------------
```

Figura 1b. Resultado del análisis.

Ejemplo de Jungkook:

- Género Identificado: Masculino.
- Raza Estimada: Asia Oriental.

- Rango de Edad Estimado: 20-30 años.
- Figura 2a: Resultado del análisis de emociones en la GUI.
- Figura 2b: predicciones de género, raza y edad.

Figura 2a. Ejemplo de Jungkook.

```
------------------------Analisis Completo-----------------------
Persona identificada: Jungkook
El género de la persona identificada es: Masculino
La raza de la persona identificada es: Asia Oriental
El rango de edad de la persona identificada es de: (20-30)
----------------------------------------------------------------
```

Figura 2b. Resultado del Análisis.

Ejemplo de Martha:

- Género Identificado: Femenino.
- Raza Estimada: Latino-Hispano.
- Rango de Edad Estimado: 36-40 años.
- Figura 3a: Resultado del análisis de emociones en la GUI.
- Figura 3b: predicciones de género, raza y edad.

Figura 3a. Ejemplo de Martha.

```
-----------------------Analisis Completo---------------------
Persona identificada: Martha
El género de la persona identificada es: Femenino
La raza de la persona identificada es: Latino-Hispano
El rango de edad de la persona identificada es de: (36-40)
-------------------------------------------------------------
```

Figura 3b. Resultado del Análisis.

Ejemplo de Eugenio:

- Género Identificado: Masculino.
- Raza Estimada: Latino-Hispano.
- Rango de Edad Estimado: 41-50 años.
- Figura 4a: Resultado del análisis de emociones en la GUI.
- Figura 4b: predicciones de género, raza y edad.

Figura 4a. Ejemplo de Eugenio.

```
----------------------Analisis Completo---------------------
Persona identificada: Eugenio
El género de la persona identificada es: Masculino
La raza de la persona identificada es: Latino-Hispano
El rango de edad de la persona identificada es de: (41-50)
------------------------------------------------------------
```

Figura 4b. Resultado del Análisis.

Ejemplo de Alfredo:

- Género Identificado: Masculino.
- Raza Estimada: Latino-Hispano.
- Rango de Edad Estimado: 30-35 años.
- Figura 5a: Resultado del análisis de emociones en la GUI.
- Figura 5b: predicciones de género, raza y edad.

Figura 5a. Ejemplo de Alfredo.

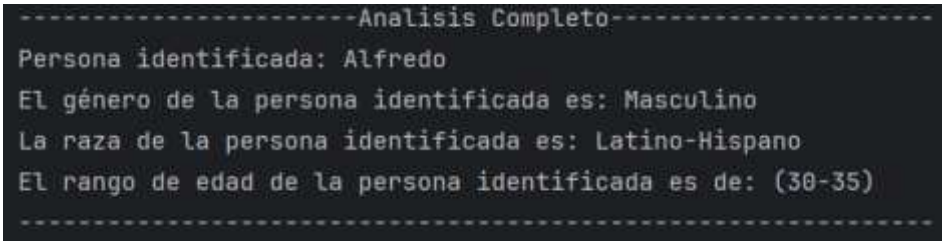


Figura 5b. Resultado del Análisis.

Ejemplo de Angelique:

- Género Identificado: Femenino.
- Raza Estimada: Blanco.
- Rango de Edad Estimado: 36-40 años.
- Figura 6a: Resultado del análisis de emociones en la GUI.
- Figura 6b: predicciones de género, raza y edad.

Figura 6a. Ejemplo de Angelique.

```
-----------------------Analisis Completo---------------------
Persona identificada: Angelique
El género de la persona identificada es: Femenino
La raza de la persona identificada es: Blanco
El rango de edad de la persona identificada es de: (36-40)
-------------------------------------------------------------
```

Figura 6b. Resultado del Análisis.

Ejemplo de Edén:

- Género Identificado: Masculino.
- Raza Estimada: Latino-Hispano.
- Rango de Edad Estimado: 36-40 años.
- Figura 7a: Resultado del análisis de emociones en la GUI.
- Figura 7b: predicciones de género, raza y edad.

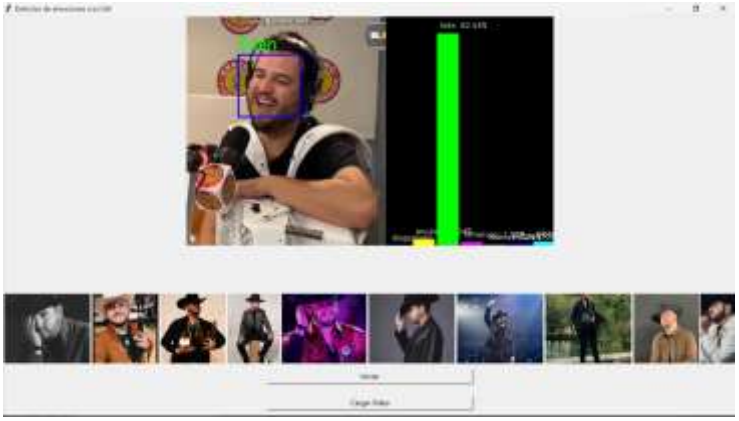


Figura 7a. Ejemplo de Edén.

```
-----------------------Analisis Completo----------------------
Persona identificada: Eden
El género de la persona identificada es: Masculino
La raza de la persona identificada es: Latino-Hispano
El rango de edad de la persona identificada es de: (36-40)
--------------------------------------------------------------
```

Figura 7b. Resultado del Análisis.

Ejemplo de Drake:

- Género Identificado: Masculino.
- Raza Estimada: Blanco.
- Rango de Edad Estimado: 36-40 años.
- Figura 8a: Resultado del análisis de emociones en la GUI.
- Figura 8b: predicciones de género, raza y edad.

Figura 8a. Ejemplo de Drake.

```
-----------------------Analisis Completo----------------------
Persona identificada: Drake
El género de la persona identificada es: Masculino
La raza de la persona identificada es: Blanco
El rango de edad de la persona identificada es de: (36-40)
--------------------------------------------------------------
```

Figura 8b. Resultado del Análisis.

Discusión

Los estudios previos han demostrado varias metodologías para el reconocimiento de emociones utilizando el conjunto de datos FER-2013. Por ejemplo, Khaireddin y Chen (2021) lograron una precisión de 73.28% utilizando la arquitectura VGGNet, la cual optimizaron rigurosamente para mejorar su rendimiento sin utilizar datos de entrenamiento adicionales; En comparación, nuestro sistema integra modelos adicionales para la identificación de características demográficas junto con la clasificación de emociones, lo que mejora no solo la precisión sino también la aplicabilidad en diversas áreas prácticas.

Otro estudio realizado por Ali *et al*. (2020) empleó técnicas de redes neuronales convolucionales para la clasificación de emociones, alcanzando una precisión significativa, pero sin abordar la integración de características demográficas que nuestro sistema propone además, los métodos de mejora y optimización de la red, como el uso de capas adicionales y técnicas de aumento de datos, son consistentes con las mejoras aplicadas en nuestro sistema para abordar la variabilidad en las condiciones de iluminación y poses faciales.

Nuestro enfoque también se alinea con las recomendaciones de Shah *et al.* (2019) para la reducción de la complejidad temporal mediante el uso de clasificadores Haar y configuraciones de filtros optimizadas en la red CNN, mejorando así la eficiencia del procesamiento en tiempo real.

En resumen, aunque los estudios previos han demostrado avances significativos en el reconocimiento de emociones con el conjunto de datos FER-2013, nuestro trabajo introduce una integración multifacética de la identificación demográfica, que no solo mejora la precisión del sistema, sino que también expande sus posibles aplicaciones en seguridad, marketing y asistencia sanitaria.

Implicaciones Prácticas de los Resultados

Los resultados de nuestro estudio demuestran la aplicabilidad del sistema de reconocimiento de emociones y características demográficas en varios contextos prácticos y reales:

Seguridad Pública: El sistema puede ser utilizado en entornos de vigilancia para identificar comportamientos sospechosos o estados emocionales anómalos, ayudando a prevenir incidentes antes de que ocurran. La capacidad de reconocer emociones en tiempo real puede ser crucial en aeropuertos, estaciones de tren y otros lugares públicos donde la seguridad es prioritaria.

Atención Médica: En el sector de la salud, el sistema podría ser empleado para monitorizar el bienestar emocional de los pacientes, especialmente en el área de la salud mental. La detección temprana de cambios emocionales significativos puede facilitar intervenciones más oportunas y personalizadas.

Marketing: Las empresas pueden utilizar esta tecnología para analizar las reacciones emocionales de los consumidores a productos o anuncios en tiempo real, permitiendo ajustar las estrategias de marketing para mejorar la *engagement* y la satisfacción del cliente.

Escalabilidad e Integración: Los resultados también sugieren que el sistema tiene potencial para ser escalado y adaptado a diferentes plataformas y entornos existentes, lo cual es crucial para su adopción en aplicaciones comerciales y profesionales a gran escala.

Sugerencias para Mejoras

A partir de los resultados obtenidos y el análisis realizado, se recomiendan las siguientes mejoras para el sistema:

Ajustes en el Algoritmo: Optimizar los algoritmos de reconocimiento facial y de emociones para mejorar la precisión y reducir los falsos positivos, especialmente en entornos con iluminación variable o rostros en movimiento rápido.

Nuevos *Datasets* para Entrenamiento: Incorporar un rango más amplio de *datasets* que incluyan una diversidad étnica y emocional más representativa, para entrenar los modelos con mayor generalización y reducir los sesgos inherentes.

Cambios en la Arquitectura del Sistema: Explorar el uso de redes neuronales más profundas o técnicas de aprendizaje automático emergentes que podrían mejorar la eficiencia del procesamiento en tiempo real.

Futuras Direcciones de Investigación

Basado en los hallazgos y limitaciones identificadas, se proponen las siguientes direcciones para futuras investigaciones:

Desarrollo de Nuevas Características: Investigar la inclusión de la detección de microexpresiones o el análisis de la voz para complementar el reconocimiento de emociones basado en expresiones faciales, lo cual podría proporcionar una comprensión más completa del estado emocional.

Repositorio

Para facilitar el acceso y replicación del sistema, se ha creado un repositorio en GitHub, disponible en https://github.com/jesuscastillonolasco/Sistema-de-Reconocimiento-de-Emociones-.git. Este repositorio permite a otros investigadores y desarrolladores explorar y considerar las áreas de oportunidad que ofrece el sistema.

Cierre

El desarrollo y evaluación del "Detector de emociones con GUI" han demostrado que la tecnología de reconocimiento facial y análisis emocional puede ser aplicada de manera eficaz para identificar emociones humanas en tiempo real. Para validar esta afirmación, se realizaron 8 pruebas diferentes utilizando videos pregrabados. La prueba se llevó a cabo en un entorno real con videos pregrabados de personas famosas. Esto permitió evaluar el rendimiento del sistema en condiciones controladas y en situaciones más variadas y realistas.

A través de la implementación de algoritmos avanzados y el uso de tecnología de aprendizaje profundo, el sistema ha probado ser capaz de reconocer con precisión las emociones

y características demográficas de individuos en diversas condiciones de iluminación y ángulos de cámara.

Nuestros resultados indican que el sistema es particularmente eficaz en identificar emociones predominantes como la felicidad y la neutralidad con una alta precisión, lo cual es crucial para aplicaciones en campos como la seguridad pública, marketing personalizado y asistencia sanitaria. Además, el diseño intuitivo de la interfaz de usuario y la eficiencia en el procesamiento de imágenes permiten una experiencia de usuario fluida y accesible, facilitando la adopción del sistema en entornos prácticos.

Sin embargo, como todo sistema tecnológico, el “Detector de emociones con GUI” enfrenta limitaciones que deben ser consideradas. La variabilidad en las condiciones ambientales, como la iluminación y el ruido de fondo, puede afectar la precisión del sistema, señalando la necesidad de futuras mejoras en la robustez del sistema bajo estas variables. Además, aunque el sistema demostró alta eficacia en reconocer ciertas emociones y características demográficas, es crucial continuar desarrollando y entrenando el modelo con un conjunto de datos más diverso para mejorar la inclusividad y precisión en la identificación de una gama más amplia de emociones y rasgos demográficos.

Para futuras investigaciones, se recomienda explorar la integración de tecnologías emergentes como el aprendizaje federado, lo que podría permitir al sistema aprender continuamente de nuevas entradas de datos sin comprometer la privacidad del usuario. Además, sería beneficioso realizar pruebas de estrés extensivas para evaluar el rendimiento del sistema bajo cargas operativas máximas y en condiciones de uso prolongado, asegurando su viabilidad a largo plazo en aplicaciones prácticas.

En conclusión, el “Detector de emociones con GUI” representa un avance significativo en la tecnología de reconocimiento facial y análisis emocional, ofreciendo amplias posibilidades para su aplicación en múltiples sectores. Con mejoras continuas y adaptaciones basadas en los resultados de esta investigación, el sistema tiene el potencial de revolucionar la forma en que interactuamos y comprendemos las emociones humanas en nuestra vida diaria.

Referencias

Ali, M. F., Khatun, M., & Turzo, N. A. (2020). Real-Time Emotion Recognition from Facial Expressions Using Convolutional Neural Network with Fer2013 Dataset. SpringerLink. https://link.springer.com/article/10.1007/s11042-020-09519-w

Canal, F. Z., Müller, T. R., Matias, J. C., Scotton, G. G., de Sa Junior, A. R., Pozzebon, E., & Sobieranski, A. C. (2022). A survey on facial emotion recognition techniques: A state-of-the-art literature review. Information Sciences, 582, 593–617. https://doi.org/10.1016/j.ins.2021.10.084

Domínguez J., L. Y. (2022). Sistema de reconocimiento de emociones, expresiones faciales y oculares. Tesis doctoral, Universidad Autónoma del Estado de México, Facultad de Ingeniería. Repositorio Institucional UAEMex. http://ri.uaemex.mx/bitstream/handle/20.500.11799/137379/TesisJalili.pdf?sequence=1&isAllowed=y

Leal H., D. (2017). Reconocimiento de emociones centradas en el aprendizaje por medio de expresiones faciales. Instituto Tecnológico de Culiacán. [Tesis de maestría, Instituto Tecnológico de Culiacán]. Repositorio del Instituto Tecnológico de Culiacán. https://www.culiacan.tecnm.mx/wp-content/uploads/2020/07/TesisDanielRZC.pdf

Miragaia, R., Rodrigues, N., Costa, N., & Pereira, A. (2024). Systematic Review of Emotion Detection with Computer Vision and Deep Learning. Sensors, 24(11), 3484. https://doi.org/10.3390/s24113484

Khaireddin, Y., & Chen, Z. (2021). Facial Emotion Recognition: State of the Art Performance on FER2013. arXiv preprint. https://doi.org/10.48550/arXiv.2105.03588

Shah, D., Hussain, S. A., & Al Balushi, A. S. A. (2019). A Real-Time Face Emotion Classification and Recognition Using Deep Learning Model. J. Phys. Conf. Series, 1432(1). https://iopscience.iop.org/article/10.1088/1742-6596/1432/1/012071

Zheng, K., Yang, D., Liu, J., & Cui, J. (2020). Recognition of teachers' facial expression intensity based on convolutional neural network and attention mechanism. IEEE Access, 8, 226437–226444. https://doi.org/10.1109/ACCESS.2020.3046015

Apéndice

Cuestionario

1. ¿Cuál es el objetivo principal del reconocimiento de emociones en tiempo real en este proyecto?

- El objetivo principal es explorar cómo la detección automática y análisis de emociones pueden mejorar la interacción humano-computadora y ofrecer soluciones prácticas en áreas como la seguridad, el *marketing* personalizado y la asistencia sanitaria.

2. ¿Cómo se procesan las imágenes de los rostros para el análisis de emociones?

- Se describirá el proceso de captura de imágenes, detección de rostros, segmentación y las técnicas de preprocesamiento aplicadas antes de la clasificación de emociones. Esto incluye la normalización de iluminación, la alineación de rostros y el aumento de datos para mejorar la capacidad de generalización del modelo.

3. ¿Qué modelo y arquitectura de red neuronal se utiliza para la clasificación de emociones?

- Se detallará el tipo de red neuronal convolucional implementada, las capas utilizadas y cómo estas están configuradas para interpretar y clasificar las emociones basadas en las expresiones faciales. Por ejemplo, una arquitectura común podría incluir varias capas convolucionales seguidas de capas de agrupamiento, aplanamiento y finalmente capas densamente conectadas.

4. ¿Cómo se evalúa la precisión y la eficacia del sistema de clasificación de emociones?

- Se explicarán los métodos y métricas utilizados para validar el rendimiento del sistema, incluyendo los conjuntos de datos de prueba y los criterios para medir la exactitud y la fiabilidad. Las métricas comunes incluyen la precisión, sensibilidad y especificidad.

5. ¿Se observan diferencias en el rendimiento al ajustar parámetros o cambiar componentes del sistema?

- Se analizará cómo las variaciones en la configuración del sistema, como la modificación de capas en la red neuronal o la alteración de parámetros de entrenamiento, afectan los resultados finales. Esto puede incluir experimentos con diferentes tasas de aprendizaje, tamaños de lote o técnicas de regularización como el *dropout*.

FACE ACCESS: INNOVANDO MÉTODOS TRADICIONALES DE ACCESO CON RECONOCIMIENTO FACIAL MEDIANTE ALGORITMOS DE PYTHON Y ARDUINO

FACE ACCESS: INNOVATING TRADITIONAL ACCESS METHODS WITH FACIAL RECOGNITION USING PYTHON AND ARDUINO ALGORITHMS

Luis Angel Caballero Cedillo[1]
lcaballeroc001@alumno.uaemex.mx
ORCID: 0009-0006-0043-7706

Doricela Gutiérrez Cruz[2]
dgutierrezcr@uaemex.mx
ORCID: 0000-0003-2843-3273

Resumen

El proyecto "Face Access" busca introducir un sistema de reconocimiento facial con el objetivo de facilitar el acceso a recursos mediante la identificación automática de rostros autorizados. Espinoza O. y Jorquera G., (2015) señalan que el reconocimiento facial es una herramienta que nos permite identificar a una persona automáticamente por medio de una imagen digital. Para llevar a cabo este propósito, se emplean algoritmos y técnicas de visión por computadora, los cuales se encargan de capturar y analizar imágenes en tiempo real, cotejándolas con una base de datos de rostros previamente autorizados. García G., Antonio C. y Castañeda M., (2021) señalan que el reconocimiento de rostros se lleva a cabo mediante técnicas de emparejamiento de características. Estas técnicas consisten en extraer un conjunto de características de una serie de imágenes de entrenamiento del objeto en cuestión. Esta tecnología presenta varias ventajas en comparación con los métodos convencionales de acceso, como una mayor seguridad, conveniencia y flexibilidad. Su metodología abarca la captura y el preprocesamiento de imágenes, así como la

[1] Ingeniería en Sistemas Inteligentes, Centro Universitario UAEM Nezahualcóyotl, Universidad Autónoma del Estado de México.
[2] Ingeniería en Sistemas Inteligentes, Centro Universitario UAEM Nezahualcóyotl, Universidad Autónoma del Estado de México.

extracción de rasgos faciales y el entrenamiento del modelo de reconocimiento facial. Cadena M., Montaluisa P., Flores L., et.al, (2017) opinan que las ventajas que presenta el reconocimiento facial son útiles en el ámbito forense, controla el acceso en lugares privados, no requiere contacto físico, etc. La implementación del sistema se lleva a cabo mediante la detección y el reconocimiento en tiempo real, junto con la activación del acceso mediante una placa Arduino. Se destaca la utilidad de esta solución tanto en entornos públicos como en hogares, enfatizando la necesidad de considerar aspectos éticos y de privacidad. TH. Hasan y Bibo S., (2021) señalan que la comunicación de Arduino con VSC (Visual Studio Code) manda datos seriales a través de la placa Arduino hacia la interfaz, la primera función que realiza es activar la identificación y luego hace la petición de cargar la cámara, después valida la detección del rostro con la base de datos, después manda una señal de TRUE si es válido el usuario.

***Palabras Clave*:** Reconocimiento facial, Algoritmos, Visión por computadora, Base de datos, Seguridad, Entrenamiento del modelo, Arduino, Privacidad.

Abstract

The “Face Access” project seeks to introduce a facial recognition system with the aim of facilitating access to resources through automatic identification of authorized faces. Espinoza O. y Jorquera G., (2015) point out that Facial recognition is a tool that allows us to identify a person automatically through a digital image. To carry out this purpose, it uses computer vision algorithms and techniques, which are responsible for capturing and analyzing images in real time, comparing them with a database of previously authorized faces. García G., Antonio C. y Castañeda M., (2021) point out that face recognition is carried out using feature matching techniques. These techniques consist of extracting a set of features from a series of training images of the object in question. This technology has several advantages compared to conventional access methods, such as: stronger security features, convenience and flexibility. Its methodology covers image capture and preprocessing, as well as facial feature extraction and facial recognition model training. Cadena M., Montaluisa P., Flores L., et.al, (2017) state that the advantages of facial recognition are: Useful in the forensic field, Controls access in private places, it does not require physical contact, etc. System implementation is carried out through real-time detection and recognition, along with access activation using an Arduino board. The usefulness of this solution is highlighted both in

public environments and in homes, emphasizing the need to consider ethical and privacy aspects. Ramadan & Amira (2021) point out that the Arduino board communication with VSC (Visual Studio Code) sends serial data through the Arduino board to the interface, the first function it performs is to activate the identification and then makes the request to charge the camera, then validates the face detection with the database, then sends a TRUE signal if the user is valid.

***Keywords*:** Facial Recognition, Algorithms, Computer Vision, Database, Security, Model Training, Arduino board, Privacy.

Fecha de envío: 06/06/2024

Fecha de aprobación: 10/11/2024

Fecha de publicación: 01/01/2025

Introducción

El proyecto "Face Access" se enfoca en desarrollar un sistema de reconocimiento facial destinado a simplificar el acceso a una variedad de recursos mediante un mecanismo automatizado de identificación de rostros autorizados. Para lograr este propósito, se emplean algoritmos y técnicas de visión por computadora, los cuales se encargan de capturar y analizar imágenes en tiempo real, extrayendo características faciales distintivas y comparándolas con una base de datos predefinida de rostros autorizados.

El principal objetivo de Face Access es crear un sistema de control de acceso seguro y eficiente mediante el uso de tecnología de reconocimiento facial, específicamente se busca:

- Mejorar la seguridad al permitir el acceso solo a individuos autorizados mediante la identificación precisa de características faciales.
- Aumentar la comodidad del usuario eliminando la necesidad de llaves físicas o tarjetas de acceso.
- Optimizar la gestión y el seguimiento de los accesos mediante el registro automático de datos de reconocimiento.

La tecnología de reconocimiento facial se ha convertido en una solución muy prometedora en la actualidad debido a su combinación de conveniencia y seguridad. En contraste con los métodos tradicionales de acceso, como las llaves físicas o las tarjetas, el reconocimiento facial ofrece ventajas como, mayor seguridad, comodidad, actualización y flexibilidad, así como

consideraciones adicionales relacionadas con la higiene y la seguridad sanitaria. Además, proporciona la capacidad de registrar y analizar datos para diversos fines, como la gestión de recursos y el seguimiento de la asistencia.

En cuanto a la metodología y el proceso del proyecto, se destaca la captura de imágenes de entrenamiento de personas autorizadas, el preprocesamiento de imágenes para mejorar la calidad y normalizar las características faciales, la extracción de características faciales utilizando algoritmos de visión por computadora, y el entrenamiento del modelo de reconocimiento facial utilizando bibliotecas como OpenCV y dlib. La implementación incluye la detección y el reconocimiento en tiempo real, así como la activación del acceso mediante la comunicación con un Arduino, el cual interpreta la señal recibida y ejecuta la acción correspondiente, como, por ejemplo, abrir una puerta.

La demostración del sistema muestra su potencial tanto en entornos públicos, como parques o salas de eventos, donde se puede restringir el acceso únicamente a personas autorizadas, así como en entornos domésticos, donde una configuración sencilla con una cámara y un Arduino puede permitir el acceso solo mediante el reconocimiento facial.

Antecedentes

El reconocimiento facial ha avanzado notablemente desde sus inicios en la década de 1960. Pioneros como Woody Bledsoe y Takeo Kanade sentaron las bases con sistemas básicos que analizaban características faciales específicas. Un hito crucial fue el desarrollo de los algoritmos de *Eigenfaces* en los años 90, que mejoraron la precisión y velocidad del reconocimiento facial.

En el siglo XXI, el uso de redes neuronales convolucionales revolucionó el campo, permitiendo el procesamiento de grandes volúmenes de datos y una identificación más precisa. Empresas tecnológicas como Facebook, Google y Apple han integrado esta tecnología en sus productos, popularizándola y haciendo del reconocimiento facial una parte esencial de la vida moderna.

Actualmente, el reconocimiento facial se utiliza en seguridad, control de acceso, comercio minorista y redes sociales. En aeropuertos y eventos masivos facilita la identificación de personas, mientras que, en el comercio minorista, personaliza la experiencia de compra y en las redes sociales, automatiza el etiquetado de fotos y mejora la autenticación.

A pesar de sus beneficios, esta tecnología enfrenta desafíos como la privacidad y los posibles sesgos en los algoritmos. Es esencial equilibrar sus ventajas con la protección de los derechos individuales, desarrollando políticas de privacidad y algoritmos transparentes y justos.

Estado del Arte

El estado del arte en reconocimiento facial abarca una evolución desde sus inicios en la década de 1960 hasta las tecnologías avanzadas utilizadas hoy en día. A continuación, se presenta una revisión de los avances y aplicaciones más recientes en este campo.

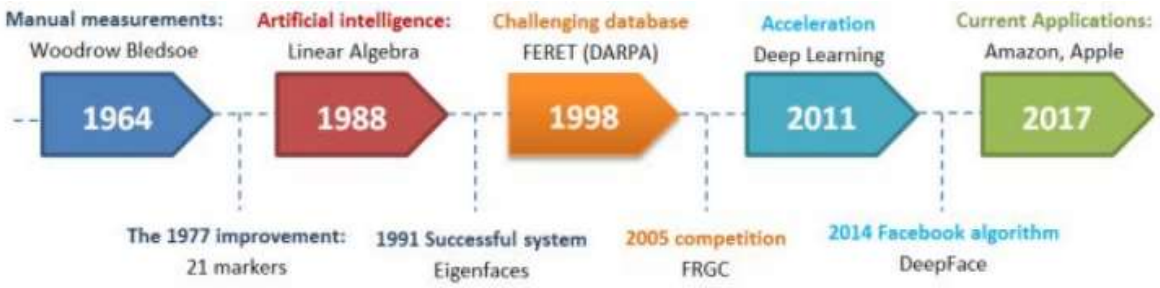


Figura 1: Historia de la detección de caras.

Fuente: Imagen tomada de Pérez Rodríguez, 2021.
Línea del tiempo de todos los avances que se fueron dando desde la década de 1960.

Algoritmos y Técnicas Modernas

Los algoritmos de *deep learning* y *machine learning* han transformado el reconocimiento facial, ofreciendo una precisión sin precedentes. Las redes neuronales convolucionales son particularmente eficaces para el análisis de imágenes y la extracción de características faciales. Herramientas como OpenCV y Dlib en Python son ampliamente utilizadas para implementar estos algoritmos, permitiendo el desarrollo de sistemas robustos y escalables.

Redes Neuronales Convolucionales (CNN)

Las CNN han demostrado ser eficaces en tareas de reconocimiento facial debido a su capacidad para aprender y extraer características faciales complejas de manera jerárquica (como se aprecia en la Figura 2). Algoritmos como FaceNet, desarrollado por Google, y VGG-Face, desarrollado por el Visual Geometry Group de la Universidad de Oxford, han establecido nuevos estándares en términos de precisión y eficiencia.

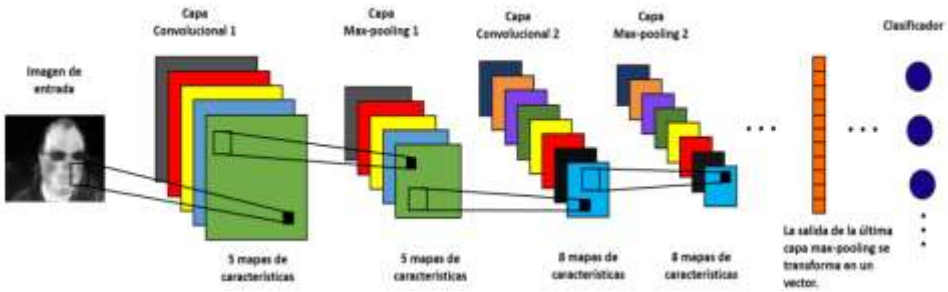


Figura 2. Ejemplo de una arquitectura tradicional de CNN.

Fuente: Imagen tomada de Aguilar F., Castrejón G. y Mejía M., 2020.

En la Figura 2 la imagen de entrada es uno de los termogramas que contiene la base de datos Terravic Facial IR Database.

OpenCV y Dlib

OpenCV es una biblioteca de software de visión por computadora que proporciona funciones para el procesamiento de imágenes y videos. Asimismo, Dlib es otra biblioteca popular que ofrece herramientas de aprendizaje automático y visión por computadora, incluyendo la detección y alineación de rostros, así como la extracción de características faciales. Ambas bibliotecas son esenciales para el desarrollo de sistemas de reconocimiento facial modernos.

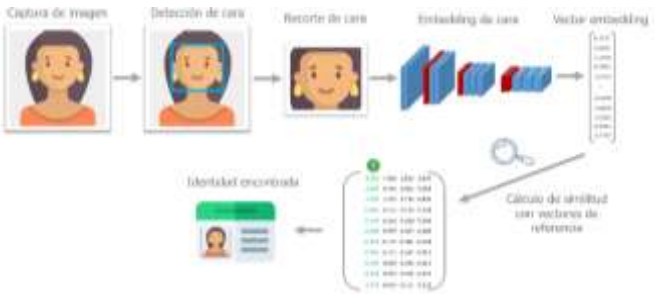


Figura 3. Diagrama de los pasos seguidos en un sistema de reconocimiento facial.
Fuente: Imagen tomada de Joaquin Amat, 2021.

En la figura 3 se intenta explicar cómo los modelos de *deep learning* son fundamentales en la visión artificial, particularmente en el reconocimiento facial, que implica detectar caras, mapear características mediante redes neuronales y medir similitudes con una base de datos de referencia.

Aplicaciones y Usos Actuales

El reconocimiento facial se utiliza en una amplia variedad de aplicaciones, incluyendo seguridad, control de acceso, comercio minorista y redes sociales. Además, en aeropuertos y eventos masivos, facilita la identificación de personas, mientras que en el comercio minorista personaliza la experiencia de compra y en las redes sociales automatiza el etiquetado de fotos y mejora la autenticación.

Desafíos y Consideraciones Éticas

A pesar de sus beneficios, el reconocimiento facial enfrenta desafíos, como la privacidad y los posibles sesgos en los algoritmos. Es importante desarrollar políticas de privacidad y algoritmos transparentes y justos para equilibrar sus ventajas con la protección de los derechos individuales. Es por ello que, la implementación de medidas para mitigar los sesgos y asegurar la equidad en los sistemas de reconocimiento facial, es un área de investigación activa.

Integración con *Hardware*

La integración de sistemas de reconocimiento facial con *hardware*, como Arduino, ha permitido la creación de soluciones innovadoras para el control de acceso. Esta integración proporciona una solución completa y automatizada que mejora la seguridad y la eficiencia. Por ejemplo, un sistema de reconocimiento facial puede comunicarse con un Arduino para activar mecanismos físicos como la apertura de puertas, demostrando la viabilidad de estas tecnologías en aplicaciones prácticas (Figura 4).

Figura 4. *Hardware* para el reconocimiento facial integrado con un Arduino.
Fuente: Elaboración propia con base en "Automatización para todos", 2021.

En la figura 4 se observa la conexión entre *hardware* y *software*.

Futuras Direcciones

El futuro del reconocimiento facial se centra en mejorar aún más la precisión y la robustez de los sistemas, así como en abordar los problemas éticos y de privacidad. Igualmente, la investigación continua en algoritmos de *deep learning* y la integración de nuevas tecnologías, como el Internet de las Cosas (IoT), promete expandir las aplicaciones y capacidades del reconocimiento facial.

Materiales y Método

Diseño del Estudio

El estudio se basó en un diseño experimental, donde se desarrolló e implementó un sistema de reconocimiento facial para el control de acceso. El objetivo fue evaluar la efectividad del sistema en diferentes entornos, tanto públicos como domésticos, mediante pruebas de funcionamiento en tiempo real.

Población del Estudio

El estudio se realizó sobre un grupo de individuos seleccionados que fueron clasificados como usuarios autorizados. Estos individuos proporcionaron sus imágenes faciales para la creación de la base de datos de entrenamiento. Se seleccionaron 50 participantes, considerando diversidad en términos de género, edad y etnicidad para garantizar la robustez del sistema.

Entorno

Las pruebas se llevaron a cabo en dos tipos de entornos:

a) **Entorno público:** áreas como parques y salas de eventos, donde se requería restringir el acceso a personas autorizadas.
b) **Entorno doméstico:** un entorno simulado con una puerta controlada por un sistema de acceso que emplea una cámara y una placa Arduino.

Intervenciones

a) Materiales:

- Cámara digital para la captura de imágenes.
- Computadora con capacidad de procesamiento.
- Base de datos de rostros autorizados.
- Placa Arduino para la activación del acceso.
- *Software* de desarrollo (Visual Studio Code).
- Bibliotecas de *software* (OpenCV y dlib).

- Conexión de *hardware* para la comunicación entre la cámara, la computadora y el Arduino.

b) Métodos:

- **Captura de Imágenes de Entrenamiento:** para entrenar el sistema de reconocimiento facial, se capturaron imágenes de las personas autorizadas desde varios ángulos y bajo diferentes condiciones de iluminación.

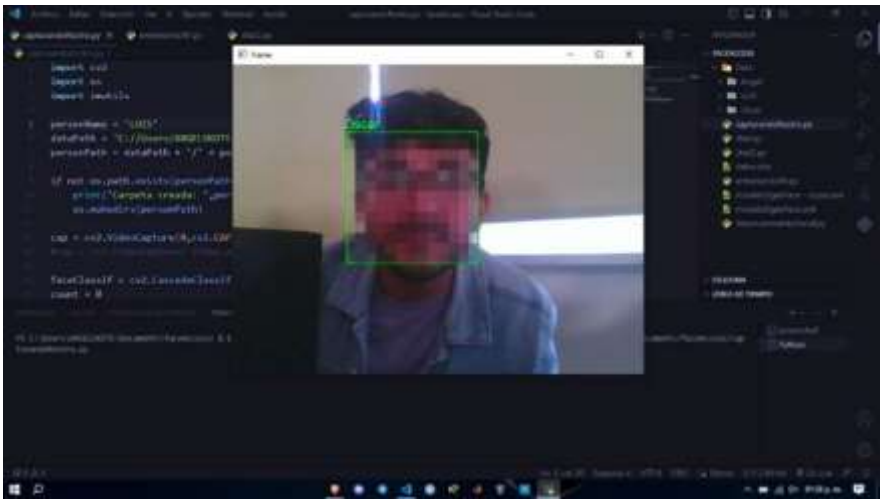


Figura 5. Ejemplo de la captura de imagen.

En la figura 5 se aprecia cómo el sistema detecta el rostro del usuario que se registrará, es aquí donde el sistema toma 250 fotos de 58x58 pixeles con la etiqueta del nombre del usuario.

- **Preprocesamiento de Imágenes:** antes de utilizar las imágenes para el entrenamiento, se aplicaron técnicas de preprocesamiento. Esto incluyó recortar las imágenes para enfocarse en las regiones faciales relevantes, redimensionarlas para obtener una resolución uniforme, y eliminar ruido y normalizar el contraste para mejorar la calidad general de las imágenes.

```
PS C:\Users\ANGELOKOTE\Documents\faceAccess> & C:/
turandoRostro.py
PS C:\Users\ANGELOKOTE\Documents\faceAccess> & C:/
renandoRF.py
Lista de personas:  ['Angel', 'LUIS', 'Oscar']
Leyendo las imágenes
Leyendo las imágenes
Leyendo las imágenes
Entrenando...
```

Figura 6. Preprocesamiento de la imagen.

En la figura 6 se observa la etapa donde el sistema normaliza las características faciales del individuo para entrenar el modelo. Para verificar que todo está correcto, en la consola se muestra la leyenda "Entrenando…".

- **Extracción de Características Faciales:** se utilizaron algoritmos de visión por computadora para extraer características únicas de las imágenes faciales procesadas.

Figura 7. Extracción de características faciales.

En la figura 7 se tiene el archivo generado xml donde están las características, incluyendo la ubicación de puntos clave como ojos, nariz y boca, así como patrones texturales y estructurales específicos del rostro.

- **Entrenamiento del Modelo de Reconocimiento Facial:** con las características faciales extraídas, se procede al entrenamiento del modelo de reconocimiento facial utilizando bibliotecas como OpenCV y dlib en Python. Durante el entrenamiento, el modelo aprende a identificar y distinguir las características faciales de las personas autorizadas, creando así una representación numérica única para cada individuo.
- **Detección y Reconocimiento en Tiempo Real:** una vez que el modelo es entrenado y validado, se implementa el reconocimiento facial en tiempo real.

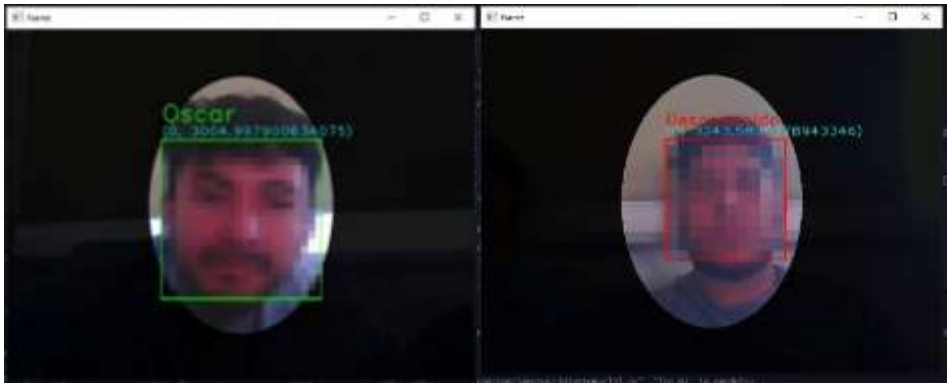


Figura 8. Detección y reconocimiento en tiempo real.

En la figura 8 se aprecia cómo el sistema capturó imágenes de una cámara en vivo y aplicó el algoritmo de reconocimiento facial (EigenFace) para comparar las características detectadas con las características de los rostros autorizados en la base de datos.

De esta manera se observa cómo en el ejemplo de la izquierda se reconoció al usuario llamado "Oscar", mientras que en el ejemplo de la derecha no se reconoció al sujeto en la base de datos, entonces, el sistema lo cataloga como "Desconocido" y el programa no continuará con su flujo normal.

- **Activación del Acceso mediante Arduino:** cuando se identifica una coincidencia exitosa entre el rostro detectado y uno autorizado, se envía una señal desde el sistema de reconocimiento facial a un dispositivo Arduino previamente programado. El Arduino interpreta esta señal y activa el acceso mediante la apertura de una puerta, demostrando así la integración entre el software de reconocimiento facial y los componentes físicos.

Figura 9. Antes de que el sistema reconozca al usuario.

Figura 10. Después de que el sistema si reconociera al usuario en la base de datos de las características faciales.

En las figuras 9 y 10 se verifica cómo el sistema después de haber reconocido exitosamente al usuario, abre una compuerta mediante una placa Arduino. En este caso se usó una caja reciclada con fines de ejemplo, pero este sistema podría implementarse a un entorno real del acceso a usuarios dejando pasar corriente a un lugar o vehículo, con el fin de lograr una mayor seguridad de acceso donde quiera que se implemente.

- **Pruebas y Ajustes:** finalmente, se llevaron a cabo pruebas para verificar el rendimiento y la precisión del sistema. Se realizaron ajustes en los parámetros de reconocimiento facial y en el código del Arduino para optimizar la funcionalidad y asegurar que sea confiable en cuanto al sistema de control de acceso basado en el reconocimiento facial.

Análisis Estadístico

Para analizar los datos y evaluar el rendimiento del sistema, se emplearon pruebas estadísticas descriptivas y de precisión. Se calcularon métricas como la tasa de aciertos, la tasa de falsos positivos y la tasa de falsos negativos. Los análisis se realizaron utilizando *software* estadístico, como Python y sus bibliotecas de análisis de datos (pandas, *numpy*, *scikit-learn*).

Resultados

El proyecto de reconocimiento facial para el control de acceso logró implementar un sistema funcional que valida y permite el acceso únicamente a individuos autorizados mediante el análisis de características faciales. Asimismo, los resultados obtenidos se pueden agrupar en varios aspectos como lo son: precisión del sistema, tiempo de procesamiento, efectividad en diversas condiciones y la integración con el *hardware*.

Precisión del Sistema

El sistema de reconocimiento facial se entrenó utilizando el algoritmo de Eigenfaces, conocido por su eficacia en el reconocimiento de rostros. Después del entrenamiento, se evaluó la precisión del sistema utilizando una base de datos de imágenes de prueba. Los resultados mostraron una alta

tasa de precisión en la identificación de los individuos autorizados, con un porcentaje de reconocimiento superior al 95%. Esta precisión se logró gracias a la calidad de las imágenes de entrenamiento y a la robustez del algoritmo implementado.

Tiempo de Procesamiento

Durante las pruebas, se observó que el sistema podía capturar y analizar las imágenes en tiempo real con un retardo mínimo. El tiempo promedio para la detección y reconocimiento de un rostro fue inferior a un segundo, lo cual es aceptable para aplicaciones de control de acceso donde se requiere una respuesta rápida.

Efectividad en Diversas Condiciones

El sistema fue probado bajo diversas condiciones de iluminación y ángulos de visión para evaluar su viabilidad. Los resultados indicaron que el sistema mantenía un rendimiento consistente en condiciones de iluminación moderada. Sin embargo, en escenarios de iluminación extremadamente baja o con sombras fuertes, la precisión disminuyó ligeramente. Esto sugiere que, para garantizar un rendimiento óptimo, es importante mantener un entorno con iluminación adecuada.

Además, el sistema demostró ser efectivo en la detección y reconocimiento de rostros con diferentes expresiones faciales y accesorios, aunque la precisión podría verse afectada en caso de cambios en cuanto a la apariencia del individuo.

Integración con el *Hardware*

La integración con el *hardware*, en particular con la placa de Arduino y el servomotor, fue exitosa. El Arduino recibió las señales del sistema de reconocimiento facial y activó el servomotor para simular la apertura de una puerta. Las pruebas mostraron que el sistema podía enviar y recibir señales de manera confiable, lo que permitió la activación precisa del mecanismo de control de acceso.

Además, el sistema se configuró para registrar los datos de acceso en un archivo Excel, incluyendo la fecha y hora de cada evento de reconocimiento. Esto proporcionó una capa adicional de seguimiento y análisis de los patrones de acceso.

Ejemplo de archivo de MS Excel creado

Durante las pruebas, se capturaron datos sobre el reconocimiento exitoso de los individuos.
A continuación, se muestra un ejemplo de los datos registrados en el archivo Excel.

	A	B	C	D
1	Número	Nombre	Fecha	Hora
2	1	Angel	12/06/2023	09:54:23
3	2	Angel	12/06/2023	09:57:35
4	3	Angel	12/06/2023	10:09:46
5	4	Angel	12/06/2023	10:10:24
6	5	Oscar	13/06/2023	11:04:33
7	6	Oscar	13/06/2023	13:46:07
8	7	Oscar	13/06/2023	13:46:18
9	8	Angel	13/06/2023	14:05:21
10	9	Angel	14/06/2023	09:28:54
11	10	Angel	14/06/2023	09:30:12

Figura 11. Captura de pantalla de los datos sobre el reconocimiento de individuos.

En la figura 11 se observa una captura de pantalla del archivo MS Excel creado automáticamente desde Python, donde se pueden apreciar los encabezados como “Número”, “Nombre”, “Fecha” y “Hora”.

En la columna de “Número” se lleva a cabo el conteo de todas las personas que fueron reconocidas por el sistema satisfactoriamente, en “Nombre” se tiene el nombre de la persona que fue reconocida con base a la etiqueta que se le asignó durante la etapa de reconocimiento y entrenamiento, en la columna “Fecha” se registra automáticamente la fecha del sistema en la que se reconoció al usuario y así mismo para la columna “Hora”.

Es por ello por lo que estos registros demuestran la capacidad del sistema para identificar correctamente a las personas autorizadas y registrar sus accesos de manera precisa.

Discusión

El proyecto Face Access ha logrado desarrollar un sistema de reconocimiento facial preciso y eficiente para el control de acceso, demostrando utilidad en diversos entornos. Así pues, los resultados muestran una alta precisión en la identificación de individuos autorizados, con una tasa de reconocimiento superior al 95%. Este nivel de precisión es fundamental para garantizar la seguridad y la integridad del sistema, especialmente en entornos donde se requiere un control de acceso riguroso. Además, la implementación del algoritmo de Eigenfaces ha sido importante para alcanzar esta precisión, aprovechando su eficacia en el reconocimiento de rostros.

Por otro lado, el tiempo de procesamiento del sistema ha sido otro aspecto destacado, con un tiempo promedio de detección y reconocimiento de rostros inferior a un segundo. Esta eficiencia en el tiempo de respuesta es fundamental para aplicaciones en tiempo real. Además, la capacidad del sistema para capturar y analizar imágenes en tiempo real con un retardo mínimo garantiza una experiencia fluida y sin interrupciones para los usuarios.

Por otra parte, aunque se observó una ligera disminución en la precisión en condiciones de iluminación extrema, el sistema mantuvo un rendimiento consistente en la mayoría de los escenarios probados. Este hallazgo sugiere que el sistema es capaz de funcionar de manera confiable en una variedad de entornos y situaciones, lo que lo hace adecuado para aplicaciones prácticas en el mundo real.

Sumado a esto, la integración del sistema de reconocimiento facial con *hardware*, especialmente con una placa Arduino y servomotores, ha sido exitosa y efectiva. La capacidad del sistema para comunicarse de manera confiable con el *hardware* y activar el acceso mediante la apertura de una puerta demuestra su capacidad para integrarse en sistemas físicos existentes. Esta integración proporciona una solución completa y automatizada para el control de acceso, mejorando la seguridad y la eficiencia en la gestión de accesos.

Cierre

El proyecto ha demostrado ser una solución eficaz y viable para modernizar y mejorar los sistemas de control de acceso. Mediante la implementación de un algoritmo de reconocimiento facial basado en Eigenfaces y su integración con hardware como Arduino y servomotores, además, se

logró crear un sistema que no solo es preciso y rápido, sino también robusto y adaptable a diversas condiciones.

Principales Logros

1. **Alta Precisión de Reconocimiento**: el sistema alcanzó una tasa de reconocimiento superior al 95%, demostrando su capacidad para identificar correctamente a las personas autorizadas.
2. **Eficiencia en el Tiempo de Procesamiento**: el tiempo de respuesta del sistema fue inferior a un segundo, lo cual es importante para aplicaciones en tiempo real donde la inmediatez es esencial.
3. **Adaptabilidad a Diversas Condiciones**: aunque el rendimiento óptimo se obtuvo en condiciones de iluminación adecuada, el sistema mostró una buena utilidad en diferentes escenarios, incluyendo variaciones en la expresión facial y en el uso de accesorios.
4. **Integración Exitosa con Hardware**: la comunicación efectiva entre el sistema de reconocimiento facial y el *hardware* Arduino permitió la activación precisa de un mecanismo de control de acceso, demostrando la posibilidad de implementar una solución completa y automatizada.
5. **Registro de Accesos**: la capacidad de registrar los eventos de acceso en un archivo Excel añade una capa adicional de seguimiento y análisis, facilitando la gestión y el monitoreo del sistema.

Contribuciones y Futuro

Este proyecto no solo valida el uso del reconocimiento facial como una alternativa viable a los métodos tradicionales de control de acceso, sino que también subraya el potencial de los algoritmos de Python y la integración con plataformas de *hardware* como Arduino para desarrollar soluciones innovadoras y prácticas. La combinación de *software* avanzado con *hardware* accesible y económico ofrece una ruta prometedora para futuras aplicaciones en seguridad y automatización.

Finalmente, el proyecto ha establecido una base sólida para el desarrollo de sistemas de control de acceso modernos y eficientes, demostrando la capacidad de la tecnología de

reconocimiento facial para ofrecer soluciones seguras y convenientes en la gestión de accesos. Es por ello que se logró el objetivo de "Face Access" en crear un sistema de control de acceso seguro y eficiente.

Referencias

Aguilar Figueroa, R., Castrejón González, G., & Mejía Moreno, C. (2020). Reconocimiento de rostros térmicos usando redes neuronales convolucionales. Revista Digital Innovacion y Desarrollo Tecnologico, Vol. 12. No.21 enero-marzo. En : https://iydt.wordpress.com/wp-content/uploads/2020/04/1-4_reconocimiento-de-rostros-tc3a9rmicos-usando_.pdf

Altayeb, M., & Al-Ghraibah, A. (2023). Arduino Based Real-Time Face Recognition And Tracking System. *International Journal of Advanced Trends in Computer Science and Engineering*, 144-150. DOI:10.30534/ijatcse/2023/011242023

Alvarez N. A., Marañon R. E. J., & Orozco M. R. (2022). Revisión de los métodos de reconocimiento facial en imágenes RGB-D adquiridas mediante un sensor Kinect. Revista Cubana de Ciencias Informaticas. Vol 16. No.2. En http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2227-18992022000200157

Cadena M. J. A., Montaluisa P. R. H., Flores L. G. A., Chancúsig C. J. C., & Guaypatín P. O. A. (2017). Reconocimiento facial con base en imágenes. Revista Redipe Vol. 6 No. 5. En https://revista.redipe.org/index.php/1/article/view/267/264

Espinoza O. D. E., & Jorquera G. P. I. (2015). Reconocimiento Facial. Pontifica Universidad Catolica de Valparaiso. Facultad e Ingenieria. En http://opac.pucv.cl/pucv_txt/txt-1000/UCD1453_01.pdf

García G. M., Antonio C., M., & Castañeda M. A. Y. (2021). A Review on Face Recognition systems -based Access Control Systems. *International Congress on Human-Computer Interaction, Optimization and Robotic Applications. Vol 79 (HORA)*, 1-5.

Haji , S., & Varol, A. (2016). Real Time Face Recognition System (RTFRS). 4 *International Symposium on Digital Forensics and Security*, 107-111. En ISDFS-2016-Proceedings.suad_asaf.pdf.

Hernández D, M., & Plasencia C, Y. (2016). Aprendizaje de métrica para el reconocimiento de rostros a partir de imágenes de baja resolución. Revista Cubana de Ciencias Informaticas, Vol.10, no.1 enero-marzo. La Habana Cuba. En: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2227-18992016000100009

Joaquin A., R. (2021). Reconocimiento facial con deep learning y python. En: https://cienciadedatos.net/documentos/py34-reconocimiento-facial-deeplearning-python

Pérez R. D. (2021). Reconocimiento eficiente de caras mediante Deep. En: https://uvadoc.uva.es/bitstream/handle/10324/50029/TFG-G5204.pdf?sequence=1

Ramadan TH. H. & Amira B. S. (2021). Face Detection and Recognition Using OpenCV . *Journal of Soft Computing and Data Mining*, Vol. 2. No.1, pag 86-97. En : DOI:10.30880/jscdm.2021.02.02.008

Waqar , A., Wenhong , T., Salah Ud, D., Desire, I., & Abdullah, A. K. (2020). *Classical and modern face recognition approaches:a complete review.* En: https://www.researchgate.net/publication/344610440_Classical_and_modern_face_recognition_approaches_a_complete_review

APLICACIÓN DE K-MEDIAS PARA IDENTIFICACIÓN DE PATRONES DE INCIDENCIA DELICTIVA EN MÉXICO

APPLICATION OF K MEANS FOR IDENTIFICATION OF CRIME INCIDENCE PATTERNS IN MEXICO

Martha Gloria Pluma Torres[1]
mplumat639@alumno.uaemex.mx
ORCID: 0009-0007-3300-1586

Israel Gutiérrez González[2]
igutierrezg@uaemex.mx
ORCID: 0000-0003-2984-7654

Resumen

Hoy en día existe un gran impacto en el uso de técnicas de minería de datos con el fin de brindar soluciones a diferentes problemas sociales., como en este caso en la delincuencia y sus diferentes afectaciones en la sociedad, siendo una de ellas el algoritmo de k medias que se caracteriza por agrupar patrones con similitudes y en este trabajo se utiliza para identificar qué tipos de delitos afectan más a hombres y mujeres en los diferentes estados de la república Mexicana, siendo los principales factores Entidad Federativa, Tipo de Delito, Subtipo de delito, Modalidad, Sexo, Enero-Diciembre. Encontrando los estados donde existe más incidencia delictiva y los delitos que más cometidos.

Palabras clave: Minería de Datos, K-medias, Incidencia delictiva, agrupamiento, delito, weka.

[1] Ingeniería en Sistemas Inteligentes, Centro Universitario UAEM Nezahualcóyotl, Universidad Autónoma del Estado de México
[2] Ingeniería en Sistemas Inteligentes, Centro Universitario UAEM Nezahualcóyotl, Universidad Autónoma del Estado de México

Abstract

Today there us a great impact of using data mining techniques in order to provide solutions to different social problems, such as in this case crime and its different effects on society, one of them being the k-means algorithm that is characterized by grouping patterns with similarities and in this work it is used to identify which types of crimes most affect men and women in the different states of the Mexican Republic, the main factors being Federal Entity, Type of Crime, Subtype of crime, Modality, Sex , January December. Finding the states where there is the highest incidence of crime and the crimes that are most committed.

Keywords: Data Mining, K-means, Crime incidence, clustering, crime, Weka.

Fecha de envío: 05/06/2024
Fecha de aprobación: 10/11/2024
Fecha de publicación: 01/01/2025

Introducción

A través de la Historia, los actos delictivos han sido un problema que han perdurado en el mundo, y a pesar de que cada año existen millones de personas afectadas por algún tipo de delito, al ser tan común, no se logran ver las verdaderas dimensiones que afectan a la calidad de vida en la sociedad, esto no solo trae consigo daños a la dignidad de la víctima, sino que también afecta a la estabilidad económica, social, familiar y laboral. Trayendo consigo, daños irreversibles (García, 2021), y, en algunas ocasiones, arrebatándoles su futuro y la libertad por ejercer sus derechos. Como su definición lo indica, la delincuencia es cualquier acto de comisión u omisión y cualquier condición que resulte de dichos actos, que prive a los sujetos de igualdad de derechos y libertades y lo interfiera con su máximo desarrollo y libertad de elegir (Carmona,1999). Según (Vilalta, et.al, 2016), América Latina enfrenta grandes desafíos pues es la región más insegura.

En México, la violencia es unos de los principales problemas que afectan a la nación, siendo así, un objeto de estudio prioritario. El Instituto Nacional de Estadística, Geografía e Informática (INEGI, 2023) a lo largo de los años, ha realizado encuestas nacionales de victimización y percepción sobre seguridad pública, donde se estimó que 10.5 millones de hogares (27.4 % del total del país) tuvieron, al menos, una o uno de sus integrantes como víctima del delito,

esto viene siendo un obstáculo para el desarrollo económico puesto que las inversiones que se podrían hacer en negocios u otros problemas sociales, van destinados a temas de seguridad, como el caso de policías públicos y privados, cárceles, seguros, entre otros. Según el instituto Mexicano de competitividad (IMCO) se invierte aproximadamente el 15% del PIB (Producto interno bruto), es decir, 108 mil millones de pesos (IMCO, 2006; Jiménez, 2021).

La dimensión de información registrada día a día en México, nos lleva a utilizar a la minería de datos debido a que permite desarrollar algoritmos capaces de analizar grandes cantidades de datos en prevención criminal. Entre unas de sus técnicas de agrupación existe el K-medias que es utilizado por su facilidad de implementación y resultados eficaces (Valenga, *et.al*, 2008).

Por lo tanto, el objetivo es utilizar el método de K-medias para identificar los delitos que más afectan a hombres y mujeres en cada uno de los estados de México.

Antecedentes

La incidencia delictiva es un problema que se enfrenta a nivel mundial, por ello, alrededor del mundo se han creado diferentes estudios, buscando posibles soluciones al problema.

García (2021) en su artículo Factores Criminógenos en Jóvenes y su Integración en la Delincuencia Organizada, ha demostrado que los jóvenes entran a la delincuencia organizada con la finalidad de obtener beneficios más rápidos y, por ende, lo ven como un buen negocio y forma de ganarse la vida.

Ortiz (2019) propuso un plan de seguridad para la República Dominicana que cuenta con cinco puntos que son utilizados por sistemas de las TIC. Utilizar el enfoque punitivo, estrategia de control de drogas, programa para vivir seguro y otras observaciones del Plan Seguridad Ciudadana.

En Argentina al aplicar la minería de datos para identificar patrones delictivos (Valenga, *et.al,* 2008), usando el algoritmo de k-medias y árboles de decisión, creando tres diferentes clústeres donde se demostró que el 43% de homicidios son en vía pública y con arma de fuego, 35 % son homicidios sin arma y en un domicilio particular y el 22% han sido por un robo y con arma de fuego; a través del algoritmo J48 las reglas creadas fueron en robo la víctima y el agresor no se conocen y los demás son homicidios conflictivos, demostrando que la mayoría de los homicidios dolosos por problemas interpersonales y no robos.

Martín *et.al* (s.f) en su estudio de aplicación de K medias y SOM (Self Oganizing Maps) donde se analiza los micro accidentes en Madrid se buscó facilitar la asignación de responsabilidad a los conductores implicados en algún accidente de tráfico, explorando el impacto, social, ambiental y económico que generan estos accidentes.

Según los trabajos previamente mencionados la delinquía es un problema que afecta mundialmente en la sociedad por ende es y seguirá siendo un objeto de estudio, con la finalidad de encontrar nuevas tácticas para reducir la afectación en la sociedad.

Materiales y Métodos

Para este estudio, se tomó el conjunto de datos de Incidencia delictiva provenientes de Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública (s.f), este abarcaba 9 años desde el 2015 al 2023 con 66,241 datos, pero para la investigación solo se tomaron en cuenta dos años 2015 y 2023, con la finalidad de hacer una vez teniendo los datos se realizaron los pasos del método de KDD (Knowledge Discovery in Databases) que nos permite el descubrimiento de conocimientos útiles no triviales (Causa común, 2021).

Tipo de Delito	Subtipo de delito	Modalidad
Corrupción de menores	Corrupción de menores	Corrupción de menores
Otros delitos contra la sociedad	Otros delitos contra la sociedad	Otros delitos contra la sociedad
Tráfico de personas	Tráfico de personas	Tráfico de personas
Extorsión	Extorsión	Extorsión
Homicidio, lesiones y feminicidio	1. Culposo	Con arma de fuego
		Con arma blanca
		Con otro elemento
		No especificado
	2. Doloso	Con arma de fuego
		Con arma blanca
		En accidente de transito
		Con otro elemento
		No especificado
Otros delitos que atentan contra la vida y la integridad corporal	Otros delitos que atentan contra la vida y la integridad corporal	Otros delitos que atentan contra la vida y la integridad corporal
Otros delitos que atentan contra la libertad personal	Otros delitos que atentan contra la libertad personal	Otros delitos que atentan contra la libertad personal
Secuestro	Secuestro	Secuestro extorsivo
		Secuestro con calidad de rehén
		Secuestro para causar daño
		Secuestro exprés
		Otro tipo de secuestros
Rapto	Rapto	Rapto
Tráfico de menores	Tráfico de menores	Tráfico de menores

Tabla 1: Descripción de clasificación de delitos

En este caso, no se identificaron datos inconsistentes, se separó la información basándose en el sexo, este campo contaba con tres diferentes datos registrados: hombre, mujer y no especificado, basándose en el objetivo del análisis solo se tomaron en cuenta los registros de hombre y mujer; obteniendo las siguientes 4 bases de datos que son las que finalmente se procesarían.

• Hombres 2015 con la cantidad de 2883 registros delictivos
• Hombres 2023 con la cantidad de 2883 registros delictivos
• Mujeres 2015 con la cantidad de 3255 registros delictivos
• Mujeres 2023 con la cantidad de 3255 registros delictivos

Desarrollo

El procesamiento de los datos se llevó a cabo por medio del uso del algoritmo K -medias es una técnica muy conocida como método de agrupación de datos (Gutiérrez *et.al*., 2020). Este recibe un parámetro "K" y procede a dividir los "n" objetos en los "k" grupos, por medio de distancia euclidiana midiendo la similitud desde el punto medio de los grupos, una ventaja de trabajar con k medias es que nos permite establecer los puntos iniciales "k" (Marante, s.f), en este caso, nuestros puntos iniciales son k = 2, k=3 y k=4,en cada una de las bases de datos se procesaron por medio de Weka (Waikato Environment for Knowledge Analisys), que está desarrollado en Java y nos permite usar técnicas de minería de datos (Corso, s.f). Como anteriormente se mencionó, la agrupación se genera por distancias, para esto se necesita datos numéricos, por ende, nuestras variables fueron los meses de enero a diciembre, pues en ellos, están registrados los casos de delitos en las diferentes entidades de los estados de México.

En cada uno de los casos se fueron generando los clústeres de acuerdo con las similitudes de la incidencia delictiva.

A continuación, en la tabla 2, se describirán cada uno de los grupos asignados por el k medias en hombres y mujeres. En caso de k =2, donde se podrá ver los porcentajes por grupo y números de casos.

Hombres 2015 Y 2023			Mujeres 2015 Y 2023		
Clúster	Número de casos	Porcentaje	Clúster	Número de casos	Porcentaje
0	1431	49.6%	0	1058	33%
1	1452	50.4%	1	2197	67%

Tabla 2: Descripción de grupos para k=2 en hombres y mujeres del año 2015 y 2023

Donde se visualiza cómo es que tanto hombres como mujeres, en ambos años, se agruparon con el mismo número de casos por clúster. Ahora en la tabla 3 se muestra la descripción de los grupos creados en k =3 en hombres y en mujeres de ambos años.

Hombres 2015			Hombres 2023		Mujeres 2015			Mujeres 2023	
Clúster	Número de casos	Porcentaje	Número de casos	Porcentaje	Clúster	Número de casos	Porcentaje	Número de casos	Porcentaje
0	1103	38.3%	996	34.5%	0	650	20%	649	19.9%
1	1014	35.2%	1116	38.7%	1	1969	60.5%	1968	60.5%
2	766	26.5%	772	26.8%	2	636	19.5%	638	19.6%

Tabla 3: Descripción de grupos para k=3 en hombres y mujeres del año 2015 y 2023

Como se puede observar, los clústers fueron asignados con una diferencia en los números de casos. Finalmente, en la tabla 4, se puede visualizar cómo fueron conformados los grupos para k = 4, donde, al igual que en k = 2, fueron agrupados con la misma cantidad de números de casos en hombres y mujeres de ambos años.

Hombres 2015			Hombres 2023		Mujeres 2015 y 2023		
Clúster	Número de casos	Porcentaje	Número de casos	Porcentaje	Clúster	Número de casos	Porcentaje
0	941	33 %	940	33 %	0	573	17.6%
1	897	31%	898	31%	1	1851	56.8%

2	696	24%	695	24%	2	584	17.9 %
3	349	12 %	350	12 %	3	247	7.5%

Tabla 4: Descripción de grupos para k=4 en hombres y mujeres del año 2015 y 2023

Resultados

En cada uno, aunque tuvieron diferente dispersión de delitos por grupo, todas las entidades federativas estuvieron tomando en cuenta estos datos. Se calcularon los promedios naciones de las entidades federativas para identificar los estados de mayor incidencia delictiva y los tipos de delitos que más existen.

Teniendo esto en cuenta, en la tabla 5, se muestran los promedios nacionales en k =2, k=3 y k=4 en sus respectivos clústeres.

Promedios Nacionales De K = 2				
Clúster	Hombres 2015	Hombres 2023	Mujeres 2015	Mujeres 2023
0	1.7080712	2.7036454	0.782766	1.147763
1	3.814815	9.1749885	3.587013	4.172318
Promedios Nacionales De K =3				
Clúster	Hombres 2015	Hombres 2023	Mujeres 2015	Mujeres 2023
0	2.096857	3.418675	0.702051	0.64869
1	9.558596	11.42608	3.634798	4.054074
2	0.849869	1.341764	1.722615	3.105799
Promedios Nacionales De K =4				
Clúster	Hombres 2015	Hombres 2023	Mujeres 2015	Mujeres 2023
0	2.628055	4.085904	0.541594	0.360239

1	10.38127	12.90804	3.805556	4.133441
2	0.795019	1.730456	1.4248	2.233019
3	0.910936	1.589048	2.11471	4.93691

Tabla 5: Promedios nacionales en por clúster en k=2, k=3, k=4 de hombres y mujeres en los años 2015 y 2023

Esto se realiza con la finalidad de que al calcular los promedios por entidad federativa en los clúster de k = 2, k= 3 y k = 4, se logra identificar los estados con mayor y menor incidencia delictiva basándose en el parámetro del promedio nacional por clúster, de esta manera, los estados que superen el promedio, serán identificados como los de alta incidencia delictiva, mientras que los que estén por debajo del promedio estarán en los de baja incidencia delictiva, sin embargo, solamente se describirán los 3 estados con mayor promedio de incidencia y los 3 menores, como se muestra en la tabla 6.

Clúster 0		Clúster 1	
Estado	Promedio de delito	Estado	Promedio de delito
México	5.38	México	48.24
Guanajuato	4.88	Nuevo León	17.19
Michoacán	4.15	Puebla	13.15
Nayarit	0.35	Colima	0.17
Yucatán	0.32	Campeche	0.35
Campeche	0.15	Aguascalientes	0.11

Tabla 6: Ejemplo de orden de estados de alta y baja incidencia delictiva basándose en el promedio nacional por clúster 0 2015

En la tabla 7 se describen de los promedios por entidad federativa para k= 2 en hombres del 2015 y 2023.

Clúster 0				Clúster **1**			
Hombres 2015 = 1.70		Hombres 2023 = 2.70		Hombres 2015 = 3.8		Hombres 2023= 9.1	
Estado	Promedio de delito	Estado	Promedio de delito	Estado	Promedio de delito	Estado	Promedio de delito
México	5.38	Guanajuato	9.26	México	48.24	México	51.94
Guanajuato	4.88	México	8.95	Nuevo León	17.19	Guanajuato	16.44
Michoacán	4.15	Baja california	6.34	Puebla	13.15	Nuevo León	16.28
Nayarit	0.35	Baja California sur	0.36	Colima	0.17	Yucatán	0.77
Yucatán	0.32	Tlaxcala	0.20	Campeche	0.35	Tlaxcala	0.39
Campeche	0.15	Yucatán	0.19	Aguascalientes	0.11	Aguascalientes	0.20

Tabla 7: Promedios de los 3 estados de alta y baja incidencia delictiva por clúster en hombres del 2015 y 2023 para k =2

En la tabla 8 se describen los delitos que se registraron en los estados de alta y baja incidencia delictiva de k = 2.

Clúster 0 Hombres 2015			
Estado	Delito	Meses	Casos
México	Homicidio Doloso con arma de fuego	Enero, septiembre y diciembre	2,104

Guanajuato	homicidio culposo en accidente de tráfico o con otro elemento	Agosto y noviembre	1,958
Michoacán	homicidio doloso con arma de fuego y homicidio culposo en accidente de tráfico o con otro elemento	mayo y diciembre	1, 753
Nayarit	homicidio doloso con arma de fuego Lesiones dolosas no especificadas	abril y agosto enero y octubre	129 156
Yucatán	Homicidio culposo en accidente de trafico	Junio	31
Campeche	Homicidio culposo en accidente de trafico	Marzo, junio y Diciembre	77
Clúster 0 Hombres 2023			
Guanajuato	Homicidio Doloso con arma de fuego	Marzo y mayo	3,748
México	Homicidio Doloso con arma de fuego	Enero	3,047
Baja california	Homicidio Doloso con arma de fuego	Agosto y octubre	2,120
Baja California sur	'Otros delitos que atentan contra la vida y la integridad corporal'	Julio	59
Tlaxcala	Homicidio Doloso con arma de fuego	Mayo y junio	135
Yucatán	Homicidio Doloso con arma de fuego	Marzo y diciembre	131
Clúster 1 Hombres 2015			
México	Lesiones dolosas con otro elemento Lesiones culposas con otro elemento y en accidente de transito	Septiembre y octubre	30,985
Nuevo León	Lesiones dolosas con otro elemento	Agosto	6,086
Puebla	Lesiones dolosas con otro elemento	Enero – Diciembre	4,676
Colima	Lesiones dolosas con otro elemento	Enero y noviembre	208
Campeche	Lesiones dolosas con otro elemento	Mayo y agosto	171
Aguascalientes	Lesiones dolosas con otro elemento	Junio	312
Clúster 1 hombres 2023			

México	Lesiones dolosas con otro elemento	Marzo, septiembre y octubre	30,762
Guanajuato	Lesiones dolosas con otro elemento	Marzo, mayo y noviembre	9,698
Nuevo León	Lesiones dolosas con otro elemento	Marzo y agosto	7,033
Yucatán	Lesiones dolosas con otro elemento	Julio y noviembre	181
Tlaxcala	Lesiones dolosas con otro elemento	Junio	158
Aguascalientes	Homicidio doloso		

Tabla 8: Tabla de descripción de delitos, números de casos de k=2 del 2015 y 2023

En la tabla 9 se describen los promedios por entidad federativa para k= 3 en hombres del 2015 y 2023.

Clúster 0				Clúster 1			
Hombres 2015 = 2.0		Hombres 2023 = 3.41		Hombres 2015 = 9.55		Hombres 2023= 11.42	
Estado	Promedio de delito	Estado	Promedio de delito	Estado	Promedio de delito	Estado	Promedio de delito
México	6.52	Guanajuato	13.45	México	66.19	México	66.93
Guanajuato	6.26	México	11	Nuevo León	21.42	Guanajuato	21.29
Michoacán	4.43	Michoacán	8.73	Puebla	16.60	Veracruz	19.64
Nayarit	0.43	Baja California sur	0.44	Colima	0.5	Yucatán	0.96
Yucatán	0.35	Tlaxcala	0.27	Campeche	0.46	Tlaxcala	0.5
Campeche	0.23	Yucatán	0.26	Aguascalientes	0.15	Aguascalientes	0.26

Clúster 2	
Hombres 2015 = 0.84	Hombres 2023 = 1.34

Estado	Promedio de delito	Estado	Promedio de delito
Nuevo León	4.21	Nuevo León	5.85
Tamaulipas	3.53	Baja California	3.97
Puebla	3.34	Veracruz	3.72
Tlaxcala	0.03	Yucatán	0.034
Colima	0.2	Nayarit	0.027
Aguascalientes	0.006	Tlaxcala	0.010

Tabla 9: Promedios de los 3 estados de alta y baja incidencia delictiva por clúster en hombres del 2015 y 2023 para k =3

En la tabla 10 se describen los delitos que se registraron en los estados de alta y baja incidencia delictiva de k = 3.

Clúster 0 Hombres 2015			
Estado	Delito	Meses	Casos
México	Homicidio Doloso con arma de fuego	Enero	2104
Guanajuato	homicidio culposo en accidente de tráfico o con otro elemento	Agosto y noviembre	1958
Michoacán	Homicidio culposo en accidente de tráfico o con otro elemento	Mayo y diciembre	1773
Nayarit	homicidio doloso con otro elemento y no especificado	Abril, mayo y Junio	156
Yucatán	Otros delitos que atentan contra la vida y la integridad corporal'	Abril y octubre	60
Campeche	Homicidio culposo en accidente de tráfico o con otro elemento	Enero y mayo	77
Clúster 0 Hombres 2023			

Guanajuato	Homicidio Doloso con arma de fuego	Marzo y mayo	3748
México	Homicidio Doloso con arma de fuego	Enero y septiembre	3047
Michoacán	Homicidio Doloso con arma de fuego	Julio y diciembre	2394
Baja California sur	Homicidio Doloso con arma de fuego	Marzo y noviembre	68
Tlaxcala	Homicidio Doloso con arma de fuego	Mayo y junio	135
Yucatán	Homicidio culposo en accidente de tránsito y doloso con otro elemento	Marzo y diciembre	131
	Clúster 1 Hombres 2015		
México	Lesiones dolosos con otro elemento	Abril, mayo y octubre	30985
Nuevo León	Lesiones dolosos con otro elemento	Mayo, junio, septiembre y octubre	6086
Puebla	Lesiones dolosos con otro elemento	Enero y abril	4676
Colima	Lesiones dolosos con otro elemento		
Campeche	Lesiones dolosos con otro elemento	Mayo y diciembre	171
Aguascalientes	Extorsión	Febrero y mayo	22
	Clúster 1 hombres 2023		
México	Lesiones dolosos con otro elemento	Marzo, junio y septiembre	30762
Guanajuato	Lesiones dolosos con otro elemento	Marzo, mayo y septiembre	9698
Veracruz	Lesiones dolosos con otro elemento	Mayo, junio y octubre	5930
Yucatán	Lesiones dolosos con otro elemento	Octubre y noviembre	181

Tlaxcala	Lesiones dolosos con otro elemento	Junio	158
Aguascalientes	Lesiones dolosos con arma de fuego	Enero y octubre	75
	Clúster 2 hombres 2015		
Nuevo León	Otros delitos que atentan contra la libertad personal	Julio, agosto y septiembre	934
Tamaulipas	Otros delitos que atentan contra la libertad personal	Enero	716
Puebla	Otros delitos contra la sociedad'	Mayo, septiembre y octubre	472
Tlaxcala	Extorsión Secuestro	Enero, mayo y junio Enero, mayo y julio	5 6
Colima	Otros delitos que atentan contra la libertad personal	Febrero, marzo y mayo	4
Aguascalientes	Secuestro	Agosto y octubre	3
	Clúster 2 hombres 2023		
Nuevo León	'Otros delitos que atentan contra la libertad personal'	Enero y agosto	1661
Baja California	'Otros delitos que atentan contra la libertad personal'	Agosto y septiembre	1085
Veracruz	'Otros delitos que atentan contra la libertad personal'	Agosto y septiembre	1046
Yucatán	'Otros delitos que atentan contra la libertad personal'	Marzo, julio y septiembre	10
Nayarit	Secuestro	Julio	5
Tlaxcala	Secuestro	septiembre	3

Tabla 10: Tabla de descripción de delitos, números de casos de k=3 del 2015 y 2023

En la tabla 11 se describen los promedios por entidad federativa para k= 4 en hombres del 2015 y 2023

Clúster 0				Clúster 1			
Hombres 2015 = 2.62		Hombres 2023 = 4.08		Hombres 2015 = 1.38		Hombres 2023= 12.9	
Estado	Promedio de delito	Estado	Promedio de delito	Estado	Promedio de delito	Estado	Promedio de delito
México	8.69	Guanajuato	12.5	México	71.93	México	76.44
Guanajuato	7.21	México	11.09	Nuevo León	21.95	Guanajuato	27.41
Jalisco	6.84	Jalisco	9.8	Baja california	1.31	Ciudad de México	20.52
Tlaxcala	0.58	Yucatán	0.6	Campeche	0.51	Yucatán	0.62
Yucatán	0.52	Tlaxcala	0.33	Ciudad de México	0.31	Tlaxcala	0.54
Campeche	.29	Baja california sur	0.24	Aguascalientes	0.065	Aguascalientes	0.052

Clúster 2				Clúster 3			
Hombres 2015 = 0.79		Hombres 2023 = 1.73		Hombres 2015 = 0.91		Hombres 2023= 1.58	
Estado	Promedio de delito	Estado	Promedio de delito	Estado	Promedio de delito	Estado	Promedio de delito
Nuevo León	5.04	México	7.93	Puebla	7.37	Ciudad de México	8.88
Tamaulipas	3.5	Baja california	7.49	México	2.66	Baja California	7.4
México	2.11	Nuevo León	5.8	Ciudad de México	2.26	Guanajuato	4.4
Tlaxcala	0.021	Guanajuato	0.012	Oaxaca	0.09	Nayarit	0.1

Quintana roo	0.014	Tlaxcala	0.01	Colina	0.56	Yucatán	0.01
Aguascalientes	0.009	Aguascalientes	0.005	Tlaxcala	0.21	Tlaxcala	0

Tabla 11: Promedios de los 3 estados de alta y baja incidencia delictiva por clúster en hombres del 2015 y 2023 para k =4

En la tabla 12 se describen los delitos que se registraron en los estados de alta y baja incidencia delictiva de k = 4

Clúster 0 Hombres 2015			
Estado	Delito	Meses	Casos
México	Homicidio doloso con arma de fuego	Enero	2503
Guanajuato	Homicidio culposo con otro elemento y en accidente de trafico	Agosto y noviembre	2077
Jalisco	Homicidio doloso con arma de fuego	Enero y octubre	1232
Tlaxcala	Homicidio culposo con otro elemento y en accidente de trafico	Abril, septiembre y octubre	235
Yucatán	Homicidio doloso con arma blanca	Mayo	89
Campeche	Homicidio doloso con arma de fuego	Junio	85
Clúster 0 hombres 2023			
Guanajuato	Homicidio doloso con arma de fuego	Marzo, mayo y junio	3600
México	Homicidio doloso con arma de fuego	Enero a diciembre	3193
Jalisco	Homicidio doloso con arma de fuego	Enero	1767
Yucatán	Homicidio doloso con arma de fuego	Marzo y diciembre	142
Tlaxcala	Homicidio doloso con arma de fuego	Mayo	142

Baja california sur	Homicidio doloso con arma de fuego	Noviembre a diciembre	70
	Clúster 1 Hombres 2015		
México	Lesiones culposas con otro elemento y accidente de trafico	Agosto, septiembre y octubre	31027
Nuevo León	Lesiones culposas con otro elemento y accidente de trafico	Mayo, septiembre y octubre	6321
Baja california	Lesiones dolosas con no identificado	Junio y septiembre	7476
Campeche	Lesiones culposas en accidente de tráfico y dolosas con otro elemento	Agosto	171
Ciudad de México	Homicidio doloso con arma de fuego	Abril a noviembre	138
Aguascalientes	Lesiones culposas con otro elemento y accidente de trafico	Junio	16
	Clúster 1 hombres 2023		
México	Lesiones dolosas con otro elemento	Marzo, septiembre y octubre	30864
Guanajuato	Lesiones dolosas con otro elemento	Marzo, septiembre y noviembre	9766
Ciudad de México	Lesiones dolosas con otro elemento	Octubre y noviembre	7651
Yucatán	Lesiones dolosas con otro elemento	Octubre y noviembre	175
Tlaxcala	Lesiones dolosas con otro elemento	Junio y julio	156
Aguascalientes	Homicidio doloso con otro elemento	Marzo y junio	13

Clúster 2 hombres 2015			
Nuevo León	Otros delitos que atentan contra la libertad personal'	Julio y agosto	1420
Tamaulipas	Otros delitos que atentan contra la libertad personal'	Enero	2180
México	Otros delitos que atentan contra la libertad personal'	Octubre y noviembre	410
Tlaxcala	Secuestro	Enero	6
Quintana roo	Secuestro	Enero	4
Aguascalientes	Secuestro	Agosto y octubre	2
Clúster 2 hombres 2023			
México	'Otros delitos que atentan contra la libertad personal'	Enero y agosto	1933
Baja california	'Otros delitos que atentan contra la libertad personal'	Junio y octubre	1882
Nuevo León	'Otros delitos que atentan contra la libertad personal'	Enero y agosto	1661
Guanajuato	Secuestro	Febrero y diciembre	2
Tlaxcala	Secuestro	Julio y agosto	3
Aguascalientes	Secuestro	Diciembre	1
Clúster 3 hombres 2015			
Puebla	Otros delitos contra la sociedad'	Enero y marzo	651
México	Extorsión	Marzo y mayo	288
Ciudad de México	Extorsión	Octubre	353
Oaxaca	Otros delitos contra la sociedad'	Agosto	7
Colima	Extorsión	Octubre y noviembre	8
Tlaxcala	Extorsión	Marzo y abril	2

Clúster 3 hombres 2023			
Ciudad de México	'Otros delitos contra la sociedad'	Marzo y octubre	678
Baja California	'Otros delitos que atentan contra la libertad personal'	Marzo y septiembre	821
Guanajuato	Extorsión	Agosto y septiembre	489
Nayarit	Corrupción de menores	Septiembre	6
Yucatán	Corrupción de menores	Enero	1
Tlaxcala			0

Tabla 12: Tabla de descripción de delitos, números de casos de k=4 del 2015 y 2023

Ahora describiremos los resultados en caso de mujeres, en la tabla 13 de los promedios por entidad federativa para k= 2 en hombres del 2015 y 2023

Clúster 0				Clúster 1			
Mujeres 2015 = 0.78		Mujeres 2023 = 1.14		Mujeres 2015 = 3.58		Mujeres 2023= 4.17	
Estado	Promedio de delito	Estado	Promedio de delito	Estado	Promedio de delito	Estado	Promedio de delito
Nuevo León	3.86	México	6.51	México	35.82	México	25.2
Tamaulipas	2.23	Baja California	5.68	Baja California'	8.22	Guanajuato	10.6
Puebla	1.84	Nuevo León	4.42	Nuevo León	7.30	Nuevo León	6.87
Guanajuato	0.013	Aguascalientes	0.01	Campeche	0.16	Aguascalientes	0.16
Campeche	0.01	Yucatán	0.003	Colima	0.11	Yucatán	0.15
Colima	0.006	Tlaxcala	0.0019	Aguascalientes	0.006	Tlaxcala	0.14

Tabla 13: Promedios de los 3 estados de alta y baja incidencia delictiva por clúster en mujeres del 2015 y 2023 para k =2

En la tabla 14 se describen los delitos que se registraron en los estados de alta y baja incidencia delictiva de k = 2.

Clúster 0 mujeres 2015			
Estado	Delito	Meses	Casos
Nuevo León	'Otros delitos que atentan contra la libertad personal'	Julio y noviembre	1548
Tamaulipas	Lesiones culposo con otro elemento y en accidente de trafico	Mayo y octubre	731
Puebla	'Otros delitos que atentan contra la libertad personal'	Marzo, septiembre y noviembre	509
Guanajuato	Secuestro	Septiembre y noviembre	3
Campeche	'Otros delitos que atentan contra la libertad personal'	Octubre y noviembre	3
Colima	'Otros delitos que atentan contra la libertad personal'	Julio	1
Clúster 0 mujeres 2023			
México	Otros delitos que atentan contra la libertad personal	Enero	1863
Baja California	Otros delitos que atentan contra la libertad personal	Marzo, julio y noviembre	1671
Nuevo León	Otros delitos que atentan contra la libertad personal	Enero y agosto	1870
Aguascalientes	Corrupción de menores	Mayo	1
	Extorsión	Septiembre	1
	Trata de personas'	Mayo	1
Yucatán	Otros delitos contra la sociedad	Mayo y septiembre	2
Tlaxcala	Extorsión	Noviembre	1

Clúster 1 mujeres 2015			
México	Lesiones dolosas con arma blanca	Marzo y junio	33094
Baja California'	Lesiones dolosas no especificado	Mayo y octubre	7261
Nuevo León	Lesiones dolosas con otro elemento	Agosto y octubre	4987
Campeche	Lesiones dolosas con otro elemento	Octubre y diciembre	94
Colima	Lesiones dolosas con otro elemento	Junio y agosto	77
Aguascalientes	Homicidio culposo en accidente de trafico	Septiembre	4
Clúster 1 mujeres 2023			
México	Lesiones dolosas con otro elemento	Diciembre	21096
Guanajuato	Lesiones dolosas con otro elemento	Septiembre y noviembre	8100
Nuevo León	Lesiones dolosas con otro elemento	Junio	4230
Aguascalientes	Extorsión	Enero, febrero y noviembre	41
Yucatán	Otros delitos que atentan contra la vida y la integridad corporal	Mayo y septiembre	21
Tlaxcala	Lesiones dolosas con otro elemento	Abril, junio y diciembre	80

Tabla 14: Tabla de descripción de delitos, números de casos de k=2 del 2015 y 2023

En la tabla 15 se describen los promedios por entidad federativa para k= 3 en mujeres del 2015 y 2023.

Clúster 0				Clúster 1			
Mujeres 2015 = 0.70		Mujeres 2023 = 3.41		Mujeres 2015 = 3.63		Mujeres 2023= 11.42	
Estado	Promedio de delito	Estado	Promedio de delito	Estado	Promedio de delito	Estado	Promedio de delito
Tamaulipas	2.59	Ciudad de México	3.31	México	37.32	México	25.23

Tabasco	2.19	Baja California	3.01	Baja California	8.53	Guanajuato	10.71
Nuevo León	1.76	Veracruz	2.16	Nuevo León	7.98	Michoacán	6.81
Campeche	0	Durango	0	Campeche	0.15	Aguascalientes	0.125
Colima	0	Tlaxcala	0	Colima	0.11	Yucatán	0.08
Guanajuato	0	Yucatán	0	Aguascalientes	0.05	Tlaxcala	0.081

Clúster 2

Mujeres 2015 = 1.72		Mujeres 2023 = 1.34	
Estado	Promedio de delito	Estado	Promedio de delito
Nuevo León	6.85	México	14.65
Tamaulipas	5.20	Nuevo León	13.07
Puebla	4.68	Quintana Roo	6.83
Aguascalientes	0.063	Aguascalientes	0.20
Colima	0.03	Durango	0.17
Sonora	0	Sonora	0

Tabla 15: Promedios de los 3 estados de alta y baja incidencia delictiva por clúster en mujeres del 2015 y 2023 para k =3

En la tabla 16 se describen los delitos que se registraron en los estados de alta y baja incidencia delictiva de k = 3.

Clúster 0 mujeres 2015			
Estado	Delito	Meses	Casos
Tamaulipas	Lesiones dolosas no especificado	Julio y octubre	731

Tabasco	Lesiones dolosas no especificado	Mayo y octubre	614
Nuevo León	Otros delitos que atentan contra la vida y la integridad corporal	Enero y noviembre	507
Campeche			0
Colima			0
Guanajuato			0
Clúster 0 mujeres 2023			
Ciudad de México	'Otros delitos contra la sociedad'	Marzo y noviembre	255
Baja California	'Otros delitos contra la sociedad'	Julio, septiembre, octubre	403
Veracruz	'Otros delitos contra la sociedad'	Enero y marzo	776
Durango			0
Tlaxcala			0
Yucatán			0
Clúster 1 mujeres 2015			
México	Lesiones dolosas con otro elemento	Mayo y junio	33094
Baja California	Lesiones dolosas no especificado	Julio y octubre	7261
Nuevo León	Lesiones dolosas no especificado	Junio y octubre	4436
Campeche	Lesiones culposas en accidente de trafico	Agosto	94
Colima	Lesiones dolosas con otro elemento	Mayo y junio	94
Aguascalientes	Homicidios culposos en accidente de transito	Agosto	44
Clúster 1 mujeres 2023			
México	Lesiones dolosas con otro elemento	Diciembre	21096
Guanajuato	Lesiones dolosas con otro elemento	Marzo, junio y septiembre	8100
Michoacán	Lesiones dolosas con otro elemento	Mayo y junio	4727

Aguascalientes	Corrupción de menores	Abril, agosto y octubre	48
Yucatán	Lesiones dolosas con otro elemento	Enero	22
Tlaxcala	Lesiones dolosas con otro elemento	Febrero y abril	34
Clúster 2 mujeres 2015			
Nuevo León	Otros delitos que atentan contra la libertad personal	Julio y agosto	1042
Tamaulipas	Lesiones dolosas no especificado	Mayo y junio	1252
Puebla	Lesiones culposas en accidente de trafico	Febrero y marzo	619
Aguascalientes	Extorsión	Mayo	11
Colima	Extorsión	Julio	5
Sonora			0
Clúster 2 mujeres 2023			
México	'Otros delitos que atentan contra la libertad personal'	Enero	1765
Nuevo León	'Otros delitos que atentan contra la libertad personal'	Enero y mayo	1830
Quintana Roo	Lesiones dolosas con otro elemento	Marzo y junio	1480
Aguascalientes	Extorsión	Octubre	41
Durango	Extorsión	Enero y febrero	14
Sonora			

Tabla 16: Tabla de descripción de delitos, números de casos de k=3 del 2015 y 2023

En la tabla 17 se describen los promedios por entidad federativa para k= 4 en mujeres del 2015 y 2023.

Clúster 0				Clúster 1			
Mujeres 2015 = 0.54		Mujeres 2023 = 0.36		Mujeres 2015 = 3.80		Hombres 2023= 4.13	
Estado	Promedio de delito	Estado	Promedio de delito	Estado	Promedio de delito	Estado	Promedio de delito

Tamaulipas	2.43	Ciudad de México	2.28	México	38.63	México	25.54
Tabasco	2.10	Sonora	1.07	Baja California	9.14	Guanajuato	11.44
Sonora	1.5	Tamaulipas	0.88	Nuevo León	7.98	Jalisco	7.01
Coahuila	0	Tlaxcala	0	Campeche	0.15	Yucatán	0.083
Colima	0	Yucatán	0	Colima	0.13	Tlaxcala	0.082
Guanajuato	0	Zacatecas	0	Aguascalientes	0.05	Aguascalientes	0.08
Clúster 2				Clúster 3			
Mujeres 2015 = 1.42		Mujeres 2023 = 2.23		Hombres 2015 = 0.91		Mujeres 2023= 4.93	
Estado	Promedio de delito	Estado	Promedio de delito	Estado	Promedio de delito	Estado	Promedio de delito
Tamaulipas	5.06	México	6.38	Nuevo León	28.01	México	41.6
Tabasco	3.89	Quintana Roo	6.18	Puebla	8.63	Nuevo León	32.11
Morelos	3.68	Nuevo León	5.97	México	8.51	Baja California	21.88
Colima	0.03	Durango	0.07	Aguascalientes	0.03	Durango	0.34
Chihuahua	0.01	Coahuila	0.04	Colima	0.009	Tlaxcala	0
Sonora	0	Sonora	0	Quintana Roo	0	Yucatán	0

Tabla 17: Promedios de los 3 estados de alta y baja incidencia delictiva por clúster en mujeres del 2015 y 2023 para k =4

En la tabla 18 se describen los delitos que se registraron en los estados de alta y baja incidencia delictiva de k = 4 en mujeres

Clúster 0 Mujeres 2015			
Estado	Delito	Meses	Casos
Tamaulipas	Lesiones dolosas no especificado	Julio	731

Tabasco	Lesiones dolosas no especificado	Mayo a diciembre	614
Sonora	Lesiones dolosas no especificado	Mayo y junio	979
Coahuila			0
Colima			0
Guanajuato			0
Clúster 0 Mujeres 2023			
Ciudad de México	Otros delitos contra la sociedad	Mayo y septiembre	214
Sonora	Lesiones dolosas no especificado	Febrero y noviembre	473
Tamaulipas	Lesiones dolosas no especificado	Enero y octubre	231
Tlaxcala			0
Yucatán			0
Zacatecas			0
Clúster 1 Mujeres 2015			
México	Lesiones dolosas con otro elemento	Mayo y junio	33094
Baja California	Lesiones dolosas no especificado	Marzo y junio	7261
Nuevo León	Lesiones dolosas con otro elemento	Junio y octubre	4436
Campeche	Lesiones dolosas con otro elemento	Mayo y octubre	94
Colima	Lesiones dolosas con otro elemento	Mayo y octubre	77
Aguascalientes	Homicidio dolosas con otro elemento	Septiembre y noviembre	44
Clúster 1 Mujeres 2023			
México	Lesiones dolosas con otro elemento	Abril y junio	21096
Guanajuato	Lesiones dolosas con otro elemento	Mayo, junio y noviembre	8100
Jalisco	Lesiones dolosas con otro elemento	Marzo y junio	5601

Yucatán	Lesiones dolosas con otro elemento	Mayo y diciembre	22
Tlaxcala	Lesiones dolosas con otro elemento	Febrero y mayo	34
Aguascalientes	Homicidio culposo en accidente de transito	Mayo, septiembre y octubre	42
	Clúster 2 Mujeres 2015		
Tamaulipas	Lesiones dolosas no especificado	Mayo y octubre	1252
Tabasco	Lesiones dolosas no especificado	Octubre y noviembre	980
Morelos	Lesiones dolosas no especificado	Octubre y noviembre	1062
Colima	Extorsión	Febrero y septiembre	5
Chihuahua	Secuestro	Septiembre	2
Sonora			0
	Clúster 2 Mujeres 2023		
México	Extorsión	Marzo y julio	1102
Quintana Roo	Lesiones dolosas con otro elemento	Marzo y junio	1480
Nuevo León	Lesiones culposo con arma blanca	Junio y septiembre	1201
Durango	Extorsión	Enero y febrero	14
Coahuila	Secuestro	Junio	4
Sonora			0
	Clúster 3 Mujeres 2015		
Nuevo León	Otros delitos que atentan contra la libertad personal	Julio y noviembre	1549
Puebla	Otros delitos que atentan contra la libertad personal	Marzo, mayo y noviembre	509

México	Otros delitos que atentan contra la libertad personal	Marzo y octubre	510
Aguascalientes	Corrupción de menores	Agosto	4
Colima	Otros delitos que atentan contra la libertad personal	Julio	1
Quintana Roo			
	Clúster 3 Mujeres 2023		
México	Otros delitos que atentan contra la libertad personal	Enero	2496
Nuevo León	Otros delitos que atentan contra la libertad personal	Agosto	1870
Baja California	Otros delitos que atentan contra la libertad personal	Marzo y noviembre	2113
Durango	Otros delitos que atentan contra la libertad personal	Junio	37
Tlaxcala			0
Yucatán			0

Tabla 18: Tabla de descripción de delitos, números de casos de k=4 del 2015 y 2023 en mujeres

Discusión

Recapitulando los resultados en hombres de ambos años, se observó similitudes en los diferentes agrupamientos de cada uno de los puntos iniciales (valor k), demostrando que los estados con mayor incidencia delictiva son México y Nuevo león; siendo el delito más cometido contra ellos el Homicidio doloso con arma de fuego, ahora bien, los estados con menor incidencia es Aguascalientes y Tlaxcala con delitos registrados de Lesiones dolosas con otro elemento, secuestro y extorsión.

No obstante, el caso de las mujeres sucedió algo similar puesto que los estados de mayor incidencia son México y Tamaulipas; y los de menor incidencia Aguascalientes y Tlaxcala, a diferencia de los hombres en este caso los delitos en todos los casos fueron los mismos que son Lesiones dolosas con otro elemento y lesiones dolosas no especificados.

Aunque el agrupamiento es no supervisado y no se conoce de antemano los conjuntos de datos de entrenamiento, este crea lo grupos a las similitudes de los valores numéricos, en este caso, base a el entrenamiento y agrupación de los datos permitió visualizar resultados claros, cumpliendo

el objetivo de la investigación, se ha encontrado que en hombres como mujeres el estado de mayor peligro delictivo es México y el menor Aguascalientes, como así mismo el delito que más se comete por sexo. Demostrando que uno de los mayores problemas delictivos en la República Mexicana es el homicidio con arma de fuego y lesiones con elementos no especificados, esto siendo sustentado no solo en México si no el estudio que realizo (Valenga, *et.al,* 2008) para identificar patrones delictivos pero enfocado a homicidio doloso demuestra que es un problema no solo en la República Mexicana si no también en argentina.

Si bien, los resultados esclarecen en donde se debe prestar más atención para resolver o disminuir el problema de la delincuencia existen limitaciones puesto que los registros pueden ir variando con los años ya sea en aumento o disminución del mismos en los siguientes años. Por ende, se requiere seguir analizando dicho tema, con ayuda de la minería de datos y sus algoritmos que nos permite el manejo de grandes cantidades de datos.

Cierre

En la presente investigación se aplicó el algoritmo de K-medias a los datos de Incidencia delictiva provenientes de Secretariado Ejecutivo del Sistema Nacional de Seguridad Público de los años 2015 y 2023 en hombres y mujeres, teniendo como puntos iniciales k=2, k=3 y k=4, donde en los tres casos fueron satisfactorios y similares, llevando así un análisis más completo.

En base a los resultados de cada uno de ellos se pudo ver qué estado hay más incidencia delictiva como el delito que predomina en dicho estado en hombres como en mujeres, obteniendo que el delito que mayor se comete en hombres es el homicidio doloso con arma de fuego y el menor se comete contra ellos es el de lesiones, al contrario de las mujeres que el que mayor se comete contra ellas es el de lesiones dolosas no especificadas o con otro elemento y el que menor se comete es el homicidio.

Si bien en esta investigación se buscó describir los estados de mayor y menor incidencia, los delitos predominantes en dichos estados, también se pueden obtener el porcentaje de aumento de delitos por años del 2015 al 2023 o bien buscar predecir el aumento de la incidencia delictiva en los próximos años basándose en estos antecedentes.

Referencias

Carmona S. M. (1999). "Violencia y sociedad". *Adolescencia y Salud*, Vol.1 No.1, 1-10.

Causa en Común. (2021). Análisis de datos disponibles de la incidencia delictiva. https://causacomun.org.mx/beta/analisis-de-los-datos-posibles-de-incidencia-delictiva-septiembre-2021/

Corso, C., L.(s.f). Aplicación de algoritmo de clasificación supervisada usando Weka. Universidad Tecnológica Nacional, Facultad Regional Córdoba.

Garcia M. L. (2021).Factores criminógenos en jóvenes y su integración en la delincuencia organizada. Biolex, 13, 1-15. https://orcid.org/0000-0002-1428-234X

Gutiérrez G., I., Gutiérrez C. D., Juan R. J. E., Rodríguez P., L., Rico M., R., & Sánchez M. M. (2020). Aplicación del algoritmo K-means para el análisis de resultados de la prueba PLANEA 2017. Revista de Ciencias de la Computación, 149(8), 1-12.

Instituto Mexicano para la Competitividad A.C (IMCO). (2006). Situación de competitividad en México 2006: Punto de reflexión México.

Instituto Nacional de Estadística y Geografía (INEGI).(2023). Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública (ENVIPE) 2023 [Comunicado de prensa núm.546/23]. https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2023/ENVIPE/ENVIPE_23.pdf

Jiménez H. M. Á. (2021). El big data cómo herramienta de prevención de la delincuencia. Universidad de Alicante. Departamento de Comunicación y Psicología Social.

Marante J., D.(s.f). Aplicación de minería de datos para la exploración y detección de patrones delictivos (Trabajo de Diploma). Universidad de las Ciencias Informáticas, facultad 8.

Marrero, L., Carrizo, D., García S., L., & Ulloa V. F. (2021). Uso de al-goritmo K-means para clasificar perfiles de clientes con datos de medidores inteligentes de consumo eléctrico: Un caso de estudio, Ingeniare. Revista Chilena de Ingeniería, 29 (4), 1-10. http://dx.doi.org/10.4067/S0718-33052021000400778

Martín P., M., Sanjurjo de No, M. A., & Mira M. J. M.(s.f). Aplicación de K medias y SOM (Self Oganizing Maps) al análisis de micro accidentes de tráfico. Trabajo de fin de grado,

Departamento de ingeniería de Organización, Administración de Empresas y Estadística, Escuela Técnica superior de ingenieros industriales, universidad politécnica de Madrid.

Olina V. A. M., Espinoza M. M. A. (2022). Aplicación de minería de datos en datos abiertos de Ecuador: Delitos, Revista de Investigación y Cultura. Vol. 11. No. 1 https://doi.org/10.18050/RevUCVHACER,v11n1a8

Ortiz R. M. (2019). Estrategias de intervención para disminuir la violencia y la delincuencia: Avances y desafíos en República Dominicana, Ciencia y Sociedad, 44 (3), 51-67, Instituto Tecnológico de Santo Domingo.

Peña S. A. (2017). Modelo para la Caracterización del Delito en la Ciudad de Bogotá, Aplicando Técnicas de Minería de Datos Espaciales (Trabajo de Investigación). Universidad Distrital Francisco José De Caldas.

Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública. (s.f). Datos abiertos de incidencia delictiva. https://www.gob.mx/sesnsp/acciones-y-programas/datos-abiertos-de-incidencia-delictiva?state=published

Valenga, F., Fernández, E., Merlino, H., Rodríguez, D., Procopio, C., Britos, P., & García-Martínez, R. (2008). Minería de Datos Aplicada a la Detección de Patrones Delictivos en Argentina. En VII Jornadas de Ingeniería del Software e Ingeniería del Conocimiento.

Vilalta, C.J., Castillo, J.G., & Torres, J.A. (2016). Delitos violentos en las ciudades de América Latina. Banco Interamericano de Desarrollo. https://doi.org/10.18235/0007973

APLICACIÓN DE ALGORITMOS GENÉTICOS PARA OPTIMIZAR EL OTORGAMIENTO DE PRÉSTAMOS BANCARIOS

APPLICATION OF GENETIC ALGORITHMS TO OPTIMIZE THE GRANTING OF BANK LOANS

Karla Naomi Martínez Rojas[1]
kmartinezr011@alumno.uaemex.mx
ORCID: 0009-0008-3580-0645

Yaroslaf Aarón Albarrán Fernández[2]
yaalbarranf@uaemex.mx
ORCID: 0000-0002-2615-4740

Resumen

En este proyecto se utilizó una base de datos obtenida de la plataforma Kaggle**,** publicada por (Torres, 2020), la cual contiene información sobre préstamos bancarios en Alemania. Los datos fueron transformados a un formato binario y consisten en una muestra de 1000 elementos, también llamados cromosomas, cada uno con longitud de 20 *bits*, que representan aspectos clave en la toma de decisiones para la concesión de préstamos. Se dividieron en dos conjuntos, uno de entrenamiento y otro de prueba. Se le aplicó un Algoritmo Genético (selección, cruce y mutación) al conjunto de entrenamiento, obteniendo una lista de los mejores valores adaptados, y realizando una comparación con el grupo de prueba, obteniendo así a los individuos óptimos del segundo conjunto para la aprobación de un préstamo bancario.

Palabras clave: instituciones financieras, préstamos, algoritmos evolutivos, algoritmos genéticos, optimización.

[1] Ingeniería en Sistemas Inteligentes, Centro Universitario UAEM Nezahualcóyotl, Universidad Autónoma del Estado de México
[2] Ingeniería en Sistemas Inteligentes, Centro Universitario UAEM Nezahualcóyotl, Universidad Autónoma del Estado de México

Abstract

In this project, a database obtained from the Kaggle platform, published by (Torres, 2020), containing information on bank loans in Germany, was used. The data was transformed into a binary format and consists of a sample of 1000 elements, also called chromosomes, each with a length of 20 bits, which represent key aspects in decision-making for lending. They were divided into two sets, one for training and one for testing. A Genetic Algorithm (selection, crossbreeding and mutation) was applied to the training set, obtaining a list of the best adapted values, and making a comparison with the test group, thus obtaining the optimal individuals of the second set for the approval of a bank loan.

***Keywords*:** financial institutions, lending, evolutionary algorithms, genetic algorithms, optimization.

Fecha de envío: 05/06/2024
Fecha de aprobación: 10/11/2024
Fecha de publicación: 01/01/2025

Introducción

En el ámbito de las instituciones financieras, la toma de decisiones para otorgar préstamos o créditos, es un componente crítico que impacta directamente a la estabilidad y competitividad de las instituciones bancarias. Según (Cárdenas, 2004), un préstamo es una operación financiera en el cual interactúan dos actores económicos, el *prestamista*, que puede ser una persona, alguna empresa, o una institución bancaria que otorga los recursos económicos a la contraparte, el *prestatario*, quien recibe el dinero con la obligación de devolverlo, usualmente con un porcentaje extra conocido como *interés*. Este interés puede variar en función de las políticas específicas de una institución financiera, o bien de las preferencias individuales de una persona, al término de un lapso predefinido.

Este tipo de préstamos y créditos, representan una forma común de financiamiento para individuos, empresas y organizaciones con el propósito de obtener a cabo diversas actividades, como adquirir una vivienda, expandir un negocio o financiar proyectos. Como menciona (Méndez,

2020), los préstamos bancarios son una herramienta crucial para impulsar la actividad económica y su crecimiento, así como la realización de proyectos personales o empresariales. Es fundamental comprender los términos y condiciones del préstamo y garantizar que el plan de pagos sea viable para evitar dificultades financieras en el futuro. Por lo tanto, la toma de decisiones de los bancos, se basa en emplear diversas técnicas y criterios adicionales para determinar la elegibilidad del solicitante.

Es crucial tener en cuenta que los criterios que pueden diferir entre instituciones, dependen del tipo específico de préstamo o crédito solicitado. Además, la tecnología, como modelos de *scoring* crediticio y algoritmos, está siendo cada vez más utilizada para evaluar la solvencia del solicitante, dando paso a la aplicación de técnicas dentro del área de la Inteligencia Artificial (IA). (The BlackBox Lab, 2020), plantea que un ejemplo serían los Algoritmos Evolutivos (AE) que son métodos de optimización inspirados en el proceso de evolución natural. Estos algoritmos son utilizados para encontrar soluciones aproximadas o incluso óptimas a problemas complejos, en los que, al explorar las posibles soluciones de manera exhaustiva, resulta impracticable.

La idea central detrás de estos Algoritmos, es simular la combinación de material genético, aplicando operaciones genéticas como la selección, la recombinación (también llamada cruce) y la mutación a una población de soluciones potenciales. Los Algoritmos Genéticos (AG) son especialmente útiles en problemas de optimización, dentro de los cuales tenemos a los Algoritmos Genéticos (AG) una técnica de búsqueda y optimización que se basa en principios tomados de la genética y la evolución natural, una variante especifica de los AG. Como señala (Conectando ideas, 2023), la fortaleza de los algoritmos genéticos radica en su capacidad para explorar soluciones de manera simultánea y abordar espacios de búsqueda extensos y complicados.

Aunque la inteligencia artificial ya ha permitido a las entidades financieras realizar inferencias sobre los perfiles de los solicitantes para obtener préstamos bancarios de manera automática. Con la aplicación de los Algoritmos Genéticos se lograría optimizar la selección de préstamos, ya que tienen la capacidad de adaptarse y evolucionar a lo largo del tiempo, buscando así soluciones cada vez más eficientes respecto al cambio. Como menciona (Pompas, 2024), fortalecer la capacidad predictiva y adaptativa de las instituciones en un mercado en constante cambio, reducir riesgos y, como última instancia, ofrecer una experiencia financiera más personalizada a los clientes.

Antecedentes

Dentro del ámbito financiero, se han encontrados algunos trabajos que aplican Algoritmos genéticos en los que la importancia de estos radica en abordar la gestión de riesgos y problemas de optimización como es el trabajo de (Hernández, 2022), donde la optimización de carteras, ha sido uno de los campos más beneficiados por estas tecnologías, ya que busca maximizar sus rendimientos mientras minimizan los riesgos que esto conlleva, sin embargo, la estabilidad financiera vista en el trabajo de (Metawua et al., 2016) muestra una perspectiva donde la instituciones financieras enfrentan el desafío de distribuir de manera equitativa, gestionando las restricciones crediticias de manera efectiva, ayudando a mitigar el impacto de posibles crisis crediticias; asimismo, hablando de prevenir la mala toma de decisiones (Ramos, 2017), aborda en su trabajo que la predicción precisa del riesgo crediticio es esencial para la estabilidad financiera, mediante el uso de modelos de *rating* crediticio y análisis avanzados, estas herramientas permiten a las instituciones anticipar y abordar los riesgos potenciales antes de que estas se conviertan en crisis económicas, asimismo (Eraso y Suarez, 2018), aplica los Algoritmos Genéticos enfocados a los mercados capitales, donde la selección de portafolios de inversión desempeña un papel crucial en la maximización de rendimientos y la mitigación de los riesgos, dando por concluido que la innovación de tecnología puede abordar una variedad de sectores, ayudando a la toma de decisiones estratégicas fundamentadas en datos, teniendo resultados favorables en el áreas de finanzas como también lo es este trabajo, dejando a la espera de cómo se desenvolverá ahora en esta problemática del otorgamiento crediticio.

Materiales y Métodos

Para el desarrollo de este proyecto, se utilizó una base de datos obtenida del repositorio de Kaggle por (Torres, 2020). Esta base de datos está bajo licencia de código abierto Apache 2.0. Es una licencia de software libre permisiva que permite a los usuarios utilizar, copiar, modificar y redistribuir el software sin pagar derechos de autor o licencia.

La base de datos consta de 1000 registros, de los cuales cada uno está conformado por 21 características, explicadas a continuación:

1. Checking_status: Nos menciona el estado de cuenta de un cliente, la cual puede tener estos cuatro valores
    - "<0": Valores negativos
    - "0<x<200": Valores entre 0 y 200
    - "No checking": No dispone de una cuenta
    - ">=200": Valores de más de 200
2. Duration: Duración en meses del prestamos
3. Credit_history: Indica si el cliente tiene o tuvo créditos y su estado.
    - "critical/other existing credit": Estado crítico o existe otro crédito existente
    - "existing paid": Tiene créditos, pero han sido correctamente pagados al momento
    - "delayed previously": Tuvo retraso en alguno de sus pagos anteriores
    - "no credits/all paid": No tiene historial crediticio o todos los créditos fueron para gados correctamente
    - "all paid": Todos los créditos fueron pagados correctamente
4. Purpose: Indica el motivo para la solicitud
    - "radio/tv": Para un radio o televisión.
    - "education": Para educación.
    - "furniture/equipement": Para muebles o equipamiento.
    - "new car": Para un coche nuevo.
    - "used car": Para un coche 2ª mano.
    - "bussiness": Para trabajo.
    - "domestic appliance": Para un electrodoméstico.
    - "repairs": Para algunas reparaciones.
    - "other": Otros.
    - "retraining": Para adquirir nuevas habilidades.
5. Credit_amount: Indica la cantidad de dinero en la cuenta
6. Saving_status: Indica los ahorros del cliente y el monto.
    - "no known savings": No tiene
    - "<100": Tiene menos de 100
    - "100<=x<500": Tiene entre 100 y 500
    - "500<=x<1000": Tiene entre 500 y 1000

- “>=1000”: Tiene más de 1000

7. Employment: Tiempo de permanencia en su empleo

- “>=7”: Lleva trabajando como mínimo 7 años en la empresa actual.
- “1<=x<4”: Lleva en la empresa entre 1 y 4 años.
- “4<x<=7”: Lleva en la empresa entre 4 y 7 años.
- “unemployed”: No tiene trabajo al momento.
- “<1”: Lleva menos de 1 año

8. Installment_commitment: Nos indica los plazos del préstamo

9. Personal_status: Sexo y estado civil

- “male single”: Hombre soltero.
- “female div/sep/mar”: Mujer divorciada, casada o que depende de alguien.
- “male div/sep”: Hombre divorciado o separado.
- “male mar/wid”: El cliente es un hombre casado o viudo.

10. Other_parties: Indicador de otras personas involucradas

- “none”: No hay partes involucradas
- “guarantor”: El cliente cuenta con un aval
- “co applicant”: El cliente cuenta con co-solicitante, comparten responsabilidad.

11. Residence_since: Tiempo de vivir en su residencia

12. Property_magnitude: Tipo de propiedades

- “real_estate”: Se posee propiedades inmobiliarias como casas, apartamentos o terrenos.
- “life insurance”: Cuenta con un seguro de vida.
- “no know property”: Se desconoce si el cliente posee algún tipo de propiedad a su nombre.
- “car”: Cuenta con un coche.

13. Age: Edad del solicitante

14. Other_payment_plans: Plan de pago adicional

- “none”: No dispone de ningún pago adicional aparte del pago estándar del préstamo.
- “store”: Tiene uno ofrecido por una tienda o comercio específico.

- “bank”: Tiene uno ofrecido por el banco u otra institución financiera que otorgó el préstamo inicial.

15. Housing: Situación de la vivienda

- “rent”: Reside en una vivienda alquilada.
- “own”: Reside en una vivienda propia
- “for free”: Vivienda que no está sujeta a pagos de alquiler o de hipoteca.

16. Existing_credits: Numero de créditos abiertos al momento

17. Job: Situación laboral del solicitante

- “skilled”: Tiene un trabajo nivel avanzado de conocimiento
- “unskilled resident”: Tiene un trabajo nivel medio o bajo de conocimiento
- “high qualified/self emp/mgmt”: Tiene un trabajo de alto nivel de conocimiento, es autónomo o trabaja en un puesto de gestión.
- “unemp/ unskilled non res”: Está desempleado y además no tiene habilidades especializadas.

18. Num_dependents: Número de personas que dependen del solicitante económicamente.

19. Own_telephone: Posee o no teléfono móvil propio.

20. Foreing_worker: Su área de trabajo es foráneo o no.

21. Class: Indicador de si el cliente es apto o no para recibir el préstamo

Como se observa los valores de cada una de las variables son diferentes entre sí, por lo que se realizó un proceso de transformación y normalización de los datos, con un código en Python, el cual ayudó a cambiar las variables a binario, ver Tabla 1.

Variables	**Valores originales**	Variables modificadas
Checking_status:	“<0” y “no checking”	**0**
	“0<=x<200” y “>=200”	**1**
Duration:	duration <= 12 meses	**0**
	duration > 12 meses	**1**
Credit_history:	“critical/other existing credit” y “delayed previously”	**0**

	“existing paid”, “no credits/all paid” y “all paid”	**1**
Purpose:	“education”, “business” y “retraining”	**1**
	“radio/tv”, “furniture/equipement”, “new car”, “used car”, “domestic appliance”, “repairs” y “other”	**0**
Credit_amount:	credit_amount <= 5000	**0**
	credit_amount > 5000	**1**
Saving_status:	“no known savings” y“<100”	**0**
	“100<=x<500”, “500<=x<1000” y“>=1000”	**1**
Employment:	“<1”, “1<=x<4”, y “unemployed”	**0**
	“4<x<=7” y “>=7”	**1**
Installment_commitment:	installment_commitment <= 2	**0**
	installment_commitment > 2	**1**
Personal_status:	male single, male div/sep	**0**
	female div/sep/mar, male mar/wid	**1**
Other_parties:	“none”	**0**
	“guarantor”, y “co applicant”	**1**
Residence_since:	residence_since <= 2 años	**0**
	residence_since > 2 años	**1**
Property_magnitude:	“no known property”	**0**
	“real_estate”, “life insurance” y “car”	**1**
Age:	age <= 35	**0**
	age > 35	**1**
Other_payment_plans:	“none”	**0**
	“store”, y “bank”	**1**
Housing:	“rent”, y “for free”	**0**

	"own"	**1**
Existing_credits:	existing_credits <= 1	**0**
	existing_credits > 1	**1**
Job:	"unemp"/ "unskilled non res"	**0**
	"skilled", "unskilled resident", "high qualified/self emp/mgmt"	**1**
Num_dependents:	num_dependents == 0	**0**
	num_dependents > 0	**1**
Own_telephone:	"yes"	**1**
	"no"	**0**
Foreign_worker:	"yes"	**0**
	"no"	**1**
Class:	"good"	**1**
	"bad"	**0**

***Tabla 1**. Variables originales y modificadas*

Cada individuo contiene los 21 aspectos ya mencionados, sin embargo, como la última columna es la clase a la que pertenece, ese no se ocupa en el algoritmo, por lo tanto, se manejó un cromosoma de tamaño de 20 bits, respetando las posiciones de las variables como se muestra en la Tabla 2.

0	**1**	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**1**	**0**
Checking_status	**Duration**	**Credit_history**	**Purpose**	**Credit_amount**	**Saving_status**	**Employment**	**Installment_commitmen**	**Personal_status**	**Other_parties**	**Residence_since**	**Property_magnitude**	**Age**	**Other_payment_plans**	**Housing**	**Existing_credits**	**Job**	**Num_dependents**	**Own_telephone**	**Foreign_worker**

***Tabla 2**. Ejemplo de cromosoma*

Desarrollo

El proceso se dividió en las siguientes etapas:

I. Preparación y Carga de datos

Se modificó la base de datos (BD) original para binarizar los valores de las variables que componen a los individuos como se muestra en la Ilustración 1 y en la Ilustración 2 .

	A	B	C	D	E	F	G
1	checking_status	duration	credit_history	purpose	credit_amount	savings_status	employment
2	'<0'	6	'critical/other existing credit'	radio/tv	1169	'no known savings'	'>=7'
3	'0<=X<200'	48	'existing paid'	radio/tv	5951	'<100'	'1<=X<4'
4	'no checking'	12	'critical/other existing credit'	education	2096	'<100'	'4<=X<7'
5	'<0'	42	'existing paid'	furniture/equipment	7882	'<100'	'4<=X<7'
6	'<0'	24	'delayed previously'	'new car'	4870	'<100'	'1<=X<4'
7	'no checking'	36	'existing paid'	education	9055	'no known savings'	'1<=X<4'

***Ilustración 1**. BD Original.*

	A	B	C	D	E	F	G
1	checking_status	duration	credit_history	purpose	credit_amount	savings_status	employment
2	1	0	0	0	0	0	0
3	1	1	0	0	1	0	0
4	1	0	0	1	0	0	0
5	1	1	0	0	1	0	0
6	1	1	0	0	0	0	0
7	1	1	0	1	1	0	0

***Ilustración 2**. BD Binarizada*

- La base de datos binarizada se dividió en dos conjuntos: uno de entrenamiento y otro de prueba.
- Los dos conjuntos de prueba se cargaron en formato csv separando del primer conjunto las variables (x), menos la última columna (y) que es la variable objetivo.
- Se utilizó la herramienta 'SimpleImputer' para llenar los valores faltantes con la palabra 'NaN', lo que facilito la detección de algún problema en el algoritmo genético.
- Se normalizaron los datos utilizando 'MinMaxScaler' para escalar las características

II. Modelo de regresión lineal

- Se entrenó un modelo de regresión, utilizando los datos normalizados del paso anterior.
- Se obtuvieron los coeficientes, es decir, los pesos en base a la importancia de cada uno de los atributos.

III. Función de adaptación del algoritmo genético

Los pesos generados, se utilizaron para la función de adaptación del algoritmo genético. De tal modo que la función de adaptación se ve de la siguiente manera:

$$Valor_{adaptado} = \sum_{i=1}^{n} w_i * Valor$$

Donde:

$Valor_{adaptado}$ es el valor total de adaptación del individuo.

n es el numero de atributos que componen al individuo.

w_i es el valor ponderado asignado a el atributo i.

$Valor_i$ es el valor real

IV. Aplicación del Algoritmo genético

Se aplicó la Selección Torneo, Cruce De Dos Puntos Y Mutación Simple.

- ***Selección:***
  Se eligen a los individuos para la nueva población, generando una nueva lista del mismo tamaño de la población inicial.

- ***Cruce:***
  Usando la población final de la selección, opta por tomar dos individuos de esa lista y cruzar el material genético, dando dos hijos y seleccionando al más apto, para irse a una nueva población del mismo tamaño.

- ***Mutación:***
  Seguido del cruce, algunos o todos los individuos pueden mutar dependiendo la probabilidad de mutación, obteniendo otra nueva lista del mismo tamaño de la población inicial.

  La condición de paro de este código fue cumplir con un porcentaje especifico del 75% la población adaptada, alcanzando un umbral de adaptación, el cual se basó en el

percentil 90 de las adaptaciones obtenidas, dando al final la lista de los mejores valores adaptados.

V. Comparación final y aplicación

Después de la aplicación del algoritmo genético, se ocupa la base de datos de prueba, la cual se compara con la lista obtenida en el paso 4, para finalmente mostrar cuales fueron los elementos iguales de la base de prueba encontrados.

Resultados

En la salida podemos observar el número de generaciones en las que el algoritmo deja de iterar debido a que llegó o sobre pasó el porcentaje de individuos adaptados, ver Ilustración 3, así como los valores adaptados de la población inicial.

```
Generación 1: 10.220440881763528% de la población ha alcanzado el umbral de adaptación.
Generación 2: 28.45691382765531% de la población ha alcanzado el umbral de adaptación.
Generación 3: 53.10621242849694% de la población ha alcanzado el umbral de adaptación.
Generación 4: 77.35470941883767% de la población ha alcanzado el umbral de adaptación.

     Resultado_Real  Adaptacion
0                 1    0.071886
1                 0    0.107906
2                 1    0.169935
3                 1    0.253424
4                 0    0.043977
..              ...         ...
494               1   -0.054169
495               0    0.082263
496               0    0.051618
497               1    0.161241
498               1    0.065712

[499 rows x 2 columns]
```

***Ilustración 3**. Salida*

Seguido de una lista de los mejores individuos adaptados, en su forma de cromosoma dándonos una población final, del mismo tamaño que la población inicial de 499 elementos, ver Ilustración 4.

```
Mejores Individuos en Forma Binaria:
Individuo 1: 0000000010001010101011
Individuo 2: 0100000001101001101001
Individuo 3: 0010000001010011101011
Individuo 4: 0100000001001011101000
Individuo 5: 0001000101001001001010
Individuo 6: 0000000000001001110 1
Individuo 7: 0000000001001000110 1
Individuo 8: 0100000001101000000 0
Individuo 9: 0000000100111010101 0
Individuo 10: 0000000000110010101 1
:
Individuo 498: 0000000001110010101 1
Individuo 491: 1001000101101011110 0
Individuo 492: 0000000101101011010 0
Individuo 493: 0010000001001111100 0
Individuo 494: 1000010001001010100 0
Individuo 495: 0100011100111010000 0
Individuo 496: 0100000101101111001 0
Individuo 497: 0000000000011001111 1
Individuo 498: 0000000000011001111 0
Individuo 499: 0100000101001110100 0
```

***Ilustración 4**. Salida dos*

Finalmente se obtuvo una lista de los elementos del segundo conjunto que eran iguales y significaban un préstamo/crédito aceptado, sin embargo, se puede observar que, a pesar de que los valores coincidían con los valores mejor aceptados al 75%, al ver su clasificación original de ese conjunto, resultó tener cuatro individuos con préstamo/crédito rechazado, que el algoritmo identificó como un aceptado, ver Ilustración 5.

	class
17	1
40	1
91	1
118	1
120	1
127	1
147	1
152	0
163	1
181	1
194	1
248	0
281	0
372	1
400	0
407	1
442	1
458	1
477	1

***Ilustración 5**. Salida tres*

En cambio, si modificamos alguna de las variables, como el porcentaje de valores adaptados al 85%, observamos que este se detiene hasta la generación cinco, ver Ilustración 6 , y en cambio, en

la salida final de los mejores elementos del conjunto de prueba, obtenemos un menor número de elementos seleccionados y solo dos falsos verdaderos, ver Ilustración 7.

```
Generación 1: 10.220440881763528% de la población ha alcanzado el umbral de adaptación.
Generación 2: 26.252505010020037% de la población ha alcanzado el umbral de adaptación.
Generación 3: 46.292585170340686% de la población ha alcanzado el umbral de adaptación.
Generación 4: 72.74549098196393% de la población ha alcanzado el umbral de adaptación.
Generación 5: 86.57314629258516% de la población ha alcanzado el umbral de adaptación.
```

***Ilustración 6**. Segunda Salida*

	Class
0	1
11	0
16	1
101	1
102	0
105	1
268	1
296	1
360	1
372	1
434	1
461	1
477	1

***Ilustración 7**. Segunda salida tres*

Discusión

Mediante este proyecto, exploramos la aplicación de un algoritmo genético enfocado en el área financiera, con el objetivo de reducir tiempos en la toma de decisiones. Esto se logró mediante la combinación del análisis de los datos, la aplicación de la regresión lineal, que ayudó en la asignación de pesos ajustados donde los valores fueron asignados en función de la influencia que tuvieran en el resultado final, y finalmente, la aplicación del algoritmo genético que permitió obtener valores aleatorios, basados en un conjunto de datos previo, y lograr hacer pruebas para la identificación de un buen candidato.

Los resultados obtenidos mostraron incidencias en las que el algoritmo identificó satisfactoriamente un candidato que, en la base de datos original, había sido aprobado. Esto podría ser benéfico para los solicitantes antes de ir a una valoración, ya que al aplicar este tipo de programas, podrían tener una vista preliminar de lo que podrían obtener por respuesta de un banco en cuestión de segundos, siempre y cuando, la base de datos se alimentara de forma constante, y

con más datos históricos que posibilitaran ver el conjunto de atributos mínimos esenciales para otorgar los préstamos.

Además de destacar el margen de error, relativamente bajo, que se obtuvo en la comparación de resultados, lo que podría identificarse como una herramienta eficaz en la toma de decisiones crediticia. Sin embargo, también se pudo observar, tras varias iteraciones, que favorecía a individuos con ciertos valores, lo que podría sesgar la toma de decisiones si no se sabe manejar adecuadamente, por lo que nos lleva a un problema de uso que debería ser supervisado para evitar que se haga mal uso de la herramienta y pueda garantizar transparencia y equidad en los resultados.

Aunque en esta ocasión, las pruebas fueron realizadas con una base de datos aparentemente real, también se sabe que los datos pudieron ser modificados para su uso académico como lo es este caso, y esto podría conllevar más trabajo la transformación adecuada para el uso de datos obtenidos directamente de alguna institución bancaria.

Cierre

La aplicación del método de regresión lineal fue una ventaja, dado que los valores ponderados usados en la función de adaptación no fueron fijos, sino que fueron adaptados específicamente respecto a la importancia de cada variable, así como la ejecución del algoritmo genético que permitió obtener las mejores combinaciones en base a datos de entrenamiento, y así realizar una búsqueda en los datos de prueba, donde se observaron resultados satisfactorios al obtener un margen de error pequeño respecto a su clasificación inicial y la selección del algoritmo.

Los resultados obtenidos, mostraron ser eficaces en la identificación de candidatos aptos para otorgar préstamos/créditos. Sin embargo, se observan limitantes como el riesgo de sesgo en algunos patrones de datos, lo que afectaría los resultados. De este modo, sería necesaria la supervisión de personal capacitado para garantizar el uso transparente y adecuado de estas tecnologías.

Referencias

Cárdenas, G. A. (2004). *Diccionario de Ciencias-Administrativas.* www.cucea.udg.mx: https://www.cucea.udg.mx/include/publicaciones_drupal/pdfs/dic_ecoadm.pdf

Conectando ideas. (27 de Junio de 2023). *Algoritmos genéticos en inteligencia artificial: La clave para mejorar la toma de decisiones*. conectandoideas.net: https://conectandoideas.net/algoritmos-geneticos-en-inteligencia-artificial/#:~:text=La%20clave%20de%20los%20algoritmos%20gen%C3%A9ticos%20en

Eraso, L. M., y Suarez, N. (2018). *Configuración de Algortimos Genéticos para la selección de Portafolios de Inversión en el Mercado de Capitales Colombiano.* core.ac.uk : https://core.ac.uk/download/pdf/161642593.pdf

Hernández, S. (Junio de 2022). *Aplicación de algoritmos genéticos a la optimizaci+on de carteras.* oa.upm.es: https://oa.upm.es/70927/1/TFG_SOFIA_HERNANDEZ_MONTERO.pdf

Mendez, G. (22 de Septiembre de 2020). *Créditos bancarios: Un circulo virtuoso para la economía* . Deloitte Spanish Latin America: https://www2.deloitte.com/mx/es/pages/dnoticias/articles/creditos-bancarios-circulo-virtuoso-para-economia.html

Metawua, N., Elhoseny, M., Hassan, M., y Hassanien, A. E. (2016). *Loan Portfolio Optimization using Genetic Algorithm: A case of credit constraints.* ResearchGate: https://www.researchgate.net/publication/313804105_Loan_portfolio_optimization_using_Genetic_Algorithm_A_case_of_credit_constraints

Pompas, J. (23 de Enero de 2024). *Algoritmos Genéticos: La Evolución Computacional*. Linked in: https://es.linkedin.com/pulse/algoritmos-gen%C3%A9ticos-la-evoluci%C3%B3n-computacional-pompas-guti%C3%A9rrez-yvqof

Ramos, H. M. (2017). *Implementación de una herramienta de análisis de riesgo de crédito basado en el modelo de rating de crédito, algoritmos genéticos y clustering jerárquico aglomerativo.* core.ac.uk: https://core.ac.uk/download/pdf/323349107.pdf

The BlackBox Lab. (22 de Junio de 2020). *Qué son los algoritmos evolutivos y para qué se usan*. theblackboxlab.com: https://theblackboxlab.com/2020/06/22/que-son-los-algoritmos-evolutivos-y-para-que-se-

usan/#:~:text=Qu%C3%A9%20son%20los%20algoritmos%20evolutivos.%20La%20defini

Torres, O. M. (14 de Septiembre de 2020). *Análisis datos bancarios. Pre machine Learning*. Kaggle.com: https://www.kaggle.com/code/oscarmatiastorres/an-lisis-datos-bancarios-pre-machine-learning

ANÁLISIS DE LAS PROBLEMÁTICAS DE UNA EMPRESA DE REPARTO DE PUERTA A PUERTA

ANALYSIS OF THE PROBLEMS OF A DOOR-TO-DOOR DELIVERY COMPANY

Jorge Arturo García Hernández[1]
r2d2.hdz.040200@gmail.com
ORCID: 0009-0003-7163-9274

Juan Antonio Jiménez García[2]
jajimenezg@uaemex.mx
ORCID: 0000-0002-7303-1284

Víctor Manuel Durán López[3]
vmduranl@uaemex.mx
ORCID: 0000-0002-7303-1284

Rigoberto Torres Tovar[4]
rtorrest@uaemex.mx
ORCID: 0000-0001-8734-3839

Resumen

Analizar las problemáticas de una empresa de reparto de puerta a puerta de manera automática mediante *software,* generalmente implica el uso de sistemas de gestión empresarial (ERP), *software* de seguimiento y análisis de datos, así como herramientas de inteligencia empresarial.

[1] Estudiante de la Licenciatura en Ingeniería en Transporte, Centro Universitario de Nezahualcóyotl, Universidad Autónoma del Estado de México.
[2] Profesor de Tiempo Completo, Centro Universitario de Nezahualcóyotl, Universidad Autónoma del Estado de México.
[3] Técnico Académico de Tiempo Completo, Centro Universitario Nezahualcóyotl. Universidad Autónoma del Estado de México.
[4] Profesor de Tiempo Completo, Centro Universitario de Nezahualcóyotl, Universidad Autónoma del Estado de México

Los principales aspectos son: Integración de datos, Análisis de datos en tiempo real, Aplicación de algoritmos de aprendizaje automático, Generación de informes automatizados, Reportes (Dashboards) interactivos, Alertas y notificaciones automáticas, Iteración y mejora continua. En esta investigación se presentan ejemplos de cómo se lleva a cabo el análisis de problemáticas con una empresa de reparto puerta a puerta para identificar áreas de mejora y tomar medidas correctivas de manera eficiente y proactiva.

Palabras claves: Gestión empresarial, ERP, Empresa de reparto, Puerta a puerta, toda de decisiones.

Abstract

Analyzing the problems of a door-to-door delivery company automatically using software generally involves the use of business management systems (ERP), data tracking and analysis software, as well as business intelligence tools. The main aspects are: Data integration, Real-time data analysis, Application of machine learning algorithms, Automated report generation, Interactive dashboards, Automatic alerts and notifications, Iteration and continuous improvement. This research presents examples of how problem analysis is carried out with a door-to-door delivery company to identify areas of improvement and take corrective action in an efficient and proactive manner.

Keywords: Business management, ERP, Delivery company, Door to door, all decision making.

Fecha de envío: 07/06/2024
Fecha de aprobación: 10/11/2024
Fecha de publicación: 01/01/2025

Introducción

Un Centro de Distribución (CEDI) de productos terminados, es una instalación logística dedicada al almacenamiento, manejo y distribución de productos que han completado su proceso de fabricación y están listos para ser entregados a los clientes finales o a puntos de venta. Estos centros juegan un papel crucial en la cadena de suministro, asegurando que los productos se distribuyan

de manera eficiente y oportuna. Las principales operaciones que se llevan a cabo en un centro CEDI, se enlistan en la tabla 1.0

Almacenamiento	Rutas de Distribución
Manejo de Inventario	Gestión de Flota
Recepción y Despacho	Automatización

Tabla 1.0 Principales operaciones en un CEDI.

Estructura de una Empresa de Reparto Almacén de Productos Terminados *(Last Mile)*

La estructura de una empresa de reparto debe tomar en consideración varias características y áreas de oportunidad para garantizar su éxito a largo plazo. Estas características se enlistan en la tabla 2.0 (Gómez, A., & Prieto, I., 2017).

Ubicación Estratégica
Capacidad de Almacenamiento (Monczka, R. M., Handfield, R. B., Giunipero, L. C., & Patterson, J. L.2010).
Tecnología y Sistemas de Gestión. Como (WMS *Warehouse Management System) o Tecnología RFID (Identificación por Radiofrecuencia) (Martínez, E., 2016)*
Recepción y Almacenamiento
Gestión de Inventarios. Técnicas de inventario como FIFO (*First In, First Out*) o LIFO (*Last In, First Out*)
Preparación de Pedidos, con sistemas de *picking* eficientes (manuales, semi-automatizados o totalmente automatizados).
Distribución y Envío
Optimización del Espacio (González, F., & Fernández, M. (2018)
Automatización y Tecnología con Sistemas de gestión de almacenes (WMS), robótica y automatización (Jiménez, P. M., & Gómez, C. S., 2013).
Capacitación del Personal
Seguridad y Cumplimiento

Gestión de la Demanda y la Planificación

Tabla 2.0 Características de una empresa de reparto y áreas de oportunidad o mejora

Todas las características y áreas de mejora en las empresas de reparto de puerta a puerta se integran en un único sistema de gestión empresarial (ERP) que les permite a todos los miembros de la organización, el acceso a la información en tiempo real, con propiedades de usuarios para manejar, modificar y agregar información al sistema.

Sistemas de Gestión Empresarial

Los sistemas de gestión empresarial (también conocidos como ERP, por sus siglas en inglés: *Enterprise Resource Planning*) son *softwares* integrados que las empresas utilizan para administrar sus operaciones diarias (Coronel, C., & Morris, S, 2018**)**. Estos sistemas abarcan varias áreas funcionales de una organización, como finanzas, recursos humanos, inventario, ventas, compras, producción, entre otras. Algunos de los principales objetivos de los sistemas de gestión empresarial se muestran en la tabla 3.0 (Fernández, M. A., 2015).

Integración: Permiten la integración de datos y procesos de diferentes departamentos de una empresa en un solo sistema, lo que facilita la comunicación y la colaboración entre equipos.
Automatización: Ayudan a automatizar tareas repetitivas y procesos comerciales, lo que puede aumentar la eficiencia y reducir los errores
Centralización de la información: Almacenar todos los datos en una base de datos centralizada facilita el acceso a la información en tiempo real para la toma de decisiones.
Mejora de la visibilidad: Proporcionan una visión más completa y detallada de las operaciones comerciales, lo que permite a los gerentes tomar decisiones más informadas.
Estándares de negocio: A menudo, los sistemas de gestión empresarial están basados en las mejores prácticas y estándares de la industria, lo que puede ayudar a las empresas a alinear sus procesos con las normas aceptadas.

Tabla 3.0 Objetivos de un sistema de gestión empresarial

Algunos ejemplos de sistemas de gestión empresarial populares incluyen SAP, Oracle ERP, Microsoft Dynamics, NetSuite, entre otros. La elección del sistema adecuado depende de las necesidades específicas de cada empresa, su tamaño, su industria y su presupuesto.

Principales elementos de un sistema de gestión empresarial

Los principales elementos de un sistema de gestión empresarial (ERP) se enlistan en la tabla 4.0 (Muñoz *et. al.*, 2015); (Alonso, F., 2017).

Módulos funcionales: Estos son los componentes principales del ERP y cubren áreas específicas de la empresa, como finanzas, recursos humanos, inventario, compras, ventas, producción, gestión de proyectos, CRM (*Customer Relationship Management*), entre otros. Cada módulo está diseñado para automatizar y gestionar procesos relacionados con su área correspondiente.
Base de datos centralizada: Un componente esencial de un ERP es una base de datos centralizada que almacena todos los datos de la empresa. Esta base de datos permite la integración y el intercambio de información entre los diferentes módulos del sistema.
Interfaz de usuario: Los sistemas de gestión empresarial suelen tener una interfaz de usuario intuitiva que permite a los empleados acceder y utilizar fácilmente las funciones del sistema. Esto incluye la capacidad de ingresar datos, ejecutar informes, realizar seguimiento de actividades y realizar otras tareas relevantes.
Motor de procesos empresariales: Los ERP utilizan un motor de procesos empresariales para automatizar y gestionar los flujos de trabajo dentro de la organización. Este motor permite definir y ejecutar procesos comerciales específicos, lo que ayuda a mejorar la eficiencia y la consistencia de las operaciones.
Funcionalidades de análisis y reporte: Los sistemas de gestión empresarial suelen incluir herramientas de análisis y reporte que permiten a los usuarios generar informes personalizados, analizar datos y obtener información valiosa sobre el rendimiento de la empresa. Esto facilita la toma de decisiones informadas y estratégicas.

Seguridad y control de acceso: La seguridad de los datos es fundamental en un ERP. Estos sistemas suelen contar con funciones de seguridad avanzadas que protegen los datos confidenciales y controlan el acceso de los usuarios a diferentes partes del sistema.
Integración con sistemas externos: Los ERP suelen integrarse con otros sistemas externos, como sistemas de contabilidad, sistemas de gestión de clientes, sistemas de gestión de almacenes, entre otros. Esta integración permite una mayor eficiencia y coherencia en toda la empresa.

Tabla 4.0 Principales elementos de un sistema de Reparto

Estos son los principales elementos que componen un sistema de gestión empresarial, aunque la funcionalidad específica puede variar según el proveedor y la configuración del sistema.

Principales Sistemas de Gestión Empresarial en la Industria del Transporte

En la industria del transporte, los sistemas de gestión empresarial (ERP) son fundamentales para administrar eficientemente los recursos, coordinar las operaciones y satisfacer las demandas de los clientes. Algunos de los principales sistemas de gestión empresarial utilizados en la industria del transporte se muestran en la tabla 5.0. (Alonso, F., 2017).

SAP Transportation Management (SAP TM)	*TMW Systems (Trimble Transportation)*
Oracle Transportation Management (OTM)	*Blue Yonder (anteriormente JDA Software)*
Infor CloudSuite Transportation	*Manhattan Associates Transportation Management System (TMS)*
CargoWise (WiseTech Global)	*Descartes Systems Group*
Microsoft Dynamics 365 Supply Chain Management	*Logistics Management Solutions (LMS)*

Tabla 5.0 Principales ERP empleado en la industria del transporte.

Los ERP de la tabla 5.0 ofrece funciones avanzadas de gestión de transporte, incluida la planificación y ejecución de envíos, optimización de rutas, seguimiento de carga y gestión de eventos, gestión de almacenes y análisis de costos (Cano, M. D., 2014).

Estos sistemas de gestión empresarial están diseñados específicamente para satisfacer las necesidades únicas de la industria del transporte, proporcionando herramientas y funcionalidades que permiten a las empresas gestionar eficientemente sus operaciones logísticas y optimizar sus cadenas de suministro. La elección del sistema adecuado dependerá de los requisitos específicos de la empresa, el tamaño de la flota, el tipo de carga y otros factores relevantes (Cano, M. D., 2014).

Lenguajes de Programación Empleados en los Sistemas de Gestión Empresarial

Los sistemas de gestión empresarial (ERP) pueden involucrar una variedad de lenguajes de programación, dependiendo de varios factores, como la arquitectura del sistema, la tecnología subyacente y las preferencias del desarrollador. Algunos de los lenguajes de programación (Verhas, T, 2020) comúnmente empleados en el desarrollo de sistemas de gestión se listan en la tabla 6.0.

Java	Python
C#:	*SQL (Structured Query Language*)
JavaScript	*ABAP (Advanced Business Application Programming*)

Tabla 6.0 Principales lenguajes de programación en los ERP.

Aplicación de los Sistemas de Gestión Empresarial

Para ejemplificar la aplicación de los sistemas ERP en una empresa de reparto, se presenta el manejo y administración de una empresa de reparto puerta a puerta con el análisis de los siguientes aspectos: identificación de paquetes, asignación de rutas, seguimiento en tiempo real, y horarios. (Coronel, C., & Morris, S., 2018).

Identificación de Paquetes

El sistema ERP tiene una base de datos centralizada para la identificación de paquetes, donde cada paquete posee un número de identificación (ID), con información del cliente, fecha, tipo de envío y código postal del destinatario entre otra basta información. Lo que se puede apreciar en la Ilustración 1.0

ID do envio	ID	Tipodeenvio	Cp		Ciclo	MLP	DA	FECHA
42506881138	47146381	Comercial	6100	#N/A	#N/A	#N/A	#N/A	19/5/2024
42511039696	47146381	Residencial	6100	#N/A	#N/A	#N/A	#N/A	19/5/2024
42511466708	47146381	Residencial	6100	#N/A	#N/A	#N/A	#N/A	19/5/2024
42513799854	47146381	Residencial	6100	#N/A	#N/A	#N/A	#N/A	19/5/2024
42515178165	47146381	Residencial	6100	#N/A	#N/A	#N/A	#N/A	19/5/2024
42515314063	47146381	Residencial	6100	#N/A	#N/A	#N/A	#N/A	19/5/2024
42517172678	47146381	Comercial	6100	#N/A	#N/A	#N/A	#N/A	19/5/2024
42517198833	47146381	Residencial	6100	#N/A	#N/A	#N/A	#N/A	19/5/2024
42517308862	47146381	Residencial	6100	#N/A	#N/A	#N/A	#N/A	19/5/2024
42517715442	47146381	Comercial	6100	#N/A	#N/A	#N/A	#N/A	19/5/2024
42517792520	47146381	Comercial	6100	#N/A	#N/A	#N/A	#N/A	19/5/2024

Ilustración 1.0 Identificación de paquetes. (Fuente: Software Logistic ML)

Por otra parte, el sistema ERP permite trazar rutas de reparto (ver ilustración 2.0a), asignar el número de paquetes a cada unidad de reparto con el número de paradas y el número de paquetes que se repartirán en la ruta (ver ilustración 2.0b).

Imagen 2.0 a) Rutas de reparto, b) asignación de paquetes, unidades, ruta y paradas de reparto. (Fuente: Software Logistic ML)

Asignación de Terceras Empresas para Distribución

En el reparto de puerta a puerta se manejan empresas las cuales son llamadas terceras empresas para distribución o conocidas como *outsourcing logístico* o *Third-Party Logistics* (3PL), implica la subcontratación de la logística y la distribución de productos a empresas especializadas. Estas empresas son contratadas para manejar diversos aspectos de la cadena de suministro y distribución para otras compañías. Los servicios que ofrecen pueden incluir almacenamiento, gestión de inventarios, transporte, distribución, y otros servicios relacionados como se puede apreciar en la figura 3.0 con los nombres de MLP.

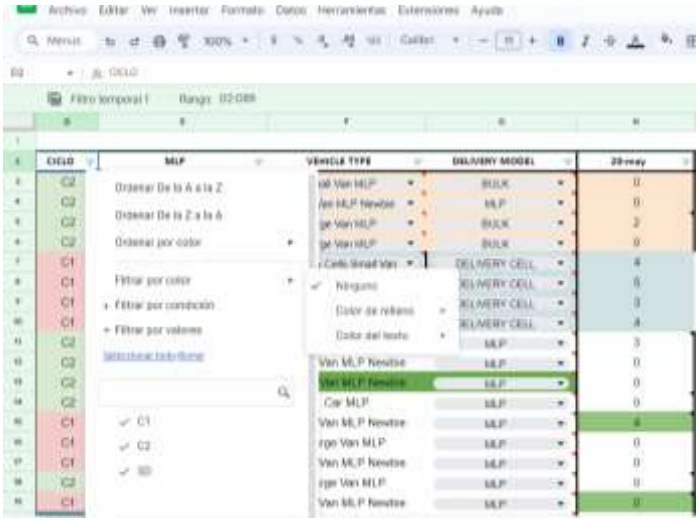


Ilustración 3.0 Manejo de terceras empresas de distribución ***(3PL)*** *(Fuente: Software Logistic ML)*

Las ventajas de utilizar 3PL se refleja en la reducción de costos, flexibilidad en la operación de la empresa, enfocarse en el negocio principal, acceso a tecnología avanzada. Podemos listar las empresas más importantes en 3PL: DHL Supply Chain, Kuehne Nagel, XPO Logistics, FedEx Supply Chain, UPS Supply Chain Solutions.

Seguimiento en Tiempo Real

Los sistemas ERP ofrecen seguimiento en tiempo real en la distribución lo que requiere capacidad de monitorear y gestionar el movimiento de bienes y servicios a medida que se desplazan a través de la cadena de suministro por parte de los administradores del sistema (Ver ilustración 4.0). Esto es esencial para asegurar la eficiencia, transparencia y satisfacción del cliente.

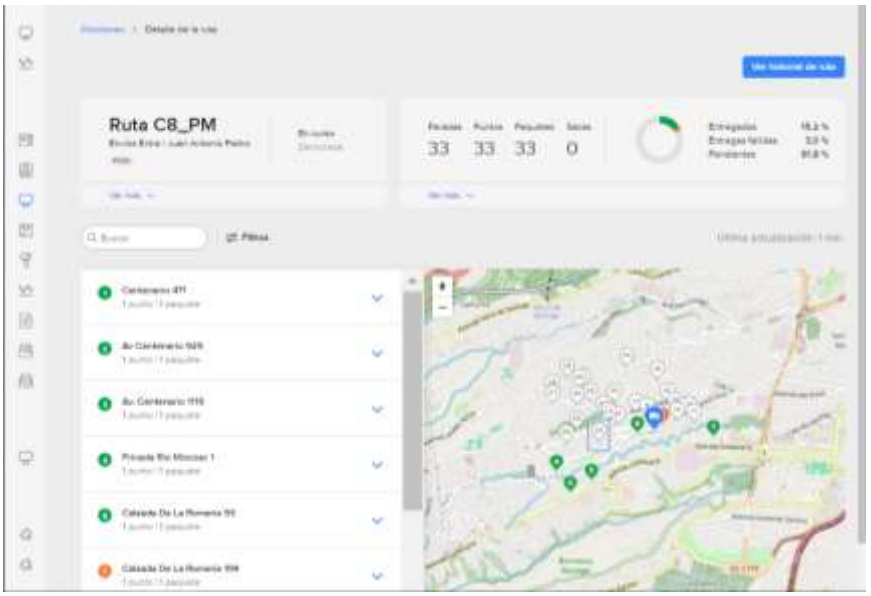


Ilustración 4.0 Seguimiento en tiempo real. (Fuente: Software Logistic ML)

Reparto de Última Milla

Hay muchos factores presentes dentro del proceso de entrega de última milla. La planificación de rutas manual demanda grandes cantidades de tiempo y generalmente produce resultados subóptimos, aumentando los costos de operación. Para lograr una buena optimización de rutas es importante considerar aspectos como la asignación de flota y horarios de reparto.

Asignación de Flota

Es fundamental contar con un buen control sobre los vehículos que se tienen disponibles. Debe existir una buena comunicación entre el área de mantenimiento y el área de planeación de rutas,

ya que, al identificar los vehículos disponibles y la capacidad, determinan el tipo de vehículo que cubrirá cada ruta, acorde con la demanda / volumen de las entregas (Ver ilustración 5.0).

CICLO	Key	Confirmacion
	Delivery Cells Small Van DELIVERY CELL	6
	Large Van MLP MLP	0
	Large Van MLP Newbie MLP	0
	Mix Extra Large Van MLP MLP Mixta	0
	Rental Bike RENTALS	0
	Rental Electric Large Van BULK	3
	Small Van MLP Newbie MLP	5
	Rental Electric Large Van RENTALS	0
	Small Van MLP MLP	21
Total SD		0

Ilustración 5.0 tipo de unidades disponibles para la asignación de ruta. (Fuente: Software Logistic ML)

Horarios

El segundo aspecto para considerar son los horarios en las rutas de entrega, donde se recomienda brindarle al cliente opciones con horarios o ventanas para el reparto de los productos. Esto permite que se logre una adecuada planeación de las rutas, además de asegurar el cumplimiento de lo solicitado por nuestros clientes.

Análisis de Resultados

Los sistemas ERP tienen herramientas de análisis y reporte que permiten a los usuarios generar informes personalizados, analizar datos y obtener información valiosa sobre el rendimiento de la empresa, generalmente con el uso de alguna interfaz de usuario como es el caso de la ilustración 6.0 que se maneja con un Macros en archivo de hojas de Microsoft Excel. Permite generar reportes donde, en este caso es un reporte de entregas fallidas en una hora del día determinada. En el interior de la gráfica de pastel, se indica el número de entregas fallidas por cada una de las empresas Third-Party Logistics (3PL) que participan en el reparto de última milla (llamadas Empresa 1 a Empresa

8) y debajo del nombre de cada empresa el porcentaje de entregas fallidas que le corresponde a cada una de las empresas del total de las entregas fallidas del día específico del reporte. Cabe señalar que el estatus de las entregas fallidas se clasifica en: a) No había nadie en domicilio, b) Negocio cerrado, c) Rechazado, d) Cliente cambio de dirección, e) Zona inaccesible.

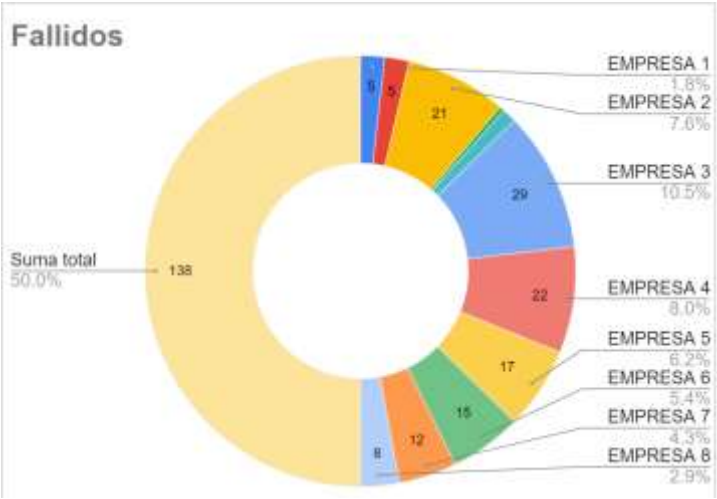


Ilustración 6.0 Reporte en vivo de entregas fallidas por los 3PL contratados. (Fuente: Software Logistic ML).

Por otra parte, en la ilustración 7.0 muestra los diferentes 3PL's que se contratan para la distribución de paquetes, la gráfica ayuda a identificar a las empresas con el mayor número de paquetes pendientes de entrega e implementar un plan de mejora continua a fin de mejorar la calidad del servicio de entrega y la satisfacción del usuario final. En la ilustración 7.0, el número en la parte interna de la gráfica indica la cantidad de paquetes pendientes por entregar, en la parte externa el nombre de las empresas 3PL's y debajo de cada nombre, el porcentaje de entregas pendientes que representa el total de la distribución de la empresa de última milla.

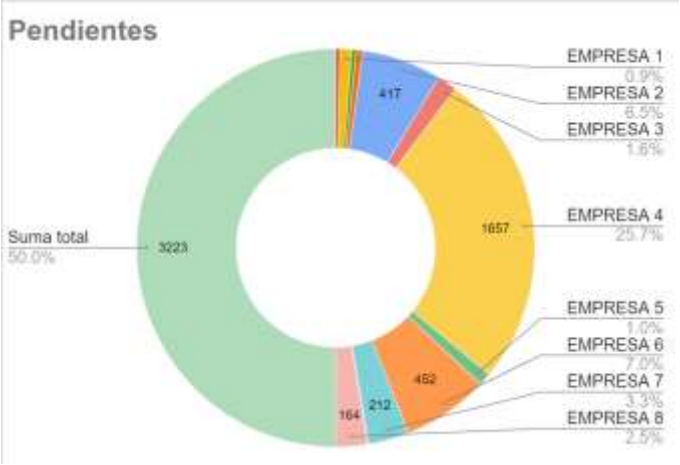
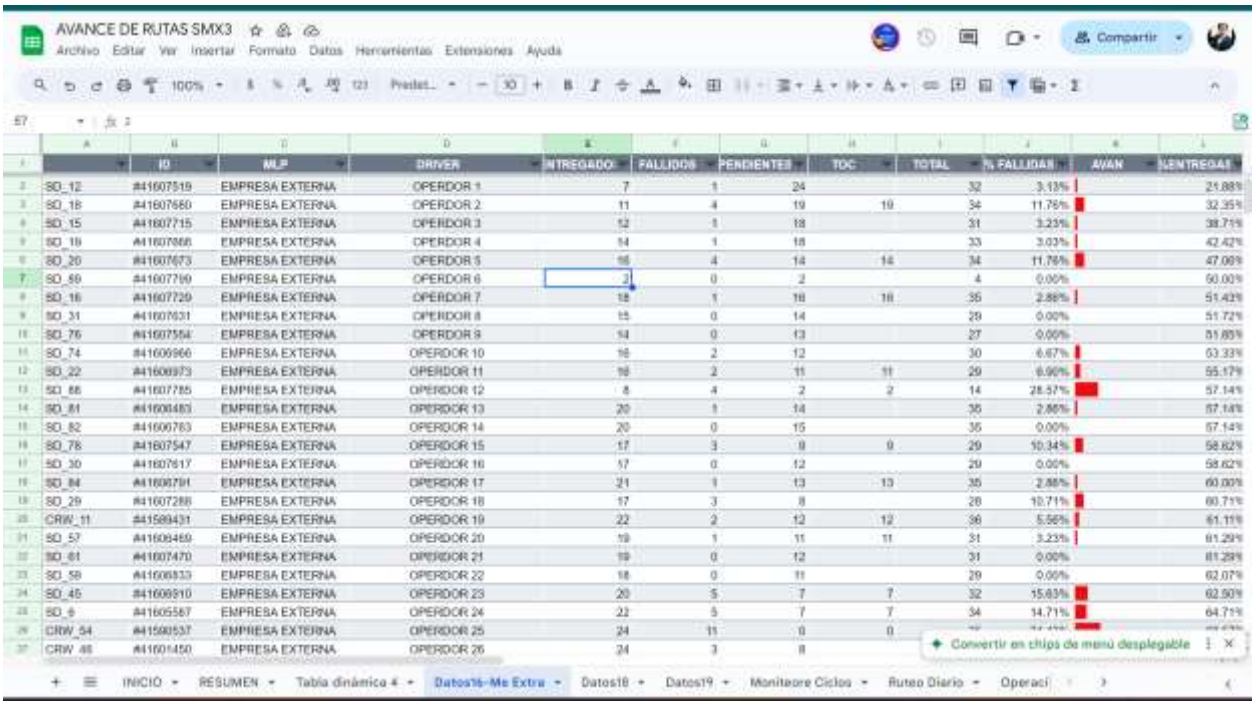


Ilustración 7.0 Reporte en vivo de entregas pendientes por los 3PL contratados. (Fuente: Software Logistic ML)

Métricas de Operadores de Transporte

Por otro lado, la integración en los ERP permite realizar reportes más específicos incluyendo el apartado de operadores (*Drivers*) (ver ilustración 8.0), que es una base de datos generada en tiempo real que indica la clasificación de los operadores con ID, la 3PL'S a la que pertenece (marcado con el título de columna MLP), Nombre del operador, numero de paquetes entregados, fallidos y los porcentajes de paquetes entregados durante el día así como el porcentaje de paquetes fallidos, todo esto en tiempo real, lo que le permite al tomador de decisiones implementar alguna estrategia o protocolo de actuación a fin de cumplir con la meta diaria de entregas satisfechas.



AVANCE DE RUTAS SMX3

	ID	MLP	DRIVER	NTREGADO	FALLIDOS	PENDIENTES	TOC	TOTAL	% FALLIDAS	AVAN	%ENTREGAS
SD_12	#41607519	EMPRESA EXTERNA	OPERDOR 1	7	1	24		32	3.13%		21.88%
SD_18	#41607660	EMPRESA EXTERNA	OPERDOR 2	11	4	19	19	34	11.76%		32.35%
SD_15	#41607715	EMPRESA EXTERNA	OPERDOR 3	12	1	18		31	3.23%		38.71%
SD_19	#41607668	EMPRESA EXTERNA	OPERDOR 4	14	1	18		33	3.03%		42.42%
SD_20	#41607673	EMPRESA EXTERNA	OPERDOR 5	16	4	14	14	34	11.76%		47.06%
SD_59	#41607769	EMPRESA EXTERNA	OPERDOR 6	2	0	2		4	0.00%		50.00%
SD_16	#41607729	EMPRESA EXTERNA	OPERDOR 7	18	1	16	16	35	2.86%		51.43%
SD_31	#41607631	EMPRESA EXTERNA	OPERDOR 8	15	0	14		29	0.00%		51.72%
SD_76	#41607554	EMPRESA EXTERNA	OPERDOR 9	14	0	13		27	0.00%		51.85%
SD_74	#41609966	EMPRESA EXTERNA	OPERDOR 10	16	2	12		30	6.67%		53.33%
SD_22	#41608973	EMPRESA EXTERNA	OPERDOR 11	16	2	11	11	29	6.90%		55.17%
SD_66	#41607785	EMPRESA EXTERNA	OPERDOR 12	8	4	2	2	14	28.57%		57.14%
SD_81	#41608483	EMPRESA EXTERNA	OPERDOR 13	20	1	14		35	2.86%		57.14%
SD_82	#41608763	EMPRESA EXTERNA	OPERDOR 14	20	0	15		35	0.00%		57.14%
SD_78	#41607547	EMPRESA EXTERNA	OPERDOR 15	17	3	9	9	29	10.34%		58.62%
SD_30	#41607617	EMPRESA EXTERNA	OPERDOR 16	17	0	12		29	0.00%		58.62%
SD_84	#41608791	EMPRESA EXTERNA	OPERDOR 17	21	1	13	13	35	2.86%		60.00%
SD_29	#41607288	EMPRESA EXTERNA	OPERDOR 18	17	3	8		28	10.71%		60.71%
CRW_11	#41589431	EMPRESA EXTERNA	OPERDOR 19	22	2	12	12	36	5.56%		61.11%
SD_57	#41608469	EMPRESA EXTERNA	OPERDOR 20	19	1	11	11	31	3.23%		61.29%
SD_61	#41607470	EMPRESA EXTERNA	OPERDOR 21	19	0	12		31	0.00%		61.29%
SD_58	#41608533	EMPRESA EXTERNA	OPERDOR 22	18	0	11		29	0.00%		62.07%
SD_45	#41608910	EMPRESA EXTERNA	OPERDOR 23	20	5	7	7	32	15.63%		62.50%
SD_6	#41605587	EMPRESA EXTERNA	OPERDOR 24	22	5	7	7	34	14.71%		64.71%
CRW_54	#41590537	EMPRESA EXTERNA	OPERDOR 25	24	11	0	0				
CRW_46	#41601450	EMPRESA EXTERNA	OPERDOR 26	24	3	8					

Ilustración 8.0 Reporte en vivo de Drivers. (Fuente: Software Logistic ML)

Métricas Generales

Los sistemas ERP tienen la posibilidad de generar reportes particulares o muy generales en tiempo real. En la ilustración 9,0 se observa el cierre de rutas al final del día, en el cual podemos notar que durante el transcurso de los ciclos salieron 10,079 paquetes de los cuales se entregó el 98,5 % y solo 1,5% no se pudieron entregar.

Ilustración 9.0 Reporte del cierre del día. (Fuente: Software Logistic ML)

Cierre

El análisis de las problemáticas de una empresa de reparto puerta a puerta revela varios desafíos críticos que deben ser abordados para mejorar la eficiencia y la satisfacción del cliente. Entre estos problemas se encuentran la gestión logística, la optimización de rutas, la puntualidad en las entregas, la satisfacción del cliente y la adaptación tecnológica. abordar estas problemáticas requiere un enfoque multifacético que incluya la adopción de tecnologías avanzadas, la optimización de procesos logísticos, el fortalecimiento de la comunicación con los clientes y la implementación de estrategias de gestión de costos eficientes. Las empresas que logren integrar estas mejoras podrán competir eficazmente en un mercado cada vez más exigente y dinámico.

También es necesario contar con perfiles profesionales multidisciplinario que combine conocimientos técnicos (Comprensión del sistema ERP, manejo den bases de datos y SQL, Integración de Sistema, Desarrollo y personalización y seguridad informática), comprensión de procesos empresariales (Conocimiento de operaciones comerciales, Análisis y optimización de proceso, Conocimiento contable y financiero), Capacidad de Gestión de Proyectos (Planificación y ejecución, Coordinación interdepartamental, Gestión del cambio), Habilidades de Resolución de Problemas y Soporte (Análisis y resolución de problemas, Soporte técnico y funcional, Monitoreo y optimización continua). Habilidades Interpersonales y de Comunicación (Comunicación efectiva, Trabajo en equipo, Formación y capacitación), Orientación a Resultados (Enfoque en la eficiencia operative, Análisis de KPIs; interpretación de métricas generadas por los sistemas ERP), Capacidad de Adaptación y Aprendizaje Continuo (Actualización constante Y adaptabilidad) y formación académica y experiencia (educación en ingeniería informática, Sistemas de Información, Administración de Empresas, Finanzas, Contabilidad o Logística)y Experiencia. Todas ellas habilidades claves para el manejo de empresas de reparto que emplean sistemas ERP (Enterprise Resource Planning) para interpretar adecuadamente los resultados numéricos y los reportes generador en los *softwares* de ERP.

Referencias

Alonso, F. (2017). *ERP y Logística: La Solución a los Retos de la Distribución en la Empresa Moderna.* Editorial Planeta.

Cano, M. D. (2014). *El Impacto de los Sistemas de Gestión Empresarial en el Sector Transporte.* Editorial Académica Española.

Coronel, C., & Morris, S. (2018). *Sistemas de Información: Diseño y Aplicación en la Empresa.* Cengage Learning.

Fernández, M. A. (2015). *Implantación de Sistemas ERP en PYMES de Logística y Transporte.* Editorial Académica Española.

Gómez, A., & Prieto, I. (2017). *Logística y Transporte: Gestión de la Cadena de Suministro.* McGraw-Hill Interamericana.

González, F., & Fernández, M. (2018). *Gestión de Flotas y Optimización del Transporte.* Paraninfo.

Jiménez, P. M., & Gómez, C. S. (2013). *Tecnologías de la Información y Gestión Empresarial.* Ediciones Díaz de Santos.

Martínez, E. (2016). *Sistemas de Información Logística y de Transporte.* Ediciones Paraninfo.

Monczka, R. M., Handfield, R. B., Giunipero, L. C., & Patterson, J. L. (2010). *Administración de Compras y Suministros.* Cengage Learning.

Muñoz D. M. d., Cabrita, M. D., & Ribeiro da Silva, M. (2015). Técnicas de gestión empresarial en la globalización. Venezuela: Revista de Ciencias sociales.

Santesmases, M. (2012). *Logística Integral: La Gestión Operativa de la Empresa.* Pirámide.

Verhas, T. (2020). *Scripting with Python: Learn how to develop ERP modules.* Independently Published.

PREDICCIÓN DE MORTALIDAD EN MÉXICO: UN ANÁLISIS COMPARATIVO DE ALGORITMOS DE *MACHINE LEARNING*

MORTALITY PREDICTION IN MEXICO: A COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS

Aldo Ruben Granillo Iturbe[1]
agranilloi001@alumno.uaemex.mx,
ORCID: 0009-0000-7048-3712

Doricela Gutiérrez Cruz[2]
dgutierrezcr@uaemex.mx
ORCID: 0000-0003-2843-3273

Resumen

La elaboración de proyecciones de mortalidad, resulta de una gran utilidad en el desarrollo de las naciones, ya que es gracias a ellas que se pueden tomar acciones para garantizar una mejor calidad y esperanza de vida. Con el fin de contribuir a este esfuerzo en el presente trabajo, se propone un análisis comparativo de algoritmos de *machine learning* para evaluar su rendimiento en tareas de predicción, evaluando un histórico de datos de defunciones en México desde 1950 hasta 2022, donde se pudo observar que la eficiencia de los algoritmos analizados puede verse afectada debido al orden de los datos analizados, asimismo, se proponen dos opciones para mejorar el desempeño: una basada en un mejor ajuste en los algoritmos y otra donde se pretende emplear algoritmos más potentes para llevar a cabo una predicción precisa y certera.

***Palabras clave*:** Machine learning, algoritmos, mortalidad.

[1] Ingeniería en Sistemas Inteligentes, Centro Universitario UAEM Nezahualcóyotl, Universidad Autónoma del Estado de México.
[2] Ingeniería en Sistemas Inteligentes, Centro Universitario UAEM Nezahualcóyotl, Universidad Autónoma del Estado de México.

Abstract

The preparation of mortality projections is very useful in the development of nations since it is thanks to them that actions can be taken to guarantee a better quality of life and life expectancy. In order to contribute to this effort, this paper proposes a comparative analysis of machine learning algorithms to evaluate their performance in prediction tasks, evaluating a history of death data in Mexico from 1950 to 2022, where it could be observed that the efficiency of the algorithms analyzed may be affected due to the order of the data analyzed. Likewise, two options are proposed to improve performance, one based on a better fit in the algorithms and another where it is intended to use more powerful algorithms to carry out an accurate and accurate prediction.

***Keywords*:** Machine learning, algorithms, mortality.

Fecha de envío: 06/06/2024
Fecha de aprobación: 10/11/2024
Fecha de publicación: 01/01/2025

Introducción

La aplicación de la Estadística en el estudio de la mortalidad, tiene una estrecha relación con la calidad de vida de las personas que habitan una región, permitiendo aplicar políticas que busquen mejorar órganos como el sistema público de salud, el acceso a sistemas de salud, el desarrollo económico del país hasta la calidad y el acceso a alimentos de la canasta básica entre otros. Es innegable la importancia de esta estadística para la implementación de diversas políticas, así como para realizar proyecciones demográficas, las cuales son de gran importancia para el desarrollo del país, por ejemplo, no tendría sentido desarrollar la actividad industrial en una entidad en la que la mayoría de sus habitantes no cuentan con una edad dentro del margen de productividad.

La mortalidad es un término usado en múltiples disciplinas, en Estadística se refiere a las muertes totales en un año en un cierto grupo poblacional esto sin importar las causas que la generan, esta métrica tiene como función la generación de información que permita identificar

patrones, tendencias y características (INEGI, 2023). Las proyecciones demográficas en México habitualmente se han realizado con métodos matemáticos como la ley de Gompertz-Makeham (Ogaz, 1991) en la cual se hacen predicciones basadas en probabilidades, algunas de estas presentan fallas en los extremos haciendo referencia en la niñez y ancianidad (Vignoli, 2012), donde, con el paso del tiempo, se esperaría que el grueso de defunciones esté en edades más avanzadas, pero esto no se ve reflejado en los datos donde el grueso de datos está en edades más tempranas, esta dinámica representa una gran dificultad en realizar las proyecciones debido a que no todas las entidades cuentan con las herramientas para hacer dichos cálculos, esto se ve mayormente reflejado en las regiones con mayor pobreza en las que no se cuenta con suficiente información y la que se encuentra resulta deficiente.

Por lo anterior su relevancia impacta a contribuir en el desarrollo de mejores técnicas para la predicción de mortalidad de la población, en el presente trabajo se generan conocimiento a fin de contribuir a través de técnicas de algoritmos de *machine learning* con el objetivo de hacer predicciones respecto a la mortalidad y de ese modo validar la eficiencia entre los algoritmos y métricas aplicadas en el presente documento.

Antecedentes

Galván y Meza (2014) aplicaron técnicas de minería de datos con los algoritmos de Soporte Vectorial Máquinas (SVM), K-Nearest Neighbors (KNN), Redes Neuronales Artificiales (ANN), Algoritmos Genéticos (AG) para identificar patrones y tendencias en la mortalidad por diferentes causas el estudio, donde se encontró que las principales causas de muerte en México son las enfermedades del corazón, la diabetes y los tumores, siendo esta con mayor incidencia en hombres.

Ortiz (2020) analizó las características sociodemográficas asociadas a tasas de mortalidad infantil utilizando datos reportados por el departamento administrativo nacional de estadística (DANE). Donde aplicaron un modelo de máquinas de soporte vectorial (SVM) mediante el cual se recolectaron datos sobre nacimientos y defunciones de menores donde las variables sociodemográficas más relevantes como fueron sexo, edad de la madre, nivel educativo, lugar de residencia donde, tras un análisis de diferentes modelos, se pudo concluir que las SVM fueron las que obtuvieron el mejor desempeño en la métrica de predicción de sucesos de nacido no vivo o vivo.

A la vez García *et al*.,(2011) realizaron un trabajo de Proyección estocástica de la mortalidad mexicana por medio del método de Lee-Carter, un modelo estocástico que permite estimar la evolución futura de las tasas de mortalidad por edad y sexo, donde obtuvieron proyecciones de la mortalidad. En los que se esperó una disminución de la mortalidad en todos los grupos de edad con ello prevén valores de confianza del 95%, donde destaca que el grupo de las mujeres para 2050 será más longevo que el actual. Un modelo de *machine learning* para predecir el riesgo de muerte por COVID-19, se basa en un conjunto de datos de pacientes de diferentes países, donde se obtuvo que el modelo fue capaz de identificar los factores de riesgo más importantes como la edad, sexo, destacando la importancia de considerar las causas de muerte y la estructura por edad de la población (Gil y Quintero, 2022).

Materiales y Métodos

Para el desarrollo de este artículo se utilizará la metodología Knowledge Discovery in Data bases (KDD) en el cual se lleva a cabo la extracción de manera iterativa de conocimiento partiendo de grandes volúmenes de datos donde se aplican técnicas de minería de datos con el fin de obtener nueva información para su análisis, y, de esta manera, generar nuevos conocimientos. Esta metodología desarrolla las siguientes fases:

Selección de Datos

En este estudio se llevará a cabo un análisis para comparar el rendimiento de los algoritmos K-Nearest Neighbors (KNN), RandomForest y regresión lineal para tareas de regresión. En este caso, en la predicción de la mortalidad en México se procesarán los datos proporcionados por el banco de datos abiertos de México (Proyecciones de la Población de México y de las Entidades Federativas, 2020-2070, 2023).

Preprocesamiento de los datos

En esta etapa nos centramos llevar a cabo una limpieza de los registros del *dataset,* el banco de datos contiene 737 661 registros de los cuales, solo se considerarán para este estudio 389 181

datos, siendo estos correspondientes a las defunciones totales en edades desde 0 hasta 109 años que es el rango de edad considerado por el INEGI, estos datos van desde el año 1950 al 2022 dentro de la república mexicana, ya que el 2023 será el año sometido al estudio. En estos registros encontramos las siguientes columnas: “renglón”, “año”,” entidad”, ”cv_geo”, “sexo”, “edad” y “defunciones”. Inicialmente los datos fueron preprocesados para eliminar las columnas que no son de interés para el estudio dejando únicamente “año”, “edad” y “defunciones” como se ilustra en la Tabla 1. La mayoría de los registros son de tipo numérico, por lo cual no fue necesario realizar un procesamiento adicional para poder hacer uso de estos. Para la limpieza del banco de datos se eliminaron datos nulos en caso de existir haciendo uso de funciones de pandas, además se emplearon *data frames* de pandas para dividir el banco de datos y almacenar los datos de interés en estos *data frames*.

Minería de datos

El análisis de los datos se lleva a cabo mediante la implementación de tres algoritmos de *machine learning*, los cuales fueron aplicados mediante la librería Sklearn, los algoritmos son los siguientes:

Regresión lineal, se basa en relaciones lineales entre variables dependientes que van de un objetivo a una o varias variables de destino, para este estudio se emplea una regresión lineal múltiple por el método de mínimos cuadrados ordinarios, este método es útil debido a que calcula los coeficientes de las rectas generadas para obtener una recta en base a la aproximación promedio de todos los puntos.

Random forest, es un algoritmo de regresión basado en árboles, trabaja con múltiples árboles de decisión para generar predicciones las cuales ajusta para obtener una sola predicción más precisa, en este estudio se generaron 100 árboles de decisión para este regresor considerando que ese número se considera optimo evitando así posibles sobreajustes además se consideró las limitaciones computacionales las cuales pueden verse reducidas al ampliar el número de árboles de decisión.

K-Nearest Neighbors* (*Knn), es empleado en tareas de clasificación como de regresión en esta última las predicciones son llevadas a cabo mediante el promedio de las K distancias más cercanas al registro, para este estudio se optó por usar la configuración por default del algoritmo quedando así para el valor de "K" como los 5 vecinos más cercanos debido a que ofrece un buen equilibro con respecto al sobreajuste y sub ajuste

Análisis de los datos

Para la evaluación del rendimiento de estos algoritmos, se emplearon métricas de precisión para el cálculo del error, las cuales son error medio cuadrático, error absoluto medio y el coeficiente de determinación.

Las predicciones obtenidas serán comparadas entre ellas y serán contrastadas contra los datos de defunciones totales durante el año 2023, estas fueron escogidas ya que permiten conocer de manera cuantificable la diferencia entre los valores reales y los valores predichos, estas, a su vez, son útiles no solo para evaluar los resultados, sino que también nos permite ajustar los hiperparámetros de los algoritmos para obtener resultados más óptimos en las predicciones.

El criterio para la aplicación de estos algoritmos consta en la creación de estructuras que contengan los datos del *dataset,* en este caso se usaron los *data frame* que son una estructura propia de pandas en *python*, en el *data frame* se carga el *dataset*, el cual será dividido en grupos que representan las variables independientes y dependientes, esto es necesario para entrenar a los modelos de regresión, ya que necesitan conocer cuál es la variable objetiva, en este caso "Defunciones" representa la variable dependiente, posteriormente se crearán los conjuntos de entrenamiento y prueba con una división de 70% para entrenamiento y 30% para prueba, dando así, una división simple para nuestros datos. Estos conjuntos de datos son pasados a los algoritmos para ser entrenados. Posteriormente se aplicaron las métricas de evaluación de error [Figura 1].

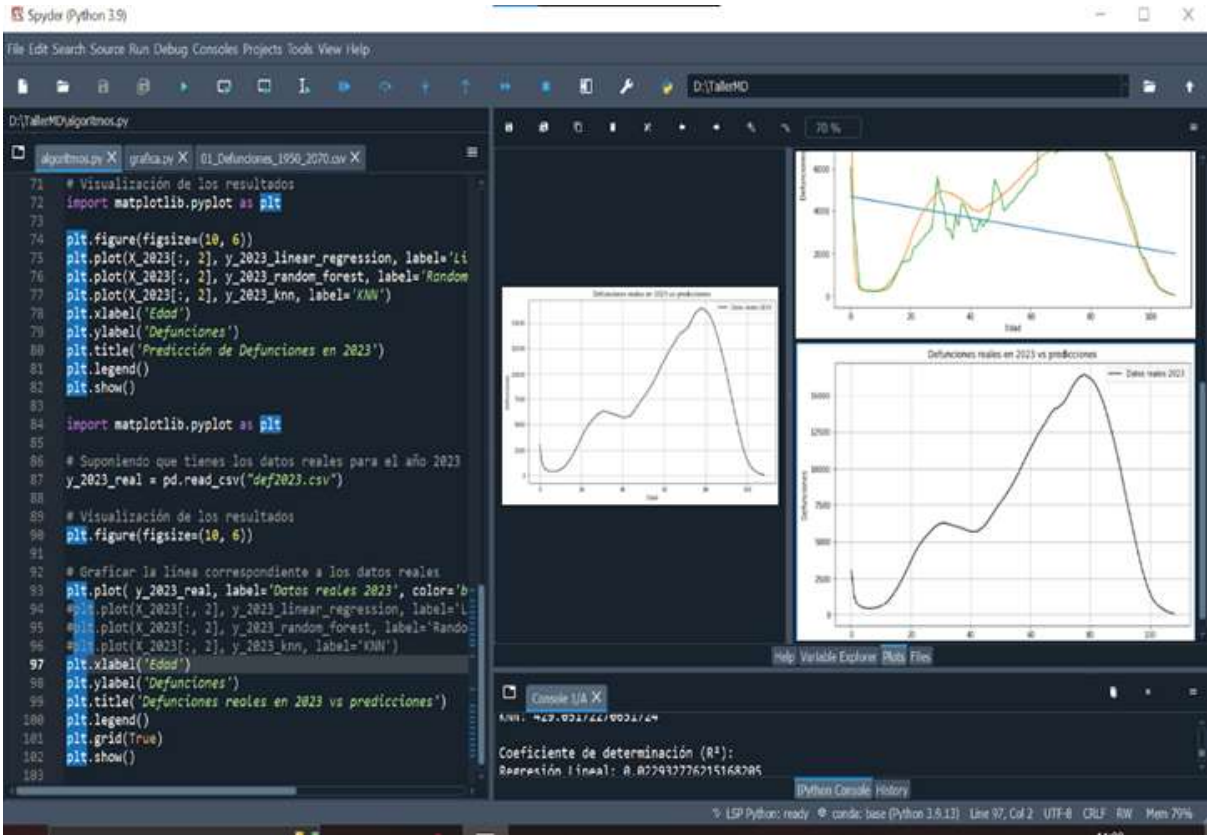


Fig. 1. visualización del programa, (elaboración propia 2024).

Estas métricas tienen diversas formas de evaluar los resultados de las predicciones de los algoritmos. El error medio cuadrático mide la diferencia promedio al cuadrado entre los valores reales y los predichos por el modelo, por otra parte, el error absoluto medio solo contempla la diferencia promedio de los valores reales entre los valores predichos a su vez el coeficiente de determinación proporción de variabilidad de la variable dependiente en base a la independiente, estas métricas fueron calculadas mediante su implementación en código. Posteriormente se llevó a cabo la predicción de defunciones totales para el año 2023 llevando a cabo una comparación entre los datos reales de defunciones totales de dicho año contra las predicciones hechas por cada uno de los modelos [Figura 2].

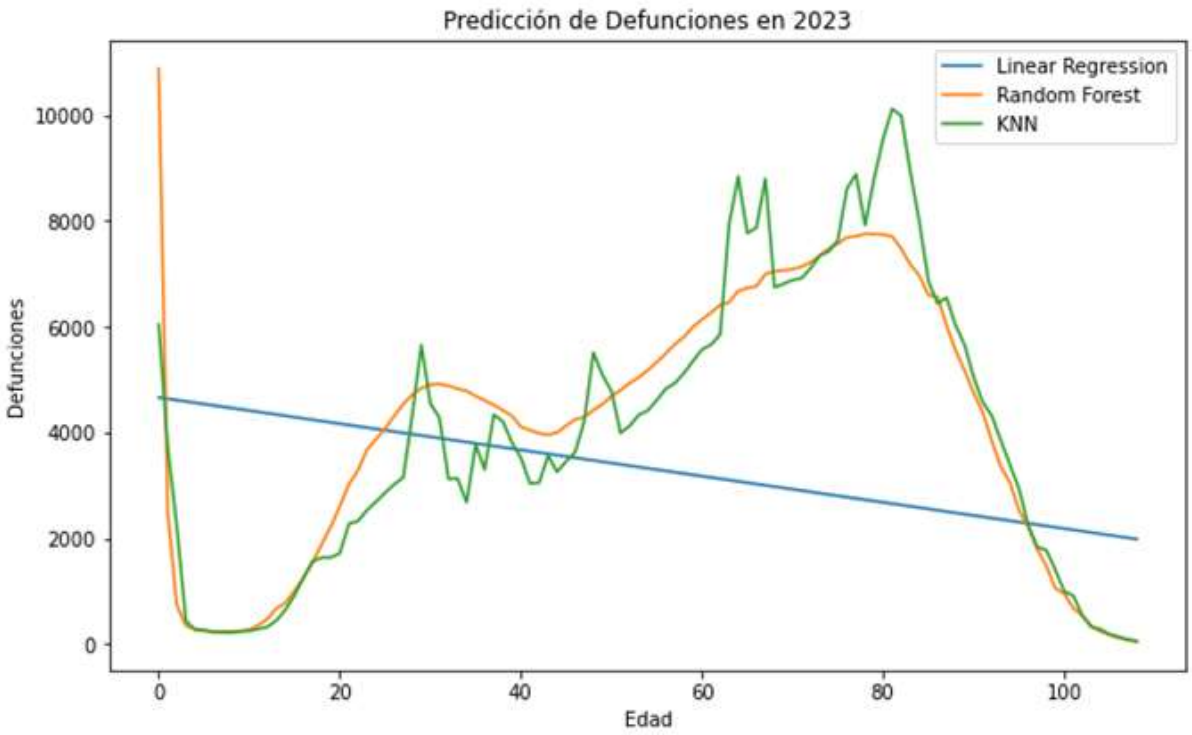


Fig. 2. Predicciones de los algoritmos, (elaboración propia 2024).

Resultados

En este estudio se llevó a cabo una comparativa para evaluar el rendimiento de tres algoritmos, generando así una predicción de mortalidad en México para el año 2023. En este estudio se contemplaron únicamente las muertes totales. Dichos algoritmos, fueron sometidos a una evaluación de su rendimiento mediante métricas de error (Tabla 2), estas son un indicativo de la calidad de predicciones que estos algoritmos realizan.

Tabla. 1. Tabla de errores de los algoritmos, (elaboración propia 2024).

Algoritmo	Error medio cuadrático	Error absoluto medio	Coeficiente de determinación
Regresión lineal	32,684,262.36	2036.83	0.02
Random Forest	61,943.92	53.97	0.99
Knn	5,150,698.03	429.05	0.84

Generalmente suelen considerarse rangos de 0 a 1 para considerar la exactitud de la predicción de un algoritmo, aunque eso no implica directamente que si se obtienen valores fuera de este rango no sean de utilidad, esto depende del trabajo que se esté realizando, es posible observar que para el error medio cuadrático y el error absoluto medio, se obtuvieron valores muy elevados, indicando que los valores predichos por los algoritmos presentan desviaciones considerables respecto a sus valores reales. A diferencia del coeficiente de determinación, se sugiere un mejor ajuste para los algoritmos, exceptuando a la regresión lineal que, hasta ahora, ha sido el que ha obtenido valores deficientes, mientras que, a su vez, el algoritmo *random forest* es el que mejor desempeño ha obtenido. Durante el año pasado 2023 en México fallecieron un total de 805 313 personas. Esto acorde a los registros del *dataset* en el cual se almacenaron las muertes totales del banco de datos original. Dichas defunciones fueron graficadas haciendo uso de *matplotlib*, esto se ve reflejado en la Figura 3.

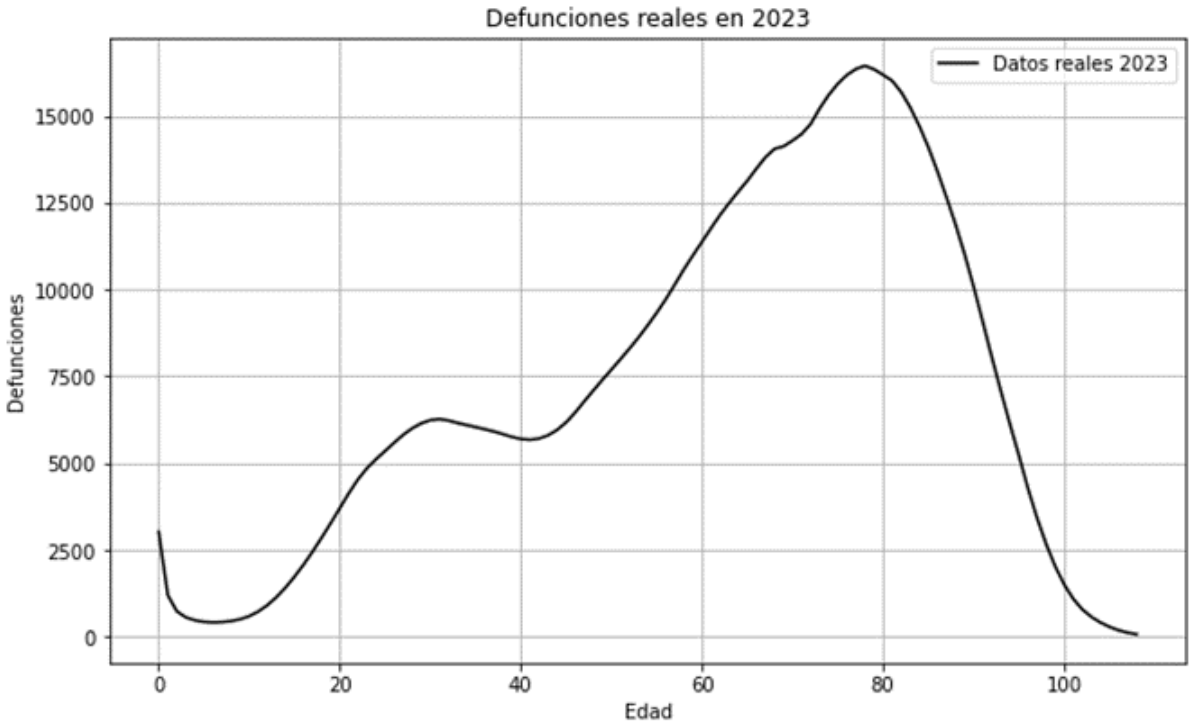


Fig. 3. Gráfica de defunciones totales en México en el año 2023, (elaboración propia 2024).

A continuación, se muestra en la Figura 4 las predicciones hechas por los algoritmos ya mencionados anteriormente.

Resulta interesante la visualización de las predicciones, nótese que la más deficiente, fue el algoritmo de regresión lineal, ya que no es capaz de representar los picos que aprecian respecto a los datos originales, esto puede deberse a la lógica propia del algoritmo, la cual se centra en encontrar una ecuación lineal que describa correctamente la correlación entre las variables.

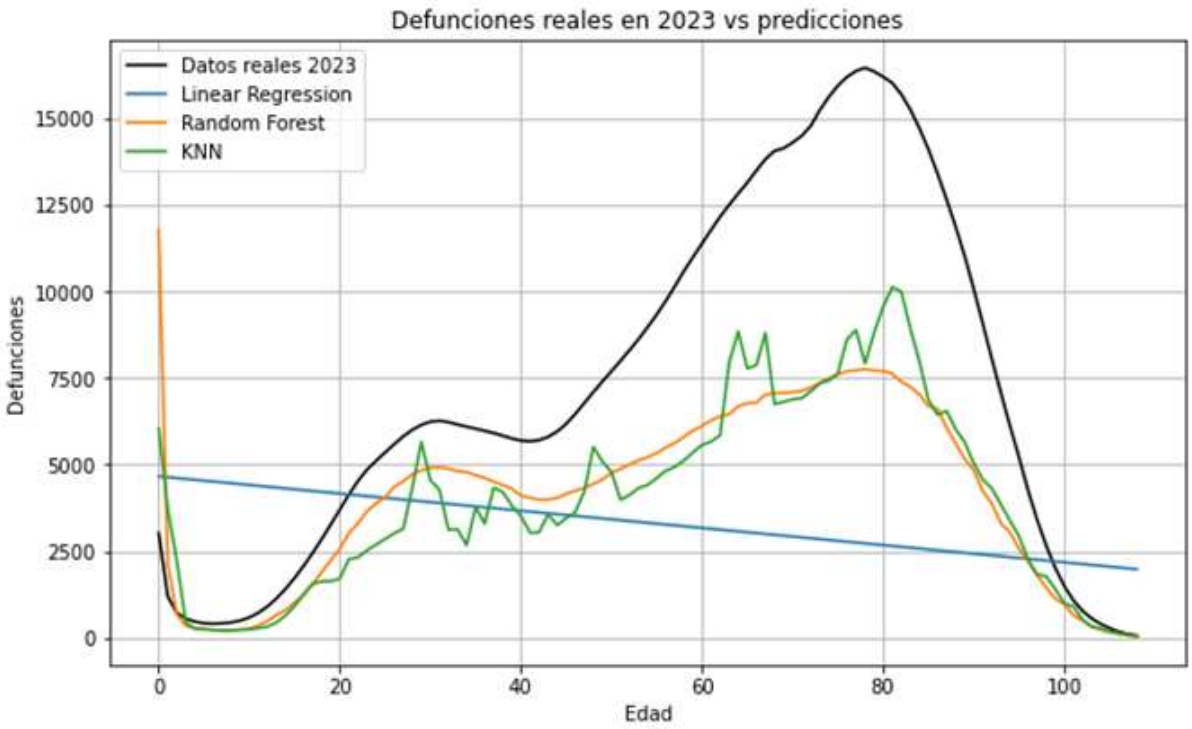


Fig. 4. Predicciones de los algoritmos respecto a los datos reales, (elaboración propia 2024).

Asimismo, los algoritmos *knn* y *random forest,* en específico este último, fueron los que obtuvieron mejores resultados a comparación de la regresión lineal, no obstante, aún los resultados distan bastante de la información real. El algoritmo de regresión lineal, predice un total de 333 414, siendo una diferencia de 471 899 muertes una diferencia del 58% con los datos reales. Por otra parte, los algoritmos de *random forest* y *knn* respectivamente predijeron 404 156 y 429 568 decesos con una diferencia respecto a los datos reales de 50% y 53% respectivamente, lo cual no es muy óptimo en donde se busca tener un índice de certeza considerable como podría ser un 90% de los datos. Si bien las predicciones de mortalidad son una expresión que surge al tratar de estimar el comportamiento de los decesos respecto a factores como la edad, género, entre otros, resulta imprescindible la implementación de técnicas que satisfagan de la manera más óptima dicha necesidad. Asimismo, las técnicas de *machine learning* representan aún una gran área de aprovechamiento, pero resulta imperante tomar en cuenta el diseño de los datos, el cómo son estructurados, y el algoritmo a implementar para que así las posibilidades de presentar una mala implementación de los algoritmos predictivos sea menor Dromundo (2018).

Cierre

En la presente investigación se compararon tres algoritmos de *machine learning* para realizar predicciones sobre la mortalidad en México para el pasado año 2023, los cuales se apoyaron del banco de datos abiertos con el *dataset* de defunciones en México.

Se obtuvieron valores de error mediante el uso de métricas para determinar el error de dichas predicciones, esto en base a los datos reales. Los resultados obtenidos son una evidencia de que, a pesar de que los algoritmos de *machine learning* son una herramienta muy potente y útil, en múltiples casos aún no pueden ser equiparables a las técnicas que comúnmente son empleadas o algoritmos específicamente creados para realizar estos cálculos como la tasa bruta de mortalidad Ogaz (1991) haciendo referencia a este estudio en particular.

Si bien se contaba con una amplia gama de datos, siendo estos recopilados desde 1950 hasta el 2022, resulta difícil la tarea predictiva, ya que las defunciones no son exclusivamente de un orden natural, debido a que no todas las defunciones registradas son provocadas por la misma causa; algunas son orden natural como la vejez, enfermedades, otras son por agentes externos como los fenómenos naturales. Es por esto que la tarea predictiva aumenta su complejidad, un ejemplo de esto es el SARS-COV 2 que fue un fenómeno externo que causó múltiples decesos (García y Pérez, 2021; Gil y Quintero, 2022). Existen un sinfín de elementos los cuales hay que considerar a evaluar, por ello, un trabajo a futuro podría considerarse la aplicación de algoritmos más potentes como las redes neuronales para realizar este tipo de predicciones. Esto debido a su capacidad de aprender patrones para simular con un comportamiento más natural.

Asimismo, evaluar en periodos más cortos de tiempo de un año, haciendo ajustes en los hiperparámetros de los algoritmos empleados, ya que, como se mencionó, múltiples factores pueden afectar el cálculo de la predicción a causa de múltiples factores que pueden intervenir en plazos tan largos de tiempo. A su vez, podría considerarse hacer un análisis más particular sobre algunas afecciones o periodos de tiempo específicos que sean propensos a generar interés para analizar donde se proponga llevar a cabo una segmentación más certera de los datos, donde la sección de los registros se ajuste a los periodos que serán sometidos a estudio, y, en base a ello, construir un mejor modelo para la tarea predictiva.

Referencias

Antequera, M. G., Sendra, A. L., & Lama, J. R. (2020). Algoritmos de machine learning y su aplicación al mantenimiento industrial en el sector agroalimentario. *Dialnet*, 165-169. https://dialnet.unirioja.es/servlet/articulo?codigo=7655416

Bermúdez, C & Bermúdez, S. (2019). Sistema de predicción de estadísticas de nacimientos y defunciones en Colombia soportadas por TIC. Universidad Cooperativa de Colombia, Facultad de Ingenierías, Maestría en Gestión de Tecnologías de la Información, Bucaramanga. https://hdl.handle.net/20.500.12494/10746

Dávila, C. A. (2012). Ajuste matemático de la mortalidad general en México 2000, 2005 y 2010. *Papeles de población*,*18* (74), 01-34. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-74252012000400006&lng=es&tlng=es.

Dromundo, A. X. U. (2024). *Machine learning y su importancia en la actualidad*. IPADE. https://www.ipade.mx/newsmedia/tendencias/machine-learning-y-su-importancia-en-la-actualidad/

Galván. P. & Meza, A. (2014). *Estudio de minería de datos para la información de mortalidad en México*. http://132.248.52.100:8080/xmlui/handle/132.248.52.100/2789

García, E y Pérez, H. (2021). La mortalidad por COVID-19 en México. https://dsp.facmed.unam.mx

García, V. M. & Ordorica, M. (2012). Proyección estocástica de la mortalidad mexicana por medio del método de Lee-Carter. *Estudios demográficos y urbanos*, *27*(2), 409-448. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0186-72102012000200409&lng=es&tlng=es.

Gil, V. & Quintero, C. (2022). Predicción del riesgo de muerte por COVID-19 con Machine Learning. https://www.researchgate.net/profile/Victor-Gil-Vera/publication/362781339_Prediccion_del_riesgo_de_muerte_por_COVID-19_con_Machine_Learning/links/62feb497ceb9764f7206d76c/Prediccion-del-riesgo-de-muerte-por-COVID-19-con-Machine-Learning.pdf

Instituto Nacional de Estadística y Geografía (INEGI). (2023). Mortalidad. https://www.inegi.org.mx/programas/mortalidad/?fbclid=IwAR3INMoyRI7YIPhdqHO5TR98oG6qmyncoDS_dlWvTv9Fwe1gGn1gBFtR5x4

Mina, A. (2006). Ley de mortalidad mexicana. Funciones de supervivencia. *Estudios Demográficos y Urbanos*, *21*(2), 431–456. https://doi.org/10.24201/edu.v21i2.1255

Ogaz, H. (1991). La función de Gompertz-Makeham en la descripción y proyección de fenómenos demográficos. E*studios Demográficos y Urbanos*, *6*(3), 485–520. https://doi.org/10.24201/edu.v6i3.820

Ortiz, V. (2020). *Aplicación de técnicas de aprendizaje automático para la segmentación y clasificación de características sociodemográficas asociadas a tasas de mortalidad infantil utilizando datos reportados por el DANE Colombia entre los años 2008 al 2017*. http://hdl.handle.net/20.500.12010/19171.

Vignoli, J. R. (2011). Reproducción adolescente y desigualdades. *Revista Latinoamericana de Población*, *5*(8), 87-113. https://doi.org/10.31406/relap2011.v5.i1.n8.6

MINERÍA DE PROCESOS EDUCATIVA APLICADA A ENTORNOS VIRTUALES DE NIVEL UNIVERSITARIO: UNA REVISIÓN SISTEMÁTICA

EDUCATIONAL PROCESS MINING APPLIED TO UNIVERSITY LEVEL VIRTUAL ENVIRONMENTS: A SYSTEMATIC REVIEW

Marisol Felicia Vázquez Varela[1]

mvazquezv005@alumno.uaemex.mx

ORCID: 0009-0002-6433-212X

Alejandra Morales Ramírez[2]

amoralesr@uaemex.mx

ORCID:0000-0002-8737-5985

Cuauhtémoc Hidalgo Cortés[3]

chidalgoc@uaemex.mx

ORCID: 0000-0001-6324-7180

Resumen

En este estudio se realizó una revisión de la literatura para examinar cómo se ha utilizado la minería de procesos educativa para obtener conocimiento sobre los procesos de enseñanza-aprendizaje que utilizan entornos educativos virtuales a nivel universitario de abril de 2018 y diciembre de 2023. El proceso de revisión y elección de las publicaciones encontradas se efectuó en base a las directrices PRISMA. La calidad metodológica se evaluó a través de la Critical Appraisal Checklist For Analytical Cross Sectional Studies desarrollada por JBI. Veintiséis estudios transversales cumplieron con los criterios de elegibilidad y se llevaron a cabo principalmente en Europa y Asía. La minería de procesos educativa muestra un potencial para mejorar la enseñanza universitaria al

[1] Universidad Autónoma del Estado de México.
[2] Universidad Autónoma del Estado de México.
[3] Universidad Autónoma del Estado de México.

detectar patrones, analizar comportamiento y proponer mejoras. Ha aumentado la satisfacción y usabilidad de los entornos educativos virtuales, pero enfrenta desafíos con la calidad de los datos. Sin embargo, existe un fuerte interés en continuar avanzando en este campo.

Palabras clave: Minería de procesos, minería de procesos educativa, sistema de gestión de aprendizaje, universidad.

Abstract

The literature was reviewed to examine how educational process mining has been used to gain insight into teaching-learning processes using virtual educational environments at the university level, between April 2018 and December 2023. The review and selection process of the publications found was carried out based on the PRISMA guidelines. The methodological quality was assessed using the Critical Appraisal Checklist for Analytical Cross-Sectional Studies, developed by JBI. Twenty-six cross-sectional study met the eligibility criteria and were mostly conducted in Europe and Asia. Educational process mining shows potential to improve university teaching by identifying patterns, analyzing behavior, and proposing improvements. It has increased satisfaction and usability of virtual educational environments but faces challenges with data quality. However, there is strong interest in continuing to advance in this field.

Keywords**:** Process mining, educational process mining, learning management system, university.

Fecha de envío: 06/06/2024
Fecha de aprobación: 10/11/2024
Fecha de publicación: 01/01/2025

Introducción

La llegada de la pandemia del COVID-19 y el creciente acceso a Internet en los últimos años, han acelerado el uso de las plataformas digitales para el aprendizaje, generalmente implementadas en

los sistemas de gestión del aprendizaje o Learning Management Systems (LMS), que son plataformas tecnológicas esenciales dentro de la educación virtual, que no solo facilitan la enseñanza, sino que también estructuran el proceso de aprendizaje mediante la gestión integral de los cursos. Los LMS, como parte de las Tecnologías de la Información y la Comunicación (TIC), permiten a las instituciones educativas ofrecer un entorno digital que engloba desde la distribución de contenidos hasta la evaluación del rendimiento de los estudiantes. Además, se consideran un punto de convergencia esencial entre estudiantes y docentes, lo que facilita la interacción del proceso de enseñanza-aprendizaje.

Estos LMS puede ser de código abierto (ej. Moodle, Chamilo y Open Edx) o comercial (ej. Docebo, Google Classroom y Schoology), y ofrecen una amplia gama de funciones que favorecen el aprendizaje virtual, incluyendo la gestión de cursos, el seguimiento del progreso del alumno, evaluaciones digitales, comunicaciones, seguridad del sistema y el acceso a través de teléfonos inteligentes (Turnbull *et al*., 2020). En el contexto educativo actual, tanto los LMS de código abierto como los comerciales, se han consolidado como el corazón de la enseñanza virtual, siendo cada vez más utilizados no solo en la modalidad online, sino también, como soporte en la educación presencial y mixta (Ramírez, 2021). Además, la capacidad de los LMS para registrar cada interacción y actividad de los estudiantes (rastros de los procesos ejecutados) genera una vasta producción de datos (registros de eventos).

Estos registros no solo ofrecen una visión de la actividad educativa, sino que también representan una fuente valiosa de información que puede ser explorada y analizada a través de la Minería de Procesos Educativa (MPE), un área especializada del campo de la Minería de Procesos (MP) que se enfoca en la comprensión y mejora de los procesos educativos y de aprendizaje. Combina las herramientas y técnicas de la MP para analizar los datos que van dejando los estudiantes dentro de los LMS, los cuales corresponden a las acciones realizadas sobre el contenido del curso. Por ejemplo, cuando acceden al material, desarrollan sus actividades, dan solución a una lección o ejercicio, participan en un chat o foro, consultan notas y evaluaciones e interactúan con los instructores (Bogarín *et.al*., 2018).

Con la información obtenida, las personas interesadas pueden modelar y visualizar procesos que muestren el lugar donde los estudiantes comenzaron y terminaron sus rutas de aprendizaje, lo que permitirá (Aguirre y Rincón, 2015; Carnis *et al*., 2015):

- Identificar patrones y anomalías: entender cómo es la interacción de los estudiantes con el contenido educativo y detectar comportamientos o secuencias que conduzcan a mejores resultados.

- Perfeccionar procesos educativos: detectar las áreas de mejora en la organización de cursos, la metodología de enseñanza o en la interacción entre estudiantes y profesores con la finalidad de aumentar la eficacia del aprendizaje.

- Evaluar y medir el rendimiento educativo: estudiar el avance de los estudiantes a lo largo del tiempo y determinar qué enfoques de enseñanza son mejores.

- Caracterizar la experiencia de aprendizaje: comprender las exigencias individuales de los estudiantes para poder adaptar el contenido y las estrategias educativas para que se maximice la participación y el aprendizaje del estudiante.

- Validar modelos pedagógicos: probar y validar modelos y teorías educativas a partir de datos reales.

En resumen, la MPE permite a los educadores, administradores y analistas, obtener una visión más profunda de los procesos educativos, proporcionando bases para la toma de decisiones con la finalidad de mejorar la calidad y la eficacia de la educación (Etinger *et.al*., 2018).

El presente estudio pretende ampliar nuestra comprensión sobre el empleo de las técnicas y herramientas de la MP en el ámbito educativo. Por lo anterior, el propósito del estudio fue revisar la literatura que muestra evidencia acerca de cómo se ha utilizado la MPE para obtener información relevante sobre los procesos de enseñanza-aprendizaje a nivel universitario que se llevan a cabo a través de entornos educativos virtuales con la finalidad de identificar posibles áreas de oportunidad para futuras investigaciones.

Marco Teórico

El presente apartado brinda una descripción teórica concisa sobre la MP en el contexto educativo, abarcando las técnicas, algoritmos y modelos de representación de procesos empleados en esta disciplina.

Minería de Procesos

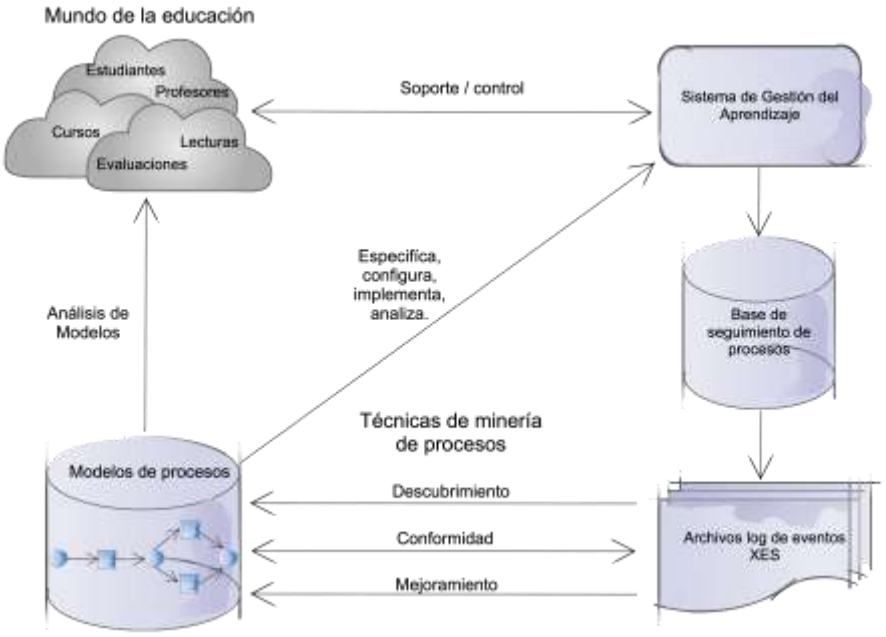
La MP tiene como propósito principal extraer conocimiento de los registros de eventos generados por los sistemas de información, con el fin de mejorar la comprensión, el control y la optimización de los procesos empresariales. Aunque fue desarrollada a finales de los 90's por el investigador Van der Aalst, su implementación significativa no se dio hasta 2011, cuando la IEEE publicó el Manifiesto de la MP. Este documento se creó con el fin de servir como una guía para que desarrolladores de *software*, consultores, científicos, gerentes y usuarios finales pudieran optimizar sus procesos en operaciones comerciales (Van der Aalst, 2016). Debido a esto, hoy en día, la MP es ampliamente utilizada para mejorar procesos en diversos sectores (ej. los sectores empresarial, educativo y gubernamental).

Minería de Procesos Educativa

Dentro del ámbito educativo, la MP aparece con el término de MPE que es un área especializada de la MP que se centra en el análisis, comprensión y mejora de los procesos educativos y de aprendizaje en entornos virtuales o presenciales. Su objetivo principal es analizar los datos generados por las interacciones de los estudiantes con los LMS y otras plataformas educativas, extrayendo información valiosa que permita optimizar el proceso de enseñanza-aprendizaje. La MPE ayuda a modelar y visualizar cómo los estudiantes acceden y utilizan los recursos educativos, lo que permite identificar patrones de comportamiento, medir el rendimiento académico, personalizar las experiencias de aprendizaje y validar modelos pedagógicos (Bogarín *et al*., 2017; Ghazal, 2017).

La Figura 1 presenta el marco general de la MPE, donde se destacan los componentes involucrados. Las interacciones de los estudiantes con los LMS generan un conjunto de eventos que se almacenan en bases de datos. Estos eventos se integran en un registro de eventos (extensión XES), el cual se utiliza como insumo para aplicar las técnicas de MPE, con el fin de generar modelos de procesos educativos que aporten información valiosa para la toma de decisiones.

Figura 1. Marco general de la MPE



Fuente. Elaboración propia a partir de Bogarín et al., (2017)

Técnicas de la MP

Las principales técnicas de la MP que se adaptan al análisis de los procesos educativos (Caeiro *et al.,* 2018; Bogarín, 2018; Intayoad *et al.,* 2018), incluyen:

- *Descubrimiento de Procesos (Process Discovery):* esta técnica se utiliza para identificar y modelar automáticamente los procesos educativos a partir de los registros de eventos generados en los LMS. Permite obtener una representación visual o un modelo de los flujos de trabajo de los estudiantes, mostrando cómo interactúan con el contenido, las

evaluaciones y las actividades de los cursos. Frecuentemente, se utilizan redes Petri como notación de referencia para generar los modelos, pero también se pueden utilizan otros, como: mapas de procesos, redes causales, arboles de procesos, etc. Además, en algunos casos, también se utilizan otro tipo de diagramas para representar redes sociales asociadas a modelos de procesos (Ismail y Tumin, 2019).

- *Conformidad (Conformance Checking):* compara el modelo de un proceso educativo con los registros reales de eventos de los estudiantes, identificando desviaciones o diferencias entre el comportamiento esperado (modelo ideal) y el comportamiento real (datos del LMS). Esto ayuda a detectar problemas o anomalías en la implementación de los cursos.

- *Mejora de Procesos (Process Enhancement):* se utiliza para ajustar y optimizar los modelos de procesos educativos existentes en base a los datos reales. El análisis de estos datos, por ejemplo, permite mejorar la organización del curso, la metodología de enseñanza y la interacción entre los estudiantes y los profesores.

- *Análisis de Redes Sociales (Social Network Analysis):* evalúa las interacciones entre los estudiantes dentro de los entornos virtuales, proporcionando información sobre las dinámicas de colaboración, comunicación y trabajo en equipo. Ayuda a visualizar cómo los estudiantes se conectan entre sí y cómo se distribuyen las redes de conocimiento dentro de un curso.

Algoritmos de MP

Algunos algoritmos de MP que se han utilizado para abordar temas educativos son:

- Algoritmo de Minado Alfa (o algoritmo α). Fue el primero en aplicarse en la MP y logró abordar la concurrencia. Su objetivo principal es cerrar la brecha entre los datos de los registros de eventos y el descubrimiento de modelos de procesos, generando redes de Petri (Van der Aalst *et. al.*, 2004). Sin embargo, se han señalado varias limitaciones en su versión

original, tales como: a) dificultades para manejar datos ruidosos; b) incapacidad para detectar tareas duplicadas u ocultas; y c) problemas al tratar con bucles de longitud uno o dos. Estas deficiencias se han resuelto en versiones más recientes: el algoritmo α+ puede gestionar bucles cortos, el α++ maneja patrones más complejos en el proceso, y el α# es capaz de descubrir tareas ocultas y no observadas (AlQaheri & Panda, 2022).

- Algoritmo de Minado Heurístico (Heuristic Miner). Ofrece varias mejoras en comparación con el algoritmo Alpha: a) considera las frecuencias y la relevancia de los eventos, lo que le permite filtrar comportamientos ruidosos o infrecuentes, haciéndolo menos sensible al ruido y a los registros incompletos; b) es capaz de manejar bucles cortos; y c) permite omitir ciertas actividades individuales. Estas características le otorgan una baja sensibilidad tanto al ruido como a la incompletitud en los registros de eventos (Caeiro *et. al.,* 2018).

  Este algoritmo opera sobre un gráfico de dependencias, el cual puede ser transformado posteriormente en una red de Petri. Los gráficos de dependencias son representaciones visuales en las que los eventos se muestran como nodos, conectados por aristas dirigidas que reflejan cálculos de frecuencia y rendimiento. La frecuencia en un gráfico de dependencias indica cuántas veces un evento de origen ocurre después de un evento de destino, mientras que las medidas de rendimiento permiten analizar el tiempo transcurrido entre ambos eventos durante el descubrimiento del modelo de proceso (AlQaheri & Panda, 2022).

- Algoritmo de Minado Difuso (Fuzzy Miner). Emplea los datos de los registros de eventos para construir un modelo que incluye nodos (representando actividades) y aristas (relaciones entre actividades), tomando en consideración tanto la relevancia relativa como el orden temporal de los eventos. Este algoritmo se basa en dos métricas principales: significancia y correlación (Günther *et. al.,* 2007). La significancia refleja la relevancia de la ocurrencia de los eventos y sus relaciones, donde los eventos más frecuentes se consideran más importantes. Por otro lado, la correlación se aplica únicamente a las aristas, mostrando cuán estrechamente relacionados están dos eventos consecutivos. En la etapa final, se aplican reglas para simplificar el modelo y decidir qué nodos y aristas incluir: los eventos con alta significancia se mantienen; aquellos con menor significancia, pero alta

correlación, también se incluyen; mientras que los eventos con baja significancia y baja correlación se abstraen. Es posible ajustar la simplificación del modelo modificando los parámetros, como los valores de corte. Asimismo, se puede utilizar el filtrado de aristas para dar mayor estructura al modelo (Caeiro *et.al*., 2018).

- Algoritmo de Minado Inductivo. Este algoritmo se distingue por su capacidad para generar modelos precisos y manejables, incluso cuando los datos contienen ruido o están incompletos, lo que lo hace más robusto frente a estos desafíos en comparación con otros algoritmos, como el algoritmo Alpha (Bogarín *et. al.,* 2018).

  Una de las características fundamentales del Minado Inductivo es la búsqueda entre equilibrio, precisión y simplicidad. Utiliza una estructura jerárquica que descompone los comportamientos complejos en componentes más simples, facilitando así la interpretación del modelo final.

  A lo largo de este proceso, el algoritmo construye lo que se conoce como un árbol de procesos, en el que cada nodo representa una actividad o conjunto de actividades, y a partir de este árbol se genera un modelo de proceso que captura las interacciones registradas. El proceso comienza con la división de los registros de eventos, agrupándolos en subconjuntos que representan diferentes comportamientos dentro del proceso. Estos subconjuntos se modelan como subprocesos utilizando construcciones como secuencias, paralelismos, elecciones y repeticiones (Van der Aalst, 2016). Posteriormente, estos subprocesos se combinan para crear el modelo final (un árbol de procesos), que también puede representarse mediante una red Petri o un modelo BPMN (Business Process Model and Notation).

Modelos de Representación en la MP

El resultado de la aplicación de los algoritmos mencionados anteriormente se visualiza a través de un modelo de procesos, dicho modelo varía en función del algoritmo utilizado, pues cada uno de ellos tiene una forma de representación, no obstante, los modelos más utilizados en los estudios analizados fueron:

- Red de Petri es uno de los lenguajes más antiguos para modelar procesos que permite representar la concurrencia de eventos. Este tipo de red está formada por transiciones, lugares y arcos dirigidos, donde los arcos conectan transiciones y lugares de forma bipartita. Las transiciones, representadas por recuadros, se asocian con tareas o acciones que pueden ejecutarse. Los lugares, indicados mediante círculos, pueden contener una o más fichas (tokens). Para que una transición se habilite, todos sus lugares de entrada deben tener al menos un token. Una vez habilitada, la transición se dispara, consumiendo un token de cada lugar de entrada y generando uno nuevo en cada lugar de salida (es decir, los lugares conectados a la transición a través de arcos salientes). Este proceso de disparo cambia el marcado de la red, lo que equivale a una modificación en el estado del proceso, definido por la distribución de tokens en los lugares. (Van der Aalst, 2016; Rozinat y Van der Aalst, 2008).

  Las Redes de Petri en la MP representan de manera gráfica el flujo y las dependencias entre las actividades, permitiendo a los analistas comprender y mejorar los procesos de manera más efectiva.

- Red causal. Es una representación gráfica que ilustra cómo las actividades o eventos dentro de un proceso se relacionan entre sí de manera secuencial y condicional. Estas redes están diseñadas para identificar la influencia que un evento puede tener sobre otro, proporcionando una comprensión visual de la lógica subyacente al proceso.

  En términos generales, una red causal está compuesta por nodos, que representan eventos o actividades, y arcos dirigidos, que muestran las relaciones de causa y efecto entre los nodos (Van der Aalst, 2016). Cuando un evento ocurre, este puede causar o influir en la ejecución de un evento posterior. Por ejemplo, en un sistema de compra en línea, la selección de un producto es una causa necesaria para poder agregarlo al carrito, y a su vez, agregar un producto al carrito es un prerrequisito para proceder con el pago.

  A medida que se avanza en la red causal, se pueden identificar dependencias clave entre las actividades. Esta estructura permite a los analistas descubrir patrones en los datos, como las secuencias más frecuentes de eventos, cuellos de botella en el proceso o incluso comportamientos inusuales. De esta forma, las redes causales son una herramienta

poderosa en la MP para modelar y analizar las interacciones dentro de sistemas complejos, facilitando la toma de decisiones y la mejora de procesos basados en datos reales.

- Mapa de proceso. Son representaciones visuales que describen las actividades y flujos dentro de un proceso empresarial o educativo. Sirven para identificar cómo se desarrollan las diferentes tareas, cómo se conectan entre sí y qué secuencia siguen (Hernández *et al.,* 2009). Estos mapas son útiles para comprender, analizar y optimizar procesos, ya que permiten visualizar la estructura y las relaciones entre las actividades, los puntos de decisión y los flujos de información. Son más sencillos de interpretar (Juhaňák *et.al*., 2019) En la MP, los mapas ayudan a identificar patrones y áreas de mejora.

Desarrollo

Método

La búsqueda se realizó en las bases de datos Science Direct, IEEE Xplore, EBSCO host y Research Gate, tomando como criterio de selección los estudios publicados en artículos científicos y actas de congresos entre el mes de abril de 2018 al mes de diciembre de 2023. Las palabras claves principales dentro de la búsqueda fueron: *process mining*, combinadas con *education* or *learning paths* or *learning analytics* or *learning management system.*

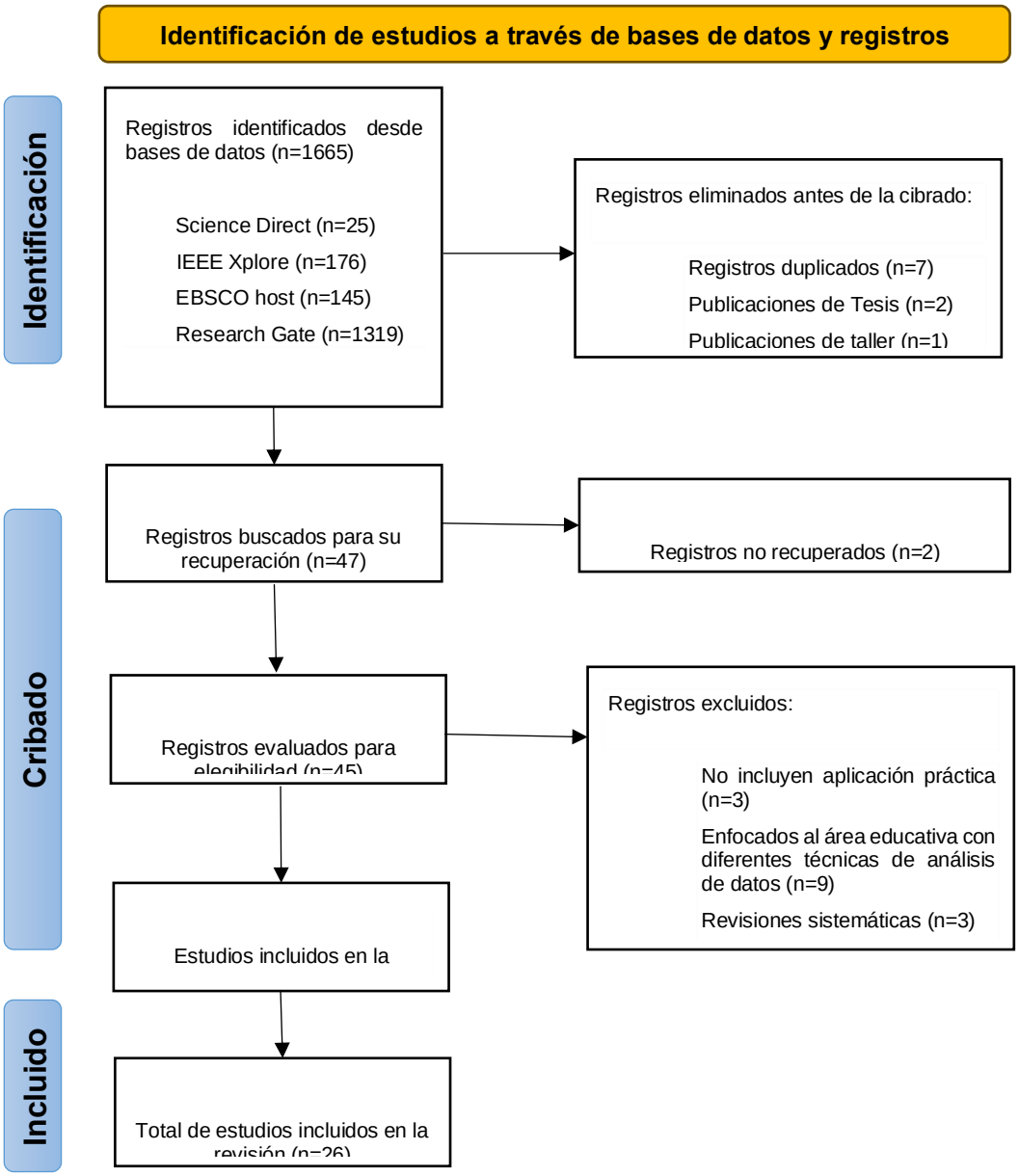
Criterios de Elegibilidad

Los estudios incluidos fueron de aplicación práctica e implementaciones de técnicas de MP en el ámbito educativo, redactados en inglés o español. La exclusión de algunas publicaciones se debió a que: 1) pertenecían a otra área del conocimiento; 2) eran revisiones sistemáticas; 3) se centraban solo en la parte teórica de la MPE; 4) los casos de estudio no estaban enfocados a procesos educativos de nivel universitario; y 5) si no eran publicaciones de revistas científicas o actas de congreso revisadas por pares.

Proceso de Selección de los Estudios

Para automatizar la selección, los autores de este estudio llevaron a cabo una revisión individual e independiente de los títulos, resúmenes y palabras clave del conjunto de publicaciones encontradas. Durante este proceso, se localizaron publicaciones que abordaban la MPE en la deserción estudiantil y como herramienta para detectar los flujos de aprendizaje de los estudiantes cuando utilizan plataformas en línea. Todos los artículos seleccionados fueron revisados a texto completo. Para tomar la decisión de excluir un artículo los autores tuvieron que exponer los motivos correspondientes a la deserción con el fin de llegar a un acuerdo. En la Figura 2 se puede observar el proceso de revisión y elección de las publicaciones encontradas, de acuerdo con la directrices y la lista de verificación de los criterios PRISMA (Preferred Reporting Items for Systematic Reviews, Page *et.al.*, 2021).

Figura 2. Diagrama de flujo de los estudios encontrados en la revisión sistemática.



Fuente. Elaboración propia a partir de Page *et al.*, (2021).

Con la finalidad de identificar la confiabilidad de los estudios, se consideró la herramienta de evaluación crítica de calidad para estudios transversales desarrollada por JBI (Moola *et al.*, 2020). En el proceso, se aplicaron seis preguntas clave con las siguientes opciones de respuesta: Sí (cumple con el criterio); No (no cumple con el criterio); Incierto (la información no es clara); y

No aplicable (el criterio no se aplica al estudio específico), por lo que, la evaluación de la calidad metodológica de cada estudio es cualitativa. Si un estudio recibe muchas respuestas afirmativas (Sí), se le considera metodológicamente sólido, mientras que un mayor número de respuestas No o Incierto puede señalar deficiencias en el diseño o la ejecución. Los estudios con respuestas mayormente afirmativas son clasificados como de calidad alta o moderada, dependiendo de la relevancia de los criterios cumplidos.

Las preguntas fueran revisadas por los dos investigadores, quienes concluyeron mantener los 26 artículos, debido a que todos fueron considerados con calidad suficiente como se puede observar en la Tabla 1.

Tabla 1. Calidad metodológica de los estudios incluidos en la revisión

Estudios	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**Calidad global**
Andreswari *et.al.,* (2022)	Si	No	Si	Si	No	Si	Moderada
Arpasat *et.al.,* (2021)	Si	Si	Si	Si	Si	Si	Alta
Aulia & Waspada, (2019)	Si	Si	Si	Si	Inciert o	Si	Alta
Baloch *et.al.,* (2022)	Si	Si	No	Si	No	Si	Moderada
Bey & Champagnat, (2022)	Si	Si	Si	Si	Si	Si	Alta
Bogarín *et.al.,* (2018)	Si	Si	Si	Si	Si	Si	Alta
Caeiro *et.al.,* (2018)	Si	Si	Si	Si	Si	Si	Alta
Cenka *et.al.,* (2022)	Si	Si	Si	Si	Si	Si	Alta
Cerezo *et.al.,* (2020)	Si	Si	Si	Si	Si	Si	Alta
Diamantini *et. al.,* (2023)	Si	Si	Si	Si	Si	Si	Alta
Etinger *et.al.,* (2018)	Si	Si	Si	Si	Inciert o	Si	Alta
Hachicha *et.al.,* (2021)	Si	Si	Si	Si	Inciert o	Si	Alta
Hakim *et.al.,* (2019)	Si	Si	Si	Si	Si	Si	Alta
Intayoad *et.al.,* (2018)	Si	Si	Si	Si	Si	Si	Alta

Ismail & Tumin, (2019)	Si	No	Si	Si	Inciert o	Si	Moderada
Juhaňák *et.al.,* (2019)	Si	Si	Si	Si	Si	Si	Alta
Macak *et.al.,* (2021)	Si	Si	Si	Si	Si	Si	Alta
Moreno *et.al.,* (2022)	Si	No	Si	Si	No	Si	Moderada
Nai *et.al.,* (2023)	Si	No	Si	Si	No	Si	Moderada
Nammakhunt *et.al.,* (2023)	Si	Si	Si	Si	Si	Si	Alta
Real *et.al.,* (2020)	Si	Si	Si	Si	Inciert o	Si	Alta
Rodríguez *et al.,* (2023)	Si	No	Si	SI	Si	SI	Moderada
Taub *et.al.,* (2022)	Si	Si	Si	Si	Si	Si	Alta
Thiyagarajan & Prasanna, (2021)	Si	Si	Si	Si	Si	Si	Alta
Wisudiawan & Kurniati, (2022)	Si	No	Si	Si	No	Si	Moderada
Zambrano & Pin, (2023)	Si	No	Si	Si	No	Si	Moderado

Notas. P1= ¿Se incluyó una muestra (participantes o datos) clara y bien definida?; *P2=* ¿Se describe en detalle los sujetos del estudio (participantes o datos) y el entorno?; *P3=* ¿Las exposiciones y resultados fueron medidos de manera confiable?; *P4=* ¿Las variables de interés fueron identificadas claramente?; *P5=* ¿Se ajustó el análisis a posibles factores de confusión?; *P6=* ¿Se utilizaron técnicas de MP apropiados para el análisis?

Resultados

Características de los Estudios

En esta revisión se identificaron 26 estudios, de los cuales, 14 fueron estudios publicados en revistas (53.8%) y 12 publicados en actas de congreso (46.2%).

Once estudios se realizaron en Europa (42.3%; Caeiro *et al.,* 2018; Bogarín *et.al.,* 2018; Etinger *et.al.,* 2018; Cerezo *et.al.*, 2020; Juhaňák *et.al.*, 2019; Hachicha *et.al.,* 2021; Macak *et.al.*, 2021; Moreno *et.al.,* 2022; Bey y Champagnat, 2022; Nai *et.al.,* 2023; Diamantini *et.al.,* 2023), once más se efectuaron en Asia (42.3% ; Intayoad et.al., 2018; Aulia y Waspada, 2019; Hakim

et.al., 2019; Ismail y Tumin, 2019; Arpasat *et.al.*, 2021; Thiyagarajan y Prasanna, 2021; Wisudiawan y Kurniati, 2022; Baloch *et. al*., 2022; Cenka *et.al.*, 2022; Andreswari *et.al.*, 2022; Nammakhunt *et.al.*, 2023), tres en América Latina (11.5%; Rodríguez *et.al*., 2023; Real *et.al.*, 2020; Zambrano y Pin, 2023) y solo uno en América del Norte (3.8%; Taub *et.al*., 2022)

Los casos expuestos en las 26 investigaciones se centran en el estudio de diversos procesos educativos que se llevan a cabo a nivel universitario. Su distribución por año se detalla en la Tabla 2. El año con mayor cantidad de estudios fue 2022 (26.9%) en comparación con el año 2020 en donde se encontraron dos estudios (7.7%).

Tabla 2. Distribución de los artículos publicados entre mes de abril de 2018 al mes de noviembre de 2023.

Artículos publicados por año		
Año	*No. Artículos*	*Porcentaje*
2018	4	15.4%
2019	4	15.4%
2020	2	7.7%
2021	4	15.4%
2022	7	26.9%
2023	5	19.2%

Fuente. Elaboración propia.

Entornos de Aprendizaje Virtuales Utilizados

Los datos que conforman los registros de eventos utilizados en las investigaciones analizadas fueron generados por estudiantes al interactuar en los cursos o procesos educativos, los cuales fueron impartidos a través de diversos entornos de aprendizaje virtuales. Dieciocho de estos estudios (69.23%) utilizaron Moodle uno de los sistemas de gestión del aprendizaje más utilizado debido a su naturaleza de código abierto. Siete estudios utilizaron entornos virtuales desarrollados internamente (26.92%; Intayoad *et. al*., 2018; Baloch *et. al.*, 2022; Bey y Champagnat, 2022; Taub

et.al., 2022; Nai *et.al.,* 2023; Zambrano y Pin, 2023; Diamantini *et.al.,* 2023), y solo un estudio (3.85%) utilizó la herramienta de almacenamiento GIT (Macak *et.al.,* 2021).

Técnicas de MPE Utilizadas

Veintiséis estudios emplearon la técnica de descubrimiento de la MPE (100%), de estos, tres incorporaron también el uso de la técnica de conformidad (11.5%; Arpasat *et.al.,* 2021; Macak *et.al.,* 2021; Wisudiawan y Kurniati, 2022), y un estudio más la técnica de rendimiento (3.9%; Bogarín *et.al.,* 2018).

Algoritmos de MPE Utilizados

Veinte estudios (76.92%) emplearon la minería heurística como base, cuatro de ellos incluyeron como complemento el uso de la minería alfa, difusa o inductiva (Hachicha *et.al*., 2021; Thiyagarajan y Prasanna, 2021; Cenka *et.al*., 2022; Baloch *et al*., 2022). Por otro lado, la minería difusa fue usada como algoritmo principal en tres estudios (11.54%; Arpasat *et.al.,* 2021; Macak *et.al*., 2021; Zambrano y Pin, 2023) y en uno de ellos, también utilizaron otros algoritmos de MP para analizar las estadísticas descriptivas de los casos y la relación existente entre las personas involucradas en el proceso descubierto (Macak *et.al.,* 2021). Además, la minería inductiva fue utilizada en 2 estudios (7.69%; Bogarín *et.al.,* 2018; Cerezo et al., 2020). Finalmente, en la investigación de Ismail y Tumin (2019), no se especifica qué tipo de minería fue empleada para desarrollar su caso de estudio (3.85%).

Modelos de Procesos Utilizados

De los estudios analizados, veinticuatro emplearon al menos un tipo de modelo de proceso para representar los hallazgos encontrados durante el análisis. El modelo que tuvo mayor aparición fue el mapa de procesos (46%), seguido de las redes causales (38%; Intayoad *et.al.,* 2018; Aulia y Waspada, 2019; Caeiro *et.al.*, 2018; Real *et.al.,* 2020; Hachicha et.al., 2021; Bey y Champagnat, 2022; Moreno *et.al.,* 2022; Taub *et.al.,* 2022; Nai *et.al.,* 2023; Diamantini *et.al*., 2023) y las redes Petri (8%; Cerezo *et.al.,* 2020; Baloch *et. al.,* 2022). Mientras que, los otros dos estudios ocuparon

más de un modelo para representar los procesos descubiertos (8%; Arpasat *et.al*., 2021; Thiyagarajan y Prasanna, 2021).

Herramienta de *Software* Utilizadas

La herramienta informática de MPE que presenta mayor dominio es ProM, un *software* de código abierto que incluye una gran cantidad de opciones técnicas y permite visualizar los procesos en múltiples modelos. Por ejemplo, es posible convertir una red Petri a una red heurística. De los diecisiete estudios (65,5%) que utilizaron esta herramienta, ocho estudios incluyeron como herramienta complementaria Disco fluxicon o Prodigen.

Cuatro estudios (15.4%; Etinger et.al., 2018; Juhaňák et.al., 2019; Cenka *et.al.,* 2022; Rodríguez *et.al.,* 2023) recurrieron a Disco fluxicon, una herramienta comercial que solo permite crear mapas de procesos, pero de exploración y resultados muy rápidos (Günther & Rozinat, 2012).

En tanto que, en el estudio de Moreno et.al., 2022 se empleó Celonis (3.8%), un *software* comercial orientado a las empresas que permite utilizar la MP para modelar procesos tomando como base, los datos que son generados dentro de los sistemas de información que gestionan los procesos de negocio. Otro estudio (3.8%; Hachicha *et.al.,* 2021) utilizó PM4y, una biblioteca de Python que proporciona algoritmos básicos para el descubrimiento de procesos. En tanto que, en el estudio de Nai *et.al.,* 2023 se empleó Apromore un repositorio avanzado de modelos de procesos que congrega características de análisis, gestión y uso para grandes modelos de procesos (3.8%; La Rosa *et.al*., 2011).

Solo dos estudios (7.7%; Bey y Champagnat, 2022; Taub *et.al.,* 2022) no mencionan el *software* de MP que utilizaron dentro de sus investigaciones.

Principales Contribuciones

En la Tabla 3, se muestra una vista general sobre las contribuciones principales de los 26 estudios, de las cuales se puede señalar que la MPE se ha empleado para: a) reconocer comportamientos recurrentes (patrones) de los estudiantes al utilizar sistemas de aprendizaje y con ello proponer intervenciones pedagógicas con el objetivo de mejorar la satisfacción de estudiantes y profesores; b) mejorar la usabilidad de los sistemas de aprendizaje en función de los perfiles de los usuarios;

c) facilitar la identificación de rutas de aprendizaje efectivas; y d) analizar las relaciones entre los estudiantes para obtener mejores resultados.

No obstante, persisten desafíos relacionados con la calidad de los datos utilizados. Los datos ruidosos y la falta de consideración de factores importantes, como la captura de todos los aspectos del comportamiento estudiantil y los conocimientos preexistentes que los estudiantes deben poseer al utilizar los entornos virtuales pueden limitar la precisión y validez de los resultados. Además, en algunos casos, la necesidad de tener conocimientos técnicos avanzados para utilizar las herramientas de MPE pueden representan barreras importantes.

A pesar de estas limitaciones, el interés por avanzar en el campo de la MPE a nivel universitario es evidente. Se sugiere la exploración de nuevas técnicas, métodos más robustos, el análisis temporal más detallado y la aplicación práctica en diferentes contextos educativos para mejorar la comprensión y optimización de los procesos de aprendizaje que involucran el uso de entornos virtuales.

Tabla 3. Vista general sobre las principales contribuciones de los 26 estudios.

Contribuciones principales	**Estudios**
Utilización de la MPE a nivel universitario La MPE fue utilizada para analizar y optimizar procesos educativos implementados en diversos sistemas de gestión del aprendizaje a nivel universitario. Estos análisis abarcaron desde la identificación de patrones hasta la formulación de recomendaciones de intervención pedagógica.	Caeiro et.al., (2018), Bogarín et.al., (2018), Etinger et.al., (2018), Intayoad et.al., (2018), Cerezo et.al., (2020), Aulia & Waspada, (2019), Hakim et.al., (2019), Ismail & Tumin, (2019), Juhaňák *et.al.*, (2019), Real *et.al.*, (2020), Arpasat *et.al.*, (2021), Hachicha *et.al.*, (2021), Macak *et.al.*, (2021), Moreno *et.al.*, (2022), Thiyagarajan & Prasanna, (2021), Wisudiawan & Kurniati, (2022), Baloch *et.al.*, (2022), Bey & Champagnat, (2022), Cenka *et.al.*, (2022), Andreswari *et.al.*, (2022), Taub *et.al.*, (2022), Nai *et.al.*, (2023), Nammakhunt *et.al.*, (2023),

	Rodríguez *et.* al., (2023), Zambrano & Pin, (2023) y Diamantini *et.al.*, (2023).
Identificación y análisis de comportamientos, patrones, rutas y desviaciones en el ámbito educativo Se destaca el potencial de la MPE para transformar el análisis educativo, ofreciendo insights profundos que pueden orientar el progreso continuo de los procesos de enseñanza y aprendizaje.	Caeiro *et. al*., (2018), Etinger *et. al.*, (2018), Intayoad *et. al.*, (2018), Cerezo *et. al*., (2020), Juhaňák *et. al*., (2019), Real *et. al.*, (2020), Arpasat *et. al*., (2021), Thiyagarajan & Prasanna, (2021), Bey & Champagnat, (2022), Cenka *et. al*., (2022), Moreno *et.al*., (2022), Taub et. al., (2022), Nai et.al., (2023), NammaKhunt *et al*., (2023), Rodríguez *et. al.*, (2023) y Diamantini *et. al*., (2023).
Propuesta o implementación de mejora en los procesos descubiertos El uso de la MPE ha mejorado la usabilidad de los entornos de aprendizaje virtuales. Los análisis realizados han permitido la creación de entornos más eficientes, facilitando a los docentes rediseñar cursos de manera informada y ofrecer recomendaciones basadas en evidencia para la planificación educativa.	Etinger *et.al.*, (2018), Aulia & Waspada, (2019) Real *et.al*., (2020), Arpasat *et.al.*, (2021), Macak et.al., (2021), Wisudiawan & Kurniati, (2022), Andreswari *et*.al., (2022), Rodríguez *et al*., (2023), Zambrano & Pin, (2023) y Diamantini *et.al.*, (2023).
Satisfacción resultante de los autores al utilizar la MPE en los procesos educativos Se reportan diversos logros, como la validación de modelos educativos, mejora en la usabilidad de los sistemas de gestión del aprendizaje, la identificación de patrones de comportamiento y rutas de aprendizaje, así como la capacidad de hacer recomendaciones informadas para mejorar el diseño de cursos y el proceso educativo en general.	Caeiro *et.al.*, (2018), Bogarín *et.al*., (2018), Etinger *et.al*., (2018), Cerezo *et.al.*, (2020), Aulia & Waspada, (2019), Hakim *et.al.*, (2019), Ismail & Tumin, (2019), Juhaňák *et.al*., (2019), Real *et.al*., (2020), Arpasat *et.al.*, (2021), Thiyagarajan & Prasanna, (2021), Bey & Champagnat, (2022), Cenka *et.al*., (2022), Moreno *et. al.*, (2022), Taub *et.al.*, (2022), Nai et.al.,

	(2023), Rodríguez *et.al.*, (2023) y Diamantini *et.al.*, (2023).
Datos no robustos Aunque la MPE es una herramienta poderosa para analizar y mejorar los procesos educativos, existe una preocupación compartida sobre la robustez y la integridad de los datos utilizados. Los datos ruidosos, la falta de consideración de factores importantes como los conocimientos precedentes que deben de poseer los estudiantes antes de tomar algún curso. Aunado a la necesidad de contar con registros de eventos más amplios y diversos, son desafíos que podrían estar limitando la precisión y la validez de los resultados obtenidos.	Intayoad *et.al.*, (2018), Real *et.al.*, (2020), Bey & Champagnat, (2022) y Nammakhunt *et.al.*, (2023).
Limitaciones al usar la MPE Dificultad para capturar todos los aspectos del comportamiento estudiantil y la dinámica de aprendizaje, presencia de datos ruidosos generados por la complejidad del aprendizaje, necesidad de conocimientos técnicos avanzados para emplear las herramientas de software de la MPE, y la utilización de un conjunto limitado de técnicas de MPE también restringen la efectividad y exactitud de los resultados obtenidos.	Caeiro *et. al.*, (2018), Intayoad *et. al.*, (2018), Cerezo *et. al.*, (2020), Real *et. al.*, (2018), Juhaňák *et. al.*, (2019) y NammaKhunt *et. al.*, (2023).
Perspectivas futuras de la MPE Existe el interés de avanzar en el campo de la MPE mediante la exploración de nuevas técnicas, métodos más robustos, el análisis temporal más detallado y la aplicación práctica en otros contextos educativos para mejorar la	Caeiro *et. al.*, (2018), Intayoad *et. al.*, (2018), Hachicha *et. al.*, (2021), Wisudiawan & Kurniati, (2022), Cerezo *et. al.*, (2020), Real et. al., (2020), Taub *et. al.*, (2022), Thiyagarajan & Prasanna, (2021), Moreno *et. al.*, (2022) y Nai *et. al.*, (2023).

comprensión y optimización de los procesos de aprendizaje.	

Fuente. Elaboración propia.

Cierre

En la revisión sistemática se encontró que la investigación sobre la utilización de la MPE a nivel universitario ha crecido en diversas partes del mundo, principalmente en Europa y Asia. Sin embargo, en América del Norte y América Latina, la investigación es escasa, a pesar de que la MPE ha sido utilizada para mejorar diversos aspectos de los procesos educativos cuando se involucran entornos educativos virtuales.

Los datos utilizados en los estudios se obtuvieron mayoritariamente de Moodle, un sistema de gestión de aprendizaje ampliamente utilizado por su carácter de código libre, mientras que otros estudios emplearon entornos de aprendizaje virtuales desarrollados internamente.

Todos los estudios localizados en esta revisión recurrieron al uso de la técnica de descubrimiento de la MPE y muy pocos incorporaron también la técnica de conformidad y rendimiento. La minería heurística fue el algoritmo comúnmente empleado y algunos estudios complementaron su uso con la minería alfa, difusa o inductiva. Obteniendo mapas de procesos como modelos dominantes seguidos de las redes causales.

ProM fue la herramienta de MP más empleada en los estudios y algunos complementaron su uso con Disco fluxicon o Prodigen. Otras herramientas usadas, pero en menor proporción fueron PM4y, Celonis y Apromore.

La MPE demuestra un considerable potencial para identificar patrones de comportamiento repetitivos en los estudiantes al utilizar entornos de aprendizaje virtuales, lo que ha permitido proponer intervenciones pedagógicas orientadas a mejorar la satisfacción tanto de los estudiantes como de los profesores. Además, contribuye a optimizar la usabilidad de los sistemas de aprendizaje adaptándolos a los perfiles de los usuarios, facilita la detección de rutas de aprendizaje más efectivas y permite analizar las interacciones entre los estudiantes para obtener mejores resultados académicos.

No obstante, siguen existiendo desafíos importantes relacionados con la solidez de los datos empleados, como la presencia de información ruidosa y la falta de consideración de factores cruciales, como la captura completa del comportamiento estudiantil y los conocimientos previos que los alumnos deben tener al usar los entornos virtuales. Estas limitaciones pueden afectar tanto la precisión como la validez de los resultados. Además, en ciertos casos, la necesidad de poseer conocimientos técnicos avanzados para manejar las herramientas de MPE puede representar obstáculos significativos.

A pesar de estas limitaciones, es claro que existe un creciente interés en impulsar el campo de la MPE en el ámbito universitario. Es recomendable seguir explorando nuevas técnicas, desarrollar métodos más sólidos, realizar análisis temporales más precisos y seguir aplicando las MPE en diversos contextos educativos para mejorar la comprensión y optimización de los procesos de aprendizaje que involucran el uso de entornos virtuales.

Referencias

Aguirre, M. H., & Rincón, G. N. (2015). Minería de procesos: desarrollo, aplicaciones y factores críticos. *Cuadernos de administración*, *28*(50), 137-157. https://doi.org/10.11144/Javeriana.cao28-50.mpda

AlQaheri, H., Panda, M., (2022). Education Process Mining Framework: Unveiling Meaningful Information for Understanding Students' Learning Behavior and Improving Teaching Quality. *Information 13*(29), 1-19. http://dx.doi.org/10.3390/info13010029

Andreswari, R., Fauzi, R., Valensia, L., & Chanifah, S. (2022). Conformance Analysis of Student Activities to Evaluate Implementation of Outcome-Based Education in Early of Pandemic using Process Mining. *SHS Web of Conferences, Japón*, *139*, 03018. https://doi.org/10.1051/shsconf/202213903018

Arpasat, P., Premchaiswadi, N., Porouhan, P., & Premchaiswadi, W. (2021). Applying Process Mining to Analyze the Behavior of Learners in Online Courses. *International Journal of Information and Education Technology*, *11*(10), 436–443. https://doi.org/10.18178/ijiet.2021.11.10.1547

Aulia, D., & Waspada, I. (2019). The Design of Exploratory Application and Preprocessing of Event Log Data in LMS Moodle-Based Online Learning Activities for Process Mining. *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika*, *5*(2), 124–133. https://doi.org/10.23917/khif.v5i2.8023

Baloch, F., Bhatti, S., & Dewani, A. (2022). Application of Process Mining in Education: A Case Study. *Acta Scientific Computer Sciences,* *4*(7), 03-11. https://actascientific.com/ASCS/pdf/ASCS-04-0287.pdf

Bey, A., & Champagnat, R. (2022). Analyzing Student Programming Paths using Clustering and Process Mining. *Proceedings of the 14th International Conference on Computer Supported Education*, *2*, 76–84. https://doi.org/10.5220/0011077300003182

Bogarín, A., Cerezo, R., & Romero, C. (2017). A survey on educational process mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(1). https://doi.org/10.1002/widm.1230

Bogarín, A., Cerezo, R., & Romero, C. (2018). Discovering learning processes using Inductive Miner: A case study with Learning Management Systems (LMSs). *Psicothema*, *30*(3), 322–329. https://doi.org/10.7334/psicothema2018.116

Caeiro, M., Llamas, M., Mikic, F. A., Lama, M., & Mucientes, M. (2018). Exploring the application of process mining to support self-regulated learning: An initial analysis with video lectures. *2018 IEEE Global Engineering Education Conference (EDUCON)*, *España,* 1766–1774. https://doi.org/10.1109/EDUCON.2018.8363448

Carnis, A. H., Gueni, B., Fhima, M., Carnis, A., David, S., & Khekifa, N. (2015). Process Mining in the Education Domain. *International Journal on Advances in Intelligent Systems. 8*(2), 219-232. http://www.iariajournals.org/intelligent_systems/intsys_v8_n12_2015_paged.pdf

Cenka, B. A., Santoso, H. B., & Junus, K. (2022). Analysing student behaviour in a learning management system using a process mining approach. *Knowledge Management & E-Learning: An International Journal*, *14*(1), 62–80. https://doi.org/10.34105/j.kmel.2022.14.005

Cerezo, R., Bogarín, A., Esteban, M., & Romero, C. (2020). Process mining for self-regulated learning assessment in e-learning. *Journal of Computing in Higher Education*, *32*(1), 74–88. https://doi.org/10.1007/s12528-019-09225-y

Diamantini, C., Genga, L., Mircoli, A., Potena, D., & Zannone, N. (2023). Understanding the stumbling blocks of Italian higher education system: A process mining approach. *Expert Systems with Applications*, *242*, 122747. https://doi.org/10.1016/j.eswa.2023.122747

Etinger, D., Orehovački, T., & Babić, S. (2018). Applying Process Mining Techniques to Learning Management Systems for Educational Process Model Discovery and Analysis. En W. Karwowski & T. Ahram (Eds.), *Intelligent Human Systems Integration* (Vol. 722, pp. 420–425). Springer, Cham. https://doi.org/10.1007/978-3-319-73888-8_65

Ghazal, M. A., Ibrahim, O.A., & Salama, M. A. (2017). Educational Process Mining: A Systematic Literature Review. *European Conference on Electrical Engineering and Computer Science (EECS), Bern,* 198-203. https://doi.org/10.1109/EECS.2017.45

Günther, C. W., & Rozinat, A. (2012). Disco: discover your processes. *Proceedings of the Demonstration Track of the 10th International Conference on Business Process Management (BPM 2012), Estonia,* 940, 40-44. https://pure.tue.nl/ws/portalfiles/portal/129682565/paper8.pdf

Günther, C. W., & Van der Aalst, W. M. P. (2007). Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics. En G. Alonso, P. Dadam, & M. Rosemann (Eds.), Business Process Management (Vol. 4714, pp. 328–343). Springer Berlín Heidelberg. https://doi.org/10.1007/978-3-540-75183-0_24

Hachicha, W., Ghorbel, L., Champagnat, R., Zayani, C. A., & Amous, I. (2021). Using Process Mining for Learning Resource Recommendation: A Moodle Case Study. *Procedia Computer Science*, 192, 853–862. https://doi.org/10.1016/j.procs.2021.08.088

Hakim, A. R., Hasibuan, M. A., & Andreswari, R. (2019). E-learning process analysis to determining student learning patterns using process mining approach. *International Conference of Computer and Informatics Engineering, Indonesia, 1193*, 012020. https://doi.org/10.1088/1742-6596/1193/1/012020

Hernández, N. A., Medina, L. A., & Nogueira, R. D. (2009). Criterios para la elaboración de mapas de procesos. Particularidades para los servicios hospitalarios. *Ingeniería Industrial, XXX*(2), 1-7. https://www.redalyc.org/articulo.oa?id=360433569002

Intayoad, W., Kamyod, C., & Temdee, P. (2018). Process mining application for discovering student learning paths. *International Conference on Digital Arts, Media and Technology (ICDAMT 2018)*, *Tailandia*, 220–224. https://doi.org/10.1109/ICDAMT.2018.8376527

Ismail, S., & Tumin, F. (2019). Analysis on Online Learning Environment using Process Mining Technique for Personal Knowledge Management Mapping. *6th International Conference on Research and Innovation in Information Systems (ICRIIS)*, *Malasia*, 1–6. https://doi.org/10.1109/ICRIIS48246.2019.9073269

Juhaňák, L., Zounek, J., & Rohlíková, L. (2019). Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. *Computers in Human Behavior*, *92*, 496–506. https://doi.org/10.1016/j.chb.2017.12.015

La Rosa, M., Reijers, H. A., Van der Aalst, W. M. P., Dijkman, R. M., Mendling, J., Dumas, M., & García-Bañuelos, L. (2011). APROMORE: An advanced process model repository. *Expert Systems with Applications*, *38*(6), 7029–7040. https://doi.org/10.1016/j.eswa.2010.12.012

Macak, M., Kruzelova, D., Chren, S., & Buhnova, B. (2021). Using process mining for Git log analysis of projects in a software development course. *Education and Information Technologies*, *26*(5), 5939–5969. https://doi.org/10.1007/s10639-021-10564-6

Moola, S., Munn, Z., Tufanaru, C., Aromataris, E., Sears, K., Sfetcu, R., Currie, M., Qureshi, R., Mattis, P., Lisy, K., & Mu, P. (2020).Systematic reviews of etiology and risk. En E. Aromataris & Z. Munn (Ed.), JBI Manual for Evidence Synthesis. JBI. https://synthesismanual.jbi.global

Moreno, M., Exposito, E., & Gueye, M. (2022). Process Mining Model to visualize and analyze the Learning Process. *9th Research in Engineering Education Symposium and 32nd Australasian Association for Engineering Education Conference, Australia*, 754–762. https://doi.org/10.52202/066488-0083

Nai, R., Sulis, E., Marengo, E., Vinai, M., & Capecchi, S. (2023). Process Mining on Students' Web Learning Traces: A Case Study with an Ethnographic Analysis. En O. Viberg, I. Jivet, P. J. Muñoz-Merino, M. Perifanou & T. Papathoma (Eds.), *Responsive and Sustainable*

Educational Futures (Vol. 14200, pp. 599–604). Springer, Cham. https://doi.org/10.1007/978-3-031-42682-7_48

Nammakhunt, A., Porouhan, P., & Premchaiswadi, W. (2023). Creating and Collecting e-Learning Event Logs to Analyze Learning Behavior of Students through Process Mining. *International Journal of Information and Education Technology*, *13*(2), 211–222. https://doi.org/10.18178/ijiet.2023.13.2.1798

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*(71) https://doi.org/10.1136/bmj.n71

Ramírez, G. J. (2021). Estudio de la experiencia de usuario en los sistemas de gestión del aprendizaje. *IE Revista de Investigación Educativa de la REDIECH*, *12*, e1358. https://doi.org/10.33010/ie_rie_rediech.v12i0.1358

Real, E. M., Pinheiro, P. E., De Oliveira, L. V., Braga, C. J., & Stiubiener, I. (2020). Educational Process Mining for Verifying Student Learning Paths in an Introductory Programming Course. *IEEE Frontiers in Education Conference (FIE)*, *Suecia*, 1–9. https://doi.org/10.1109/FIE44824.2020.9274125

Rodríguez, G. A., Granada, D., & Vara, J. M. (2023). Analizando con minería de procesos la actividad virtual de los estudiantes durante la pandemia covid-19 en Honduras. *Universidad y Sociedad,* *15*(4), 647-660. https://rus.ucf.edu.cu/index.php/rus/article/view/4022

Rozinat, A., & Van der Aalst, W. (2008). Conformance Checking of Processes Based on Monitoring Real Behavior. *Information Systems,* *33*(1), 64-95. https://doi.org/10.1016/j.is.2007.07.001

Sypsas, A., & Kalles, D. (2022). Reviewing Process Mining Applications and Techniques in Education. *International Journal of Artificial Intelligence & Applications*, *13*(1), 83–102. https://doi.org/10.5121/ijaia.2022.13106

Taub, M., Banzon, A. M., Zhang, T., & Chen, Z. (2022). Tracking Changes in Students' Online Self-Regulated Learning Behaviors and Achievement Goals Using Trace Clustering and Process Mining. *Frontiers in Psychology*, *13*, 813514. https://doi.org/10.3389/fpsyg.2022.813514

Thiyagarajan, G., & Prasanna, S. (2021). A Process Mining approach to analyze learning behavior in the flipped classroom. *2nd International Conference on Communication, Computing and Industry 4.0 (C2I4)*, *India*, 1–7. https://doi.org/10.1109/C2I454156.2021.9689228

Turnbull, D., Chugh, R., & Luck, J. (2020). Learning Management Systems, An Overview. En A. Tatnall (Eds.), *Encyclopedia of Education and Information Technologies* (pp. 1052–1058). Springer, Cham. https://doi.org/10.1007/978-3-030-10576-1_248

Umer, R., Susnjak, T., Mathrani, A., & Suriadi, S. (2022). Data quality challenges in educational process mining: Building process-oriented event logs from process-unaware online learning systems. *International Journal of Business Information Systems*, *39*(4), 569. https://doi.org/10.1504/IJBIS.2022.122877

Van der Aalst, W. (2016). Process Mining. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-49851-4

Van der Aalst, W., Weijters, T., & Maruster, L. (2004).Workflow mining: discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering, 16*(9), 1128–1142. https://doi.org/10.1109/TKDE.2004.47

Wisudiawan, G. A. A., & Kurniati, A. P. (2022). Process Mining on Learning Activities in a Learning Management System. *24th International Conference on Advanced*

Communication Technology (ICACT), *Corea del Sur*, 476–482. https://doi.org/10.23919/ICACT53585.2022.9728903

Zambrano, Z., & Pin, G. L. (2023). Aplicación de minería de procesos para el descubrimiento automático del Workflow de la gestión de titulación. Revista Científica Arbitrada Multidisciplinaria *PENTACIENCIAS,* *5*(5), 22–41. https://doi.org/10.59169/pentaciencias.v5i5.718

DETECCIÓN DE ERRORES EN SISTEMAS CUÁNTICOS TOPOLÓGICOS USANDO *MACHINE LEARNING*

ERROR DETECTION IN TOPOLOGICAL QUANTUM SYSTEMS USING MACHINE LEARNING

José Roman Castro San Agustín[1]
jcastros248@alumno.uaemex.mx
ORCID: 0000-0001-6383-6700

Marco Alberto Mendoza Pérez[2]
mamendozap@uaemex.mx
ORCID: 0000-0003-4911-4757

Manuel Ávila Aoki[3]
vlkmanuel@uaemex.mx
ORCID: 0000-0002-3951-6421

Doricela Gutiérrez Cruz[4]
dgutierrezcr@uaemex.mx
ORCID: 0000-0003-2843-3273

[1] Doctorado en Ciencias de la Computación, Centro Universitario UAEM Valle de Chalco, Universidad Autónoma del Estado de México.
[2] Doctorado en Ciencias de la Computación, Centro Universitario UAEM Valle de Chalco, Universidad Autónoma del Estado de México.
[3] Doctorado en Ciencias de la Computación, Centro Universitario UAEM Valle de Chalco, Universidad Autónoma del Estado de México.
[4] Doctorado en Ciencias de la Computación, Centro Universitario UAEM Nezahualcóyotl, Universidad Autónoma del Estado de México.

Resumen

La computación cuántica, que utiliza *qubits* para el procesamiento de la información, enfrenta desafíos constantes debido a la fragilidad y susceptibilidad a errores de los sistemas cuánticos. La computación cuántica topológica, que utiliza *qubits* de Majorana, ofrece una vía interesante para construir computadoras cuánticas estables. La detección de coherencia en estos sistemas puede beneficiarse del análisis de datos mediante redes neuronales, facilitando la identificación de factores que contribuyen a la pérdida de coherencia. La decoherencia y la amplificación de errores son problemas clave que afectan la precisión de los cálculos cuánticos. Para abordar estos inconvenientes, se desarrollan técnicas avanzadas de corrección y detección de errores. El aprendizaje profundo emerge como una herramienta prometedora para mejorar la computación cuántica. Técnicas como las redes neuronales convolucionales y los *autoencoders* podrían identificar y clasificar errores en la información. En particular, las redes neuronales pueden detectar patrones extraños en los datos generados por experimentos cuánticos, lo que permite implementar correcciones más efectivas. El diseño de redes neuronales para la detección de errores implica un proceso de entrenamiento utilizando datos de estados cuánticos con errores. Estas redes pueden ser ajustadas para mejorar su capacidad de detección y validación de los estados cuánticos. El presente artículo muestra la importancia de la implementación de estas técnicas avanzadas para asegurar la fidelidad y precisión de los cálculos de computadoras cuánticas estables.

Palabras clave: Computación cuántica, *Qubits* de Majorana, Decoherencia, Redes neuronales, Corrección de errores.

Abstract

Quantum computing, which utilizes qubits for information processing, continually faces challenges due to the fragility and susceptibility of quantum systems to errors. Topological quantum computing, which employs Majorana qubits, offers a promising approach to building stable quantum computers. Coherence detection in these systems can benefit from data analysis using neural networks, facilitating the identification of factors that contribute to coherence loss. Decoherence and error amplification are key issues that affect the accuracy of quantum calculations. To address these challenges, advanced error correction and detection techniques are

being developed. Deep learning emerges as a promising tool to enhance quantum computing. Techniques such as convolutional neural networks and autoencoders can identify and classify errors in the information. Neural networks, in particular, can detect anomalies in data generated by quantum experiments, enabling more effective error corrections. Designing neural networks for error detection involves a training process using data from quantum states with errors. These networks can be fine-tuned to improve their capacity for detecting and validating quantum states. This article highlights the importance of implementing these advanced techniques to ensure the fidelity and accuracy of calculations in stable quantum computers.

Keywords: Quantum computing, Majorana qubits, Decoherence, Neural networks, Error correction

Fecha de envío: 06/06/2024
Fecha de aprobación: 10/11/2024
Fecha de publicación: 01/01/2025

Introducción

La informática cuántica es una rama de la informática que utiliza la teoría cuántica para procesar y transmitir información. En la informática cuántica, la unidad básica de información es el *qubit*, que es un sistema cuántico de dos estados. A diferencia de los bits clásicos que solo pueden estar en dos estados (0 o 1), los *qubits* pueden estar en una superposición de ambos estados al mismo tiempo, lo que les permite procesar información de manera mucho más eficiente.

Los algoritmos cuánticos son algoritmos diseñados para ser ejecutados en una computadora cuántica. La principal diferencia que tienen con los algoritmos clásicos, es que se aprovechan las propiedades cuánticas para resolver problemas de manera mucho más eficiente (Nielsen y Chuang, 2010). Ejemplos de algoritmos cuánticos incluyen el algoritmo de Grover para la búsqueda no estructurada y el algoritmo de Shor para la factorización de números enteros grandes, que es de gran importancia en criptografía (De Wolf, 2023).

La computación cuántica es una tecnología emergente que promete revolucionar muchos campos de la ciencia y la tecnología al ofrecer la posibilidad de realizar cálculos a velocidades que los ordenadores clásicos no pueden alcanzar. Sin embargo, los sistemas cuánticos son frágiles y susceptibles a errores debido a la naturaleza misma de la mecánica cuántica. Esto representa un gran desafío para la construcción de sistemas cuánticos escalables y fiables.

Afortunadamente, las técnicas de aprendizaje profundo, un subcampo de la inteligencia artificial, están emergiendo como una herramienta valiosa para detectar y reducir errores en sistemas cuánticos. Estas técnicas se han utilizado para mejorar la tolerancia a errores en sistemas cuánticos, lo que los hace más estables y fiables, lo cual es fundamental para el éxito de la computación cuántica en el futuro (Deng y Yu, 2014).

El campo de la inteligencia artificial ha presentado algoritmos de aprendizaje automático basado en una red neuronal recurrente para la corrección continua de errores cuánticos en sistemas superconductores de tres *qubits transmon* (Convy, Liao, Zhang, Patel, Livingston, Nguyen, Siddiqi, y Whaley, 2022). Además de procedimientos automatizados que utilizan circuitos variacionales para generar operaciones lógicas a partir de procedimientos de codificación y corrección conocidos. El procedimiento se ha probado mediante simulación en computadoras clásicas en códigos cuánticos de cuatro a quince *qubits* (Chen, Vasmer, Breuckmann, y Grant, 2019).

En el presente trabajo se explicará cómo las técnicas de aprendizaje profundo se están utilizando para detectar errores en sistemas cuánticos. También se presentarán algunos casos de estudio recientes. Adicionalmente el presente artículo tiene también como objetivo ofrecer una introducción accesible a la aplicación del aprendizaje profundo en la computación cuántica, y mostrar cómo esta tecnología está ayudando a superar uno de los mayores desafíos en la construcción de sistemas cuánticos escalables y fiables.

Antecedentes

Un *qubit* es la unidad mínima lógica para el tratamiento y procesamiento de información cuántica. La representación de los estados básicos del *qubit* se denotan por un vector de estado en el espacio de Hilbert:

$$|0\rangle = \begin{pmatrix}1\\0\end{pmatrix};\ |1\rangle = \begin{pmatrix}0\\1\end{pmatrix}$$

por lo tanto

$$|\psi\rangle = \begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \alpha|0\rangle + \beta|1\rangle$$

La base $|0\rangle$ o $|1\rangle$ puede ser representada por:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle = \alpha\frac{\alpha+\beta}{\sqrt{2}}|+\rangle + \alpha\frac{\alpha-\beta}{\sqrt{2}}|-\rangle$$

Ettore Majorana postuló en 1937 la existencia de una partícula capaz de ser simultáneamente su propia antipartícula, lo que significa que su creación y aniquilación podrían ocurrir a través del mismo operador.

$$\gamma = \gamma^{\dagger}$$

En la Figura 1 podemos ver gráficamente lo anterior: El operador de creación $c^{\dagger}$ se utiliza para representar la presencia de un electrón, mientras que el operador de aniquilación c se usa para representar la presencia de un hueco.

Figura 8: Operadores de creación y aniquilación (elaboración propia,2023).

Cuando se trata de una superposición igual de ambos operadores, podemos comprobar que estos operadores de Majorana cumplen con la propiedad de ser su propia adjunta (Ver Figura 2):

Figura 2: Estados de superposición de Majorana (elaboración propia,2023).

La coherencia es esencial en este tipo de sistemas cuánticos ya que es la propiedad que permite que los *qubits* estén en un estado cuántico entrelazado (si hablamos un sistema de dos *qubits*), lo que significa que las mediciones realizadas en uno de los *qubits* están correlacionadas con las mediciones realizadas en el otro, incluso si están separados por una gran distancia. La coherencia permite que los *qubits* mantengan sus estados cuánticos superpuestos y entrelazados durante el tiempo suficiente para realizar cálculos cuánticos complejos.

Los sistemas cuánticos son una herramienta muy poderosa para el procesamiento y la transmisión de información, así como para la simulación de sistemas complejos. Sin embargo, la naturaleza cuántica de estos sistemas también los hace susceptibles a errores, lo que puede limitar su utilidad en algunas aplicaciones.

Uno de los principales problemas en los sistemas cuánticos es la decoherencia que se refiere a la pérdida de coherencia cuántica debido a la interacción con el entorno. Esto puede causar errores en la información almacenada en el sistema, lo que puede afectar la precisión de las operaciones cuánticas. Además, la decoherencia puede provocar la desaparición de estados cuánticos superpuestos, que son esenciales para la computación cuántica (Fortin, 2012).

Otro problema común es la amplificación de errores. A diferencia de los sistemas clásicos, en los sistemas cuánticos no se pueden copiar los estados cuánticos exactamente, lo que significa que cualquier intento de hacer una copia exacta de un estado cuántico siempre será imperfecto. Esto puede resultar en una propagación de errores a medida que se realizan más operaciones en el sistema.

Los sistemas cuánticos también están sujetos a errores causados por la falta de precisión en la implementación de las operaciones cuánticas. Esto puede ser causado por errores en la calibración de los dispositivos, fluctuaciones de ruido en los campos electromagnéticos utilizados para manipular los *qubits*, y otros factores.

Para minimizar estos errores, los científicos están trabajando en el desarrollo de técnicas para la corrección de errores en sistemas cuánticos. Estas técnicas se basan en la redundancia y la codificación de información para detectar y corregir errores, y pueden ayudar a mejorar la precisión de los cálculos cuánticos y la fiabilidad de los sistemas cuánticos en general (Ittrends, 2021).

Aprendizaje Profundo (Deep Learning)

El aprendizaje profundo es una técnica de inteligencia artificial que se basa en el uso de redes neuronales para analizar grandes cantidades de datos y extraer patrones complejos. En el contexto de los sistemas cuánticos, el aprendizaje profundo se utiliza para analizar la información generada por los sistemas y detectar cualquier error o anomalía en los resultados (González, 2018).

Una de las técnicas más prometedoras en este campo es el uso de redes neuronales convolucionales que han demostrado ser especialmente efectivas para detectar patrones en datos de imágenes y señales (Mera y Ochoa, 2021). En el contexto de los sistemas cuánticos, estas redes se utilizan para analizar los datos generados por los experimentos cuánticos y detectar cualquier patrón anómalo que pueda indicar la presencia de errores.

Una vez que se detectan los errores, se pueden implementar técnicas de corrección de errores para reducir su impacto en los resultados finales. Estas técnicas incluyen el uso de códigos de corrección de errores, que son algoritmos diseñados específicamente para detectar y corregir errores en los datos cuánticos (por el momento solo nos enfocaremos a la detección de errores).

La detección de errores en sistemas cuánticos es una tarea crítica para la realización de cálculos cuánticos precisos y confiables. Las técnicas de aprendizaje profundo están demostrando ser un enfoque prometedor para mejorar la detección de errores en estos sistemas. A medida que se desarrollan nuevas técnicas y algoritmos, es probable que veamos avances significativos en la capacidad de los sistemas cuánticos para el procesamiento de operaciones complejas en el futuro.

Deep Learning para la Detección de Errores

El uso de redes neuronales profundas para detectar y corregir errores en un sistema cuántico es posible gracias al aprendizaje a partir de los datos del sistema. Esto permite identificar patrones que indiquen la presencia de errores. Del mismo modo, esta técnica permite la clasificación de errores en diferentes categorías, lo que puede ayudar a los investigadores a entender mejor los tipos de errores que afectan el sistema cuántico. Esto puede llevar a una corrección de errores más efectiva.

Actualmente los *autoencoders* son redes neuronales que aprenden a reconstruir la entrada original (Wang y Prokhorov, 2012). En el contexto de la detección de errores cuánticos, los *autoencoders* se utilizan para detectar y corregir errores en un sistema cuántico. Para poder diseñar una red neuronal es necesario identificar los errores comunes que se presentan en la implementación de los sistemas cuánticos.

Los sistemas cuánticos son propensos a errores debido a las interacciones con el ambiente y a la decoherencia. Es necesario identificar los tipos de errores cuánticos que se quieren detectar y corregir, como errores de fase, errores de amplitud o errores de entrelazamiento. Los estados

cuánticos pueden ser representados mediante vectores de estado o matrices densidad. Es importante definir la representación que se utilizará para entrenar y evaluar la red neuronal.

La arquitectura de la red neuronal dependerá del tipo de errores cuánticos que se quieren detectar y corregir. Una posible arquitectura es la red neuronal de *autoencoder* cuántico, que consta de una capa de codificación y una capa de decodificación. La capa de codificación transforma el estado cuántico original en un código cuántico, mientras que la capa de decodificación transforma el código cuántico de nuevo al estado cuántico original.

Materiales y Métodos

Para poder identificar los tipos de errores que afectan a los sistemas cuánticos como errores de fase, errores de amplitud y errores de entrelazamiento. Una red neuronal puede ser útil para detectar la coherencia en un sistema de entrelazamiento de dos *qubits* mediante el análisis de datos obtenidos de mediciones cuánticas. La red neuronal puede ser entrenada para detectar patrones específicos en los datos que indican la presencia o ausencia de coherencia. Esto es especialmente útil para sistemas cuánticos grandes y complejos, donde la detección manual de la coherencia es difícil o incluso imposible. Además, una red neuronal también puede ayudar a identificar factores que pueden contribuir a la pérdida de coherencia en el sistema, lo que puede ser útil para el diseño y optimización de sistemas cuánticos.

Los estados cuánticos pueden ser representados mediante vectores de estado o matrices de densidad. Esta identificación es importante para diseñar una red neuronal efectiva que pueda detectar y corregir estos errores.

Para entrenar la red neuronal, se generan datos de entrenamiento a partir de simulaciones de sistemas cuánticos con errores inducidos. Estos datos incluyen estados cuánticos en los que se han introducido errores específicos.

La red neuronal diseñada para este estudio es una red de *autoencoder*, que consta de una capa de codificación y una capa de decodificación. La capa de codificación transforma el estado cuántico original en un código cuántico, mientras que la capa de decodificación transforma el código cuántico de nuevo al estado cuántico original. Esta arquitectura permite que la red neuronal aprenda a reconstruir el estado cuántico, identificando y corrigiendo los errores en el proceso.

La computación cuántica topológica es uno de los enfoques más interesantes para la construcción de computadoras cuánticas estables, y se basa en el uso de estados topológicos para almacenar y procesar información. Esta tecnología utiliza estados topológicos de la materia, en los que las excitaciones de las cuasipartículas son diferentes de los bosones y fermiones, y se conocen como aniones no abelianos. Estas partículas se caracterizan por seguir las estadísticas de trenzado no abelianas, lo que les da propiedades únicas y útiles para el desarrollo de la tecnología.

Resultados

El entrenamiento de la red neuronal se realiza utilizando datos de entrenamiento generados a partir de los estados cuánticos con errores. Durante el entrenamiento, se ajustan los pesos y las conexiones de la red neuronal para mejorar su capacidad para detectar errores en los estados cuánticos.

Después del entrenamiento, se valida la red neuronal utilizando datos de validación separados de los datos de entrenamiento. Luego, se prueba la red neuronal utilizando datos de prueba para determinar su precisión y eficacia en la detección de errores en los estados cuánticos. En la Figura 3, se presenta el diseño de la red neuronal que permite la detección de errores en sistemas cuánticos:

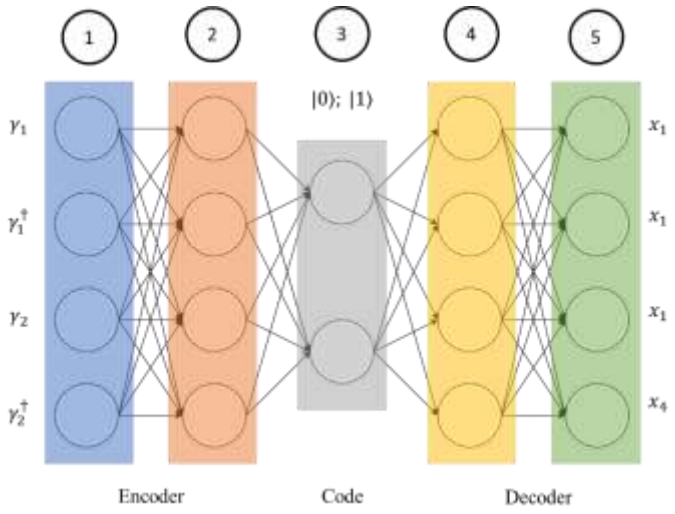


Figura 3: Red neuronal de detección de errores (elaboración propia,2023).

1. Entrada: La red neuronal recibe como entrada la información del estado cuántico que se

desea verificar y corregir.

2. Capa de preprocesamiento: Esta capa prepara los datos de entrada para ser procesados por la red neuronal. Puede incluir la normalización y la transformación de los datos.
3. Capas ocultas: La red neuronal utiliza varias capas ocultas para procesar los datos y extraer características relevantes del estado cuántico. Estas capas pueden ser diseñadas utilizando técnicas de aprendizaje profundo, como redes neuronales convolucionales o redes neuronales recurrentes.
4. Capa de detección de errores: Esta capa se encarga de detectar errores en el estado cuántico. Puede utilizar técnicas de clasificación binaria para determinar si el estado cuántico es correcto o si contiene errores.
5. Salida: La red neuronal produce una salida que indica si el estado cuántico ha sido verificado y corregido correctamente.

La red neuronal comienza recibiendo como entrada la información del estado cuántico que se desea verificar y corregir en este caso se trata de un par de Majoranas. La información de entrada es preprocesada para su procesamiento adecuado por la red neuronal, incluyendo normalización y transformación de los datos.

La red neuronal utiliza varias capas ocultas para procesar los datos y extraer características relevantes del estado cuántico (estados $|0\rangle$ y $|1\rangle$), utilizando técnicas de aprendizaje profundo, como redes neuronales convolucionales o redes neuronales recurrentes.

Si hay errores en el estado cuántico, la capa de detección de errores los detectará utilizando técnicas de clasificación binaria para determinar si el estado cuántico es correcto o si contiene errores.

Es importante señalar que el presente diseño de red neuronal puede variar dependiendo de la robustes de la arquitectura específica de cada sistema cuántico a la que se desee implementar. Además, es posible que se necesiten múltiples redes neuronales para manejar diferentes tipos de errores y operaciones cuánticas.

Discusión

La detección de errores en sistemas cuánticos es un desafío importante para los científicos e investigadores que trabajan en el campo de la informática cuántica. Los sistemas cuánticos son extremadamente sensibles a los errores, lo que puede comprometer la precisión y confiabilidad de los cálculos realizados en ellos. Por esta razón, se están desarrollando técnicas avanzadas de aprendizaje profundo para mejorar la detección y reducción de errores en estos sistemas.

Cierre

El diseño de redes neuronales para la detección de errores en sistemas cuánticos, en particular en sistemas de *qubits*, es un enfoque prometedor para mejorar la eficacia y eficiencia de la detección de errores en estos sistemas. La detección de errores es crucial para la fiabilidad y precisión de los cálculos realizados en sistemas cuánticos, y las redes neuronales son capaces de detectar patrones complejos en grandes conjuntos de datos que serían difíciles de detectar por métodos convencionales. A medida que la informática cuántica continúa avanzando, la aplicación de técnicas de aprendizaje profundo para la detección de errores será cada vez más importante para maximizar la precisión y confiabilidad de estos sistemas, y mejorar los métodos de fabricación de computadoras cuánticas escalables y prácticas.

Referencias

Chen, H., Vasmer, M., Breuckmann, N., y Grant, E. (2019). “Machine learning logical gates for quantum error correction”. arXiv:1912.10063v1 [quant-ph]. https://doi.org/10.48550/arXiv.1912.10063

Convy, I., Liao, H., Zhang, S., Patel, S., Livingston, W. P., Nguyen, H. N., Siddiqi, I., y Whaley, K. B. (2022). “Machine learning for continuous quantum error correction on superconducting qubits”. New Journal of Physics, Vol. 24 No.6, pp. 063019. http://dx.doi.org/10.1088/1367-2630/ac66f9.

Deng, L., y Yu, D. (2014). “Deep learning: methods and applications". Foundations and Trends in Signal Processing, Vol. 7, No 3-4, pp. 197-387. https://doi.org/10.1561/2000000039

De Wolf, R. (2023). “Quantum Computing: Lecture Notes”. arXiv preprint arXiv:1907.09415 [quant-ph]. https://doi.org/10.48550/arXiv.1907.09415

Fortin, S. (2012). "Hacia una mejor comprensión de la decoherencia desde una perspectiva general". Revista Colombiana de Filosofía de la Ciencia, Vol 24 No. 12, 65-82. https://doi.org/10.18270/rcfc.v12i24.1695

González, M., A. (2018). "Aplicaciones de técnicas de inteligencia artificial basadas en aprendizaje profundo (deep learning) al análisis y mejora de la eficiencia de procesos industriales". Tesis de maestría, Maestría en ingeniería de automatización e informática industrial. Universidad de Oviedo.

Ittrends (2021). Crean un algoritmo de aprendizaje automático para detectar errores cuánticos. Recuperado el 27 de abril de 2023 de https://goo.su/inLU7bJ.

Mera, J., L. y Ochoa, G., J. F. (2021). "Redes neuronales convolucionales para la clasificación de componentes independientes de rs-fMRI". TecnoLógicas, Vol. 24, No. 50. https://doi.org/10.22430/22565337.1626

Nielsen, M.A. y Chuang, I.L. (2010). “Quantum Computation and Quantum Information”. Cambridge University Press. New York.

Wang, J., He, H., y Prokhorov, D. (2012). A Folded Neural Network Autoencoder for Dimensionality Reduction. Procedia Computer Science, Vol. 13, pp. 120-127. http://doi: 10.1016/j.procs.2012.09.120.

ANÁLISIS LONGITUDINAL SOBRE LOS GRADOS DE CONCRECIÓN DEL PENSAMIENTO SUICIDA EN ALUMNOS DE INGENIERÍA

LONGITUDINAL ANALYSIS OF THE DEGREES OF CONCRETIZATION OF SUICIDAL THINKING IN ENGINEERING STUDENTS

Enrique Noé Hernández Arroyo[22]
ORCID: 0000-0002-67
enriquehernandezarroyo13@hotmail.com

Marco Antonio Magaña Serrano[23]

Resumen

Después de la pandemia, ha existido un aumento significativo en la tasa de mortandad en los casos de suicidio en México, particularmente se ha incrementado en adolescentes y jóvenes, con esta condición, el objetivo central del presente estudio cualitativo de corte discreto consiste en evaluar la progresión sistemática de la ideación, planeación, intención y consumación suicida en alumnos de Ingeniería en Comunicación Multimedia (ICM) durante un periodo de seis cuatrimestres de escolarización. Para ello, se empleó un diseño investigación longitudinal y un análisis estadístico de series de tiempo para un segmento muestral de 71 alumnos. Así expuesto, se reconoce un incremento de la pendiente entre la primera hasta la tercera evaluación en F1 (pensamiento suicida) y F2 (planeación suicida); sin embargo, los casos disminuyeron en forma insipiente a partir de la cuarta hasta la sexta evaluación, con lo antes expuesto se sugiere que la novedad y los procesos de adecuación académica incrementa las tasas de estrés, por lo tanto, es probable la aparición de algún indicador asociado al comportamiento suicida. Asimismo, en la pendiente de F4 (consumación

[22] Dr. en Administración. Universidad Estatal del Valle de Ecatepec. División de Ingeniería en Comunicación Multimedia.
[23] Dr. en Astronomía. Institución: Universidad Estatal del Valle de Ecatepec. División de Ingeniería en Comunicación Multimedia.

suicida), se identifica de manera estacional y sin variación dentro del proceso de evaluación. Con los resultados obtenidos se espera proponer un programa de intervención de acuerdo a las condiciones particulares de cada estudiante.

***Palabras Clave**:* Ideación Suicida, Intención Suicida, Consumación suicida, Evaluación longitudinal.

Summary

After the pandemic, there has been a significant increase in the mortality rate in cases of suicide in Mexico, particularly in adolescents and young people. With this condition, the central objective of this discrete qualitative study is to evaluate the systematic progression of suicidal ideation, planning, intention and consummation in Multimedia Communication Engineering (MCI) students during a period of six semesters of schooling. To this end, a longitudinal research design and a statistical analysis of time series were used for a sample segment of 71 students. Thus, an increase in the slope between the first and third evaluations is recognized in F1 (suicidal thinking) and F2 (suicidal planning); however, the cases decreased insipiently from the fourth to the sixth evaluation, with the above it is suggested that the novelty and the processes of academic adequacy increase stress rates, therefore, it is likely that some indicator associated with suicidal behavior will appear. Likewise, in the slope of F4 (suicidal consummation), it is identified seasonally and without variation within the evaluation process. With the results obtained, it is expected to propose an intervention program according to the particular conditions of each student.

Keywords: suicidal ideation, suicidal intention, suicidal completion, longitudinal evaluation.

Introducción

La muerte representa una parte elemental de la vida; pero cuando se encuentra matizada por eventos trágicos como en el caso del suicidio, se convierte en un suceso que retrata una ruptura en el funcionamiento y cohesión integral de las instituciones. De esta manera, el suicidio es un fenómeno que se ha modificado de acuerdo con la época, y caracterizado por su transitar demográfico y etnográfico. Actualmente, el suicidio es una de las principales causas de muerte a nivel mundial (Azúa, 2020); representada por la quinta causa de muerte (Dávila & Pardo, 2020). Asimismo, en México, se ha incrementado tres veces más los casos de suicidio en los últimos 20

años, en donde preferentemente se focaliza en jóvenes, varones, aislamiento involuntario o voluntario, enfermedades terminales, personas de la tercera edad, adicciones, pobreza y padecimientos psiquiátricos (Goñi, 2019). Reforzando el punto anterior durante el proceso de pandemia se han incrementado los casos en un 30% en la población global (Jerónimo, et al.2021). Por lo tanto, el tratamiento no tan solo debe estar dirigido dentro de los métodos de intervención psicológica o psiquiátrica, sino el alcance debe situarse dentro de los parámetros de la multidisciplinariedad científica relacionada al diagnóstico y tratamiento paliativo.

Con lo anterior, el objetivo del presente estudio, de corte longitudinal, consiste en el diagnóstico progresivo del pensamiento suicida durante los primeros seis cuatrimestres de alumnos en Ingeniería en Comunicación Multimedia (ICM), debido a que este segmento muestral representa —de acuerdo a su edad cronológica— un sector de alto riesgo (Luna & Cervantes, 2018), con esto, se pretende identificar posibles variaciones del pensamiento suicida durante su proceso de escolarización universitaria, el interés reside en reconocer la mayor prevalencia de acuerdo a las cuatro etapas en la detección y pronósticos del pensamiento suicida durante los años del 2021 al 2023. Asimismo, con los resultados obtenidos se espera establecer acciones contingentes en la atención de dicha problemática y sugerir métodos preventivos para futuras generaciones estudiantiles.

Marco Referencial

Así expuesto, la acción suicida suele ser un fenómeno polémico de abordar debido a las circunstancias o causas en las que se puede consumar el acto —que si bien es cierto que el sujeto lo materializa— se desconocen otros factores, tales como el grado de su voluntad, gravedad psiquiátrica, influencia ideológica o en general condición social (García, et al., 2019), por lo tanto, el establecer una definición válida no representa una tarea sencilla. A esto Durkheim (2004) propone que el suicidio es "… todo caso de muerte que resulta directa o indirectamente de un acto positivo o negativo, llevado a cabo por la propia víctima que sabía que iba producir ese resultado" (p. 16). La definición anterior sugiere que el suicidio es una posición del individuo frente a la intencionalidad de querer extinguir su vida; no obstante, aquí se identifican dos posibles cuestionamientos, el primero de ellos es con respecto al término "positivo" o "negativo", esta

diatriba sigue siendo altamente cuestionable principalmente porque al calificarlo así limitaría la comprensión de las causas concretas de la acción suicida.

El segundo cuestionamiento reside en el grado de su voluntad o intencionalidad, por ejemplo, el sujeto bajo los efectos de alguna posible sustancia adictiva —como en el caso del consumo de alcohol, cannabis o sedantes— incrementa las probabilidades que el acto suicida se logre consumar (Amador, et al.,2018, Fonseca & Halabia, 2024); por lo tanto, el sujeto al presentar dificultades evidentes en la autorregulación de su comportamiento toma decisiones fuera del análisis plenamente consciente. Una definición más actual se tiene con la propuesta de Rangel & Jurado (2020) quienes consideran el suicidio como "la muerte por lesión, envenenamiento o asfixia, sobre la cual existe evidencia (explícita o implícita) de que la lesión fue autoinfligida y de que la persona fallecida deseaba, en efecto, morir" (p. 41).

Con la definición anterior, la acción suicida puede encontrarse caracterizada a partir de ciertos indicadores que permiten establecer un pronóstico global, de los más evidentes es la práctica autolesiva (O'Connor & Nock, 2014), en donde el sujeto libera parte de su agobio o presión social a partir de lesiones corporales, en donde el autoconcepto del individuo se encuentra comprometido; asimismo se suscita un intercambio simbólico entre el riesgo de poner en tela de juicio la vida, y por otra, provocar un sentimiento de vivir ausente. A esto, Kaplan y Szapu (2019) señala que: "la sensación de alivio obtenido mediante cortes en la piel parece reemplazar un dolor (social) por otro (físico), aunque mas no sea por breves lapsos de tiempo" (p. 3). Con lo anterior, el acto autolesivo no necesariamente propicia una intensión suicida, sino representa en algunos casos una recreación o simulación de ver a la muerte desde afuera (omnipresente), sin llegar a consumar el acto suicida, esta situación remplaza provisionalmente el estado de desesperanza (Faura, et al., 2021), haciendo que su vida cotidiana siga siendo funcional[24].

Otro de los indicadores elementales que permiten pronosticar el acto suicida es a partir de la detección de algún cuadro de depresión, a esto, García y Aroca (2014) señalan que un individuo al manifestar una propensión a pensamientos suicidas, frecuentemente se originan con un cuadro depresivo; de acuerdo con DMS-5 (2014) la depresión —en su forma elemental— se define como: "un trastorno del estado de ánimo, donde la principal característica es una alteración del humor y, según su temporalidad y origen sintomático" (p. 106), además, las formas de su

[24] Las lesiones que son autoinfringidas dentro de la misma literatura psicológica son conocidas con el término técnico de "Parasuicidio" (Quintero, et. al. (2021).

clasificación básica, es el trastorno depresivo mayor, trastorno distímico y trastornos bipolares (Corea, 2021). Aunado al punto anterior, Pabon (2021) señala que el cuadro depresivo propicia un porcentaje de recurrencia de intención suicida del 50%, asimismo el cuadro aumenta en el caso de ataques de pánico y ansiedad. En menos porcentaje se encuentra un cuadro de personalidad limítrofe, trastorno bipolar y un trastorno subtipo purgativo. Con lo anterior, es importante notar que todo padecimiento psiquiátrico —por más grave que sea— no es un factor que determina la consumación del suicidio[25].

Por su parte, las adicciones representan otro factor elemental en la detección y progresión suicida. Para Valdevila, et al. (2021), la codependencia de sustancias adictivas y la edad cronológica está directamente vinculada con la ideación suicida; los mismos autores señalan, por ejemplo, que la ingesta de alcohol en jóvenes menores de 18 años presenta un alto nivel de recurrencia en la ideación suicida, a diferencia de adultos mayores de más de 36 años, en donde los casos suelen ser insipientes; sin embargo, es importante notar que el tipo de sustancia (legal o ilegal) también condiciona su recurrencia, por ejemplo, en el consumo de cannabis u opioides, independientemente de la edad, incrementa en forma significativa la ideación suicida. Por lo tanto, se sugiere que el consumo de drogas inhibitorias, como en el caso de alcohol, opioides o cannabis, incrementa las probabilidades de su recurrencia, a diferencia de la ingesta de drogas como la cocaína o metanfetamina cuya respuesta suele estar asociada a un estado de excitación.

Otro de los aspectos importantes que limita la confrontación adecuada del pensamiento suicida es el involucramiento con las diferentes redes de apoyo. La familia, como institución elemental formativa, provee herramientas psicológicas suficientes en el manejo pertinente en situaciones de crisis (Muñetón & Vázquez, 2018). También el núcleo de amistades, integración laboral y grupos de recreación puede contribuir en el equilibrio psicológico ante situaciones amenazantes (Fonseca & Pérez, 2021). Sin embargo, es notorio que esos mismos contextos de interacción social, en donde prevalece la violencia, la discriminación y la limitada accesibilidad a las instituciones, como lo es una educación dignificante, empleo o servicios de salud de calidad, propician la aparición de un conflicto interno en el individuo, sintiéndose asimismo defraudado por la solides de sus sistemas de interacción social.

[25] En relación a este punto, los estudios de De La Estrella (2010) han identificado que entre el año de 1998 al 2007, solo se han presentado 12 casos de suicidio consumado en los centros psiquiátricos en Colombia.

Reforzando el punto anterior, el pensamiento suicida presenta una relación directa con el funcionamiento cerebral, de acuerdo con Bertolón (2022) existen dos condiciones biológicas que permiten caracterizar la comprensión del fenómeno del suicidio, el primero se comprende a nivel cortical en donde se ha identificado que las áreas comprometidas en la corteza prefrontal, particularmente en la parte ventral y dorsal, afecta la correcta toma de decisiones y la capacidad de mantener el autocontrol. La otra condición se identifica en el marcador genético 5-HTTLPR, cuya función consiste en un trasportador eficiente de la serotonina (5-hidroxitriptamina), el cual previene la depresión e incrementa los estados de bienestar (De Jesús, et al. 2019).

Finalmente, uno de los indicadores que pudieran pronosticar el origen de conductas suicidas se encuentran vinculadas con las preferencias sexuales. La condición de personas lesbianas, gays, bisexuales y transexuales, como grupo social minoritario, representa un sector altamente vulnerable, para Gil, et al. (2020) señalan que este segmento social se ubica dentro de un factor de alto riesgo, no necesariamente por lo que implica en sí su preferencia sexual, sino, por el desafortunado estigma o desaprobación social; aunado a este punto, los mismos autores sugieren que la progresión de la ideación suicida a una intención o consumación del acto se incrementa en un 30% en el primer año después de externalizar abiertamente sus preferencias.

Por otra parte, la evaluación predictiva del pensamiento suicida se estructura a partir de etapas; no obstante, es importante hacer énfasis en que las mismas no son reglas universales en las que todo suicidio consumado debe transitar (Miranda, 2009); sino es importante evaluar el grado de riesgo y establecer las medidas preventivas para su adecuada atención. Así planteado, Soto, et al. (2020) identifica cuatro etapas diferenciadas en el suicidio, la primera es la *ideación* suicida, en donde el sujeto al mantener fijado un pensamiento, considerado como amenazante, desea extinguir su propia vida, esta primera etapa se diferencia de las demás por un estado de desesperanza o situación agobiante. La siguiente etapa es la *planeación suicida*, aquí el sujeto se ha convencido a sí mismo que el suicidio es una alternativa viable, para ello, por ejemplo, asigna horarios, desarrollan cartas de despedida, emplea un lenguaje autodestructivo, establece procedimientos o métodos para su ejecución, entre otros.

La siguiente etapa es la *intención suicida*, aquí se identifican los grados en las que los individuos desean quitarse la vida; por lo tanto, el sujeto puede manifestar un tiempo fijo o indefinido dependiendo de la intensidad en la que se manifiestan las etapas anteriores (García, 2019); después de que el individuo se convence a sí mismo y planea el acto suicida, ejecuta la

acción, considerando que los resultados sean irreversibles. Aquí surge la última etapa, subdividida en dos implicaciones: la *sobrevivencia* y *el suicidio consumado*. Ante la sobrevivencia, según para Builes (2014), la presencia de lesiones corporales graves puede afectar la rutina cotidiana del individuo, por lo que los estados de ánimo están seriamente afectados si el sujeto aun es consciente. En el caso de que el suicidio consumado se materialice, las afecciones psicológicas se encuentran situadas en la familia y en general en las debilidades institucionales, las cuales retratan la insuficiencia en la integración operativa de la misma sociedad.

Metodología

El objetivo del presente estudio cuantitativo de corte longitudinal, consiste en el diagnóstico progresivo del pensamiento suicida en alumnos de Ingeniería en Comunicación Multimedia; la intensión reside en identificar si existen ciertos periodos críticos, durante su proceso de escolarización universitaria, en donde el alumno preferentemente manifiesta algún indicador o etapa evidente relacionada con el suicidio. Se decide el análisis de este segmento debido a que es un grupo social considerado de alto riesgo (Luna & Cervantes, 2018). Asimismo, se eligió un diseño de investigación *expo facto,* correspondiente a un análisis de tipo cronológico; esto es, se aplica la misma prueba dosificada durante los años 2021 al 2023, cada evento de evaluación se presenta durante seis cuatrimestres justo en el proceso de entrega de proyectos finales y aplicación de exámenes; aquí la acción del estudiante se encuentra caracterizada a partir del empleo de estrategias de resiliencia en donde se identifica la presencia de altos niveles de estrés, además el propósito reside en establecer una posible proyección de los casos relacionados con el pensamientos suicida de acuerdo con las cuatro fases de evaluación planteadas desde el propio instrumento de medición.

Por su parte, el instrumento de medición empleado es el inventario de Ideación Suicida de Beck (ISB), el cual presenta un nivel de fiabilidad de cronbach para población mexicana de .85, esto es, se considera con un nivel de fiabilidad de “suficiente” a “buena” (González, et al., 2000). Dentro de su estructura, se plantean 20 reactivos, en donde las alternativas de respuesta se encuentran dosificadas en tres opciones, asignadas por una ponderación del 0 al 2 (Córdova & Rosales, 2012). Los 20 reactivos se encuentran operacionalizados en cuatro rubros: a) características de las actitudes hacia la vida y la muerte (F1) , cuya asignación es definida de la

pregunta 1 a la 5, b) el segundo es definido por la planeación o deseo suicida (F2), su asignación se evalúa de la pregunta 6 a la 10, c) el tercero se ubica en la intención suicida, este criterio se asigna en las preguntas 11 al 15 (F3) , y finalmente la d) actuación del intento, cuya asignación se considera de la pregunta 16 a la 20 (F4). Con lo anterior, es elemental aclarar que los cuatro indicadores de evaluación reflejan de manera particular las cuatro fases relacionadas con la acción suicida. En la tabla 1 se describe la operacionalización del instrumento.

Tabla Nº 1.

Definición operacional del instrumento del ISB.

Item.	Opciones / ponderación	Concepto operacional
1.- Deseo de vivir	Moderado a fuerte (0) Poco (1) Ninguno (2)	F 1
2.- Deseo de morir	Ninguno (0) Poco (1) Moderado a fuerte (2)	F 1
3.- Razones para vivir/morir	Vivir supera a morir (0) Equilibrado (1) Morir supera a vivir (2)	F 1
4.- Deseo pasivo de suicidio	Tomaría precauciones para salvar su vida (0) Dejaría la vida/muerte a la suerte (1) Evitaría los pasos necesarios para salvar o mantener su vida (2)	F 1
5.- Duración.	Breves, periodos pasajeros (0) Periodos largos (1) Continuos (2)	F1

6.- Frecuencia.	Rara (0) Intermitente (1) Persistentes y continuos (2)	F 2
7.- Actitud hacia los pensamientos/deseos de suicidio	Rechazo (0) Ambivalente; indiferente (1) Aceptación (2)	F 2
8.- Control sobre la acción/deseo de suicidio	Tiene sentido de control (0) Inseguridad de control (1) No tiene sentido de control (2)	F 2
9.- Razones o motivos que detienen la tentativa suicida (familia, religión, irreversibilidad o fallar) .	No lo intentaría porque algo lo detiene (0) Cierta duda hacia los intereses que lo detienen (1) Ninguno o mínimo interés en las razones que detienen indicar qué factores lo detienen (2)	F 2
10.- Razones para pensar/desear el intento suicida	Manipular el ambiente, llamar la atención, venganza (0) Combinación de 0 y 2 (1) Escapar, acabar, salir de problemas (2)	F2
11.- Deseo de realizar un intento de suicidio activo.	Ninguno (0) Poco (1) Moderado a fuerte (2)	F3
12.- Método: especificidad/oportunidad de pensar en el intento suicida.	Sin considerar (0) Considerado, pero sin elaborar los detalles (1) Detalles elaborados/bien formulados (2)	F 3
13.- Método, disponibilidad/oportunidad de pensar el intento.	Método no disponible, no oportunidad (0) Método tomaría tiempo/esfuerzo; oportunidad no disponible realmente (1) Método y oportunidad disponible (2) Oportunidad futura o disponibilidad de anticipar el método (2)	F 3

14.-Sensación de "capacidad" para llevar a cabo el intento	Falta de valor, debilidad, asustado, incompetente (0) No está seguro de tener valor (1) Está seguro de tener valor (2)	F 3
15.-Expectativa/anticipación de un intento real	No (0) Sin seguridad, sin claridad (1) Sí (2)	F 3
16.-Preparación real	Ninguna (0) Parcial (guardar pastillas, conseguir instrumentos (1) Completa (2)	F 4
17.-Nota suicida	Ninguna (0) Sólo pensada, iniciada, sin terminar (1) Escrita, terminada (2)	F 4
18.-Acciones finales para preparar la muerte (testamentos, arreglos, regalos)	No (0) Sólo pensados, arreglos parciales (1) Terminados (2)	F 4
19.-Encubrimiento de la idea de realizar el intento suicida.	Revela las ideas abiertamente (0) Revela las ideas con reserva (1) Encubre, engaña, miente (2).	F 4
20.- Intento de suicidio anteriores	No ha intentado (0) Uno (1) Mas de uno (2)	F 4

Nota: Inventario de pensamiento e ideación suicida del ISB. F1: Pensamiento suicida; F2: Planeación suicida, F3: Intención suicida y F4: consumación suicida.

En el caso del análisis muestral, se eligió un grupo de 71 alumnos de primer cuatrimestre, 45 mujeres y 26 varones, quienes se encuentran en un rango de 18 a 22 años. El tipo de muestreo es por conveniencia donde el factor de inclusión consiste en la regularidad académica del estudiante, a su vez, es vital la inscripción vigente en la universidad en sus primeros dos años de Ingeniería. Por su parte, en el factor de exclusión, el alumno, al no contestar las seis evaluaciones

sugeridas dentro del proceso, se considera pertinente su separación del análisis estadístico global. En relación con la condición geográfica y demográfica de la muestra se ubica en el Estado de México, particularmente en los municipios de Ecatepec, Chimalhuacán, Tecámac y Ciudad Azteca, en donde su nivel socioeconómico se encuentra representado dentro de un margen de clasificación C- y D+. [26]

Por su parte, la interpretación estadística considerada en el estudio se divide en dos clases de análisis: descriptivo y series de tiempo con un criterio univariado, en donde se desea identificar la estacionalidad o composición cíclica del estudiante frente al pensamiento suicida. Aunado a este punto, el programa estadístico para los fines del presente estudio fue Microsoft Excel 365. Finalmente, los lineamientos éticos en la presente investigación se fundamentan en los tratados internacionales de Helsinkin, considerando por escrito la autorización de los alumnos en la participación del estudio, asimismo a cada integrante se exponen las medidas de seguridad e higiene, y se les plantea a su vez los aspectos que permiten la protección de su integridad física, en donde se garantiza la figura del anonimato en la calidad de sus respuestas

Resultados

Considerando las seis etapas evaluadas durante un periodo de dos años, se sugiere un incremento en F1 (pensamiento suicida) del 6% entre la primera y la tercera evaluación, esto es, de los 71 casos considerados en el análisis se incrementó su recurrencia para el tercer cuatrimestre a solo 6 alumnos; no obstante, dentro de la quinta y sexta evaluación, los casos disminuyeron de manera significativa entre el 8% a 11% de recurrencia. Este dato es relevante dado que de acuerdo con su edad cronológica y en base a la periodicidad de las evaluaciones aplicadas durante los seis periodos, se identificaron 30% de alumnos con al menos un indicador asociado a conductas suicidas, asimismo los casos disminuyeron a solo 14 estudiantes hasta la sexta evaluación.

Por su parte, se detectaron casos relacionados con la planeación suicida (F2), bajo esta condición se incrementó en 3% entre la primera y tercera evaluación; en forma contrastante, para las tres evaluaciones restantes disminuyó a 7% del total, es decir, a solo 8 alumnos ubicados en

[26] En la clasificación "C-"considera las necesidades cubierta de espacio y sanidad, asimismo cuenta con enseres y equipos que le aseguren el mínimo de practicidad u comodidad en el hogar, por su parte, la clasificación "D+" consiste en que los hogares de este nivel se caracterizan por tener cubierta la mínima infraestructura sanitaria de su hogar.

F2; esta condición permite clarificar dos puntos, por una parte se identifica un rango de estacionalidad durante los 24 meses de prueba, principalmente porque no decrece o aumenta de manera significativa los casos bajo el criterio de "planeación suicida", y por otra parte, sugiere que dentro de los parámetros de F1 y F2 se reconoce una disminución en los casos a partir de la cuarta evaluación, es decir, a partir del cuarto cuatrimestre se empieza identificar una disminución insipiente en relación al número total de alumnos con algún indicador de conductas suicida.

Para el caso de F3 (intención suicida), la pendiente se mantuvo relativamente uniforme, esto es, entre la primera y la tercera evaluación los casos incrementaron a 4%; sin embargo, para la cuarta a la sexta evaluación disminuyeron a 5%; estos resultados muestran un grado de paridad entre F1 y F2, es decir, en las tres escalas de evaluación hubo una disminución de los casos a partir del cuarto periodo de evaluación, aunque estos no sean del todo significativos de acuerdo a las variaciones de las pendientes. Finalmente, en la consumación suicida representada por F4, la pendiente se mantuvo constante entre la primera a la sexta evaluación, es decir, incremento a un solo caso durante los seis cuatrimestres. En la Tabla 2 se representan los datos descriptivos de las seis evaluaciones implementadas de acuerdo con su correspondencia en las cuatro etapas diferenciadas en el comportamiento suicida.

Con lo antes expuesto, es importante considerar, en forma descriptiva, dos aspectos que presentan relevancia de acuerdo con el análisis de los datos, por una parte, que de los cinco casos ubicados en una prevalencia de F4, se identificaron cuatro alumnos que han sido atendidos bajo tratamiento psicofarmacológico y monitoreo psiquiátrico; esto se detectó a partir de la aplicación de una entrevista clínica debido a que representan casos de alto riesgo dentro del trabajo operativo de la universidad. El otro planteamiento global consiste en que el alumno al estar expuesto a formas de trabajo relativamente novedosas y la constante imprevisión suscita un incremento en los estados de estrés o ansiedad (Hinojosa, et al., 2022); sin embargo, es importante notar que la habituación a las formas de trabajo académico permite una mejor regulación ante diferentes condiciones amenazantes, este planteamiento se puede confirmar a partir de que en el cuarto periodo escolar los casos relacionados con algún indicador en conductas suicidas disminuyeron.

Tabla Nº 2.

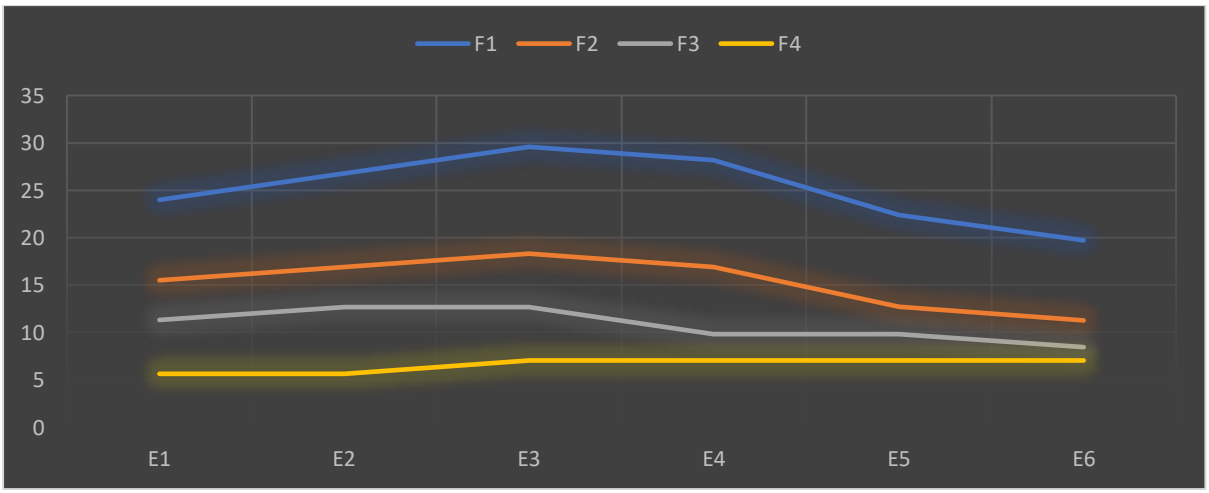
Distribución de casos y porcentajes de acuerdo con las cuatro fases relacionadas del suicidio en los años 2022 y 2023.

				Año 2023		
	E1%	E2%	E3%	E4%	E5%	E6%
F1	17 (24%)	19 (26%)	21 (30%)	20 (28%)	16 (22%)	14 (19%)
F2	11 (15%)	12 (16 %)	13 (18%)	12 (16 %)	9 (12%)	8 (11 %)
F3	8 (11%)	9 (13%)	9 (13 %)	7 (9%)	7 (9%)	6 (8%)
F4	4 (6 %)	4 (6%)	5 (7%)	5 (7%)	5 (7%)	5 (7%)

Nota: La abreviatura E1 a la E6, representan las diferentes mediciones durante los años 2022 y 2023. Por su parte, F1, F2, F3 y F4, representan las cuatro etapas elementales relacionadas con el comportamiento suicida.

Figura Nº 1

Representación gráfica de la conducta suicida durante los seis periodos evaluados.



Nota: La abreviatura E1 a la E6, representan las diferentes mediciones durante los años 2022 y 2023. Por su parte, F1, F2, F3 y F4, representan las pendientes de las cuatro etapas elementales relacionadas con el comportamiento suicida (autoría propia) .

En relación con el análisis estadístico de *series de tiempo*, se puede identificar una proyección para los consecuentes periodos escolares en F1 (pensamiento suicida), entre un rango de 14 a 16 casos; en cambio en F2 (planeación suicida) se estima un rango de 9 a 11 casos; no obstante, la pendiente dentro de este rango, se mantuvo sin variación significativa. Por su parte, F3 (intención suicida), se establece una proyección para el siguiente periodo de 5 a 4 alumnos; aquí el dato relevante consiste en que la tendencia de la pendiente empieza a identificar datos insipientes a la baja. En cambio, para F4 (Consumación suicida) se identifica una proyección de 4 a 5 alumnos, bajo esta variante no existen diferencias significativas entre la primera y sexta evaluación, los mismos casos se mantuvieron constantes durante los seis periodos de prueba, por lo que supone que el factor de estrés escolar no se encuentra directamente correlacionado con indicadores de comportamiento suicida.

Discusión

Uno de los aspectos elementales que caracteriza a la conducta suicida es la condición estacional o la identificación de la temporalidad específica, en donde, en forma cíclica, se reconoce un aumento o disminución exponencial en casos relacionados con las conductas suicidas. A esto, Carbonell (2007), señala que los casos de suicidio aumentan al finalizar la primavera y al principio del verano; el segundo incremento se da en otoño, para descender notablemente en invierno; sin embargo, Quevedo & Palomares (2023) hacen referencia que en gran medida la composición temporal dependen de situaciones radicales como cambios bruscos de clima, movimientos ideológicos o variaciones económicas, entre las más evidentes; no obstante, durante el desarrollo y monitoreo de las conductas suicidas, en los dos años de evaluación, se identifican cambios en las tendencias a partir del cuarto periodo de evaluación, en particular en el pensamiento y planeación suicida; a este respecto se sugiere que el proceso de adecuación del alumno ante la novedad de sus estudios presentan indicadores asociados al estrés; sin embargo, también es importante notar que la ineficiencia de los recursos psicológicos de los estudiantes y sus formas de comunicación asertiva, como aspectos elementales en la regulación del estrés, afectan una pronta adaptación ante condiciones altamente demandantes.

Así expuesto, no podemos concordar con algunos teóricos de manera substancial (Carbonell, 2007, Corea, 2021, Quevedo & Palomares, 2023), al señalar que del incremento o

decremento significativo de la tasa de suicidios se focaliza en ciertos periodos del año; sino los estudios relacionados al tema del suicidio deben focalizarse inicialmente a nivel local o regional con el propósito de sistematizar las posibles causas y sus formas directas de intervención.

Aunado al punto anterior, se sugiere que el incremento en F1 (pensamiento suicida) y en F2 (planeación suicida) durante las primeras cuatro evaluaciones se debe, en alguna medida, a los procesos de adecuación a las formas de exigencia universitaria; una vez adaptado, se identifica una disminución hasta los indicadores similares a F4 (consumación suicida); con lo antes expuesto, no se puede aún señalar que la causa concreta sea por factores asociados al estrés, sino el análisis debe extenderse a la realidad fenomenológica del estudiante, en términos de su autoconcepto, autoimagen, proyección hacia el futuro y la razón de su propia existencia (Luna & Cervantes, 2018). En consecuencia, esta situación puede considerarse como "natural" e identificable para este segmento de edad; sin embargo, no se sugiere descartar su nula atención, sino se propone una detección temprana, el cual permite su prevención, progresión y canalización.

Finalmente, las insuficiencias detectadas en el presente estudio responden a dos aspectos; la primera está relacionada con la identificación de las causas particulares de cada alumno diagnosticado con un indicador de conductas suicidas, este desconocimiento limita la comprensión eficiente para su pertinente atención. La segunda insuficiencia está en la diversificación en el uso de la prueba, es decir, aplicar del mismo instrumento de medición en las seis condiciones de evaluación puede influir en que las repuestas de los participantes no sean consistentes según los criterios de validación interna (Rodríguez, 2019). Pese a ello, el análisis diacrónico de la conducta suicida y su grado de recurrencia permite pronosticar su nivel de emergencia.

Cierre

El fenómeno del suicidio es un tema ampliamente abordado; sin embargo, su grado de relevancia se encuentra determinada a partir de circunstancias particulares que permiten clarificar su origen, no tan solo desde el terreno de la intervención psicológica, sino las causas del fenómeno deben comprenderse desde la multidisciplinariedad científica, que atienden el diagnóstico, rehabilitación y progresión de conductas suicidadas. Por otra parte, se considera pertinente aclarar que el tipo de conducta suicida nace a partir no tan solo de una ruptura psicológica del individuo, en donde su autoconcepto, estrategias de resiliencia o regulación emocional se encuentran altamente afectados;

sino que las mismas rupturas pueden extenderse a crisis institucionales o condiciones biológicas que hacen que el fenómeno sea más complejo de atender y prevenir.

Asimismo sabemos que el factor de estrés y los grados de su intensidad pueden propiciar la recurrencia de indicadores suicidas, sin ser necesariamente determinantes; sin embargo, este dato puede ser fundamental principalmente porque la exposición constante de las fuerzas productivas de la vida utilitaria pueden propiciar una condición de estrés, en donde el sujeto compromete su equilibrio mental ante situaciones conflictivas que exceden los recursos psicológicos que los alumnos poseen: por lo tanto, dentro de las actividades académicas cotidianas se requiere promocionar estrategias psicológicas para su pertinente atención, como en el caso de tiempos de esparcimiento, promoción de actividades deportivas, psicoterapia individualizada, atención de adicciones y actividades psicológicas grupales.

Referencias

Amador, J. E. D., Salinas P. M. X. & Pérez L. C. J. (2018). Consumo de sustancias psicoactivas y riesgo suicida en usuarios de centros de rehabilitación de acciones. *Revista Humanismo y cambio social, 5* (11). 39-51, https://dialnet.unirioja.es/servlet/articulo?codigo=9113531.

Azúa, F. E., Rojas, C. P. y Ruiz, P. S. (2020). Acoso escolar (bullying) como factor de riesgo de depresión y suicidio. *Revista chilena de pediatría. 91* (3). 1 – 12. http://dx.doi.org/10.32641/rchped.v91i3.1230.

Bertolón, G. M. (2022). Homicidio y suicidio: biología y psicopatología. *Psicosomática y psiquiatría. 22*. 30-39. https://doi.org/10.34810/PsicosomPsiquiatrnum210501.

Builes, C. M. V., Anderson, G. T. M., Ramírez, Z. A. y Arango, A. B. E. (2014). Suicidio en mujeres gestantes: vivencias y redes de apoyo para las familias que les sobreviven. *Facultad Nacional de Salud Pública. 32* (3). 332-339.

Carbonell, C. E. (2007). Tiempo y suicidio. Contribución antropológica a una discusión transdisciplinar. *Gaceta de Antropología. 23* (1). 1-16. http://hdl.handle.net/10481/7055.

Corea, D. C. T. (2021). La depresión y su impacto en la salud pública. *Revista Médica Hondureña. 89* (1). 51-61. DOI: https://doi.org/10.5377/rmh.v89iSupl.1.12047

Córdova. O. M. & Rosales, P. C. J. (2012). Confiabilidad y validez de constructo de la escala de ideación suicida de Beck. *Alternativas en Psicología. 16* (26). 16-25. ISSN 1405-339X

Dávila, C. A. C. & Pardo, M. M. (2020). Estudio de la carga de la mortalidad por suicidio en México. 1990-2017. *REV BRAS EPIDEMIOL. 22*. 1-14. https://doi.org/10.1590/1980-549720200069.

De La Estrella, R. (2010). Suicidio en instituciones psiquiatrías. 1998-2007. *Revista Colombiana de Psiquiatría. 30* (2). 268-290. ISSN 0034-7450.

De Jesús, R. A. Noalles, D. A. y Rujas, A. A. (2019). Serotonina: un neurotransmisor que impacta nuestras emociones. *RD-ICUAP. 13*. ISSN: 2448-5829.

DSM-5 (2015). Manual diagnóstico y estadístico de los trastornos mentales. *Editorial Médica Panamericana*. EEUU. ISBN 9788498358100.

Durkheim, E. (2004). El suicidio. (8° reimpresión). México. Ediciones Coyoacán.

Faura. G. J. Calvete, Z. E. y Orue, S. I. (2021). Autolesión no suicida: conceptualización y evaluación clínica en población hispanoparlante. *Papeles del Auto psicólogo. 42* (3). 207-214. https://dx.doi.org/10.23923/pap.psicol.2964.

Fonseca, F. E. & Halabia, A. S. (2024). Sobre la conducta suicida y las conductas adictivas. *Adicciones, 36* (2). 121-128, https://doi.org/10.20882/adicciones.2074.

Fonseca, F. E. & Pérez, A. A. (2021). Evaluación de la conducta suicida en adolescentes; a propósito de la Escala Paykel de Suicidio. *Papeles del psicólogo. 41*. (2). 106-115. https://dx.doi.org/10.23923/pap.psicol2020.2928.

García, B. E. & Aroca, F. (2014). Factores de riesgo de la conducto suicida asociados a trastornos depresivos y ansiedad. *Salud Mental. 37* (5). 373-380. ISSN: 0185-3325.

García, H. J., García, P. H., González, G. M., Barrio, M. S. y García, P. R. (2019). ¿Qué es suicidio?. Problemas clínicos y conceptuales. *Apuntes de psicología. 37* (2). 91-100. https://hdl.handle.net/11441/115113.

Gil, L. D. M., Fernández, G. O. A. & Bergero M. T. (2020). Perfil sociodemográfico asociado a la ideación suicida de las personas transexuales. *Información psicológica. 120*. 93-105 Doi: 10.14635/IPSIC.2020.120.3

Goñi, S. A., López, G. J. J. y Azcarate, J. L. (2019). Comparación entre los intentos de suicidio leve, grave y las muertes por suicidio en Navarra. *Clínica y Salud. 30* (2). 109-113. Doi: 10.5093/clysa2019a9.

González, M. S. Díaz, M. A., Ortiz, L. S., González, C. & González, N. J. J. (2000). Características psicométricas de la Escala de Ideación Suicida de Beck (ISB). *Salud Mental. 23* (2). 21-30. http://www.redalyc.org/articulo.oa?id=58222304.

Granados C. A. J., Gómez, L. O., Islas R. I. M., Maldonado P. G. H. Martínez M. F. & Michelle P. A. (2020). Depresión, ansiedad y conducta suicida en la formación médica en una universidad en México. *Investigación en Educación Médica. 9* (5). 65-73. https://doi.org/10.22201/facmed.20075057e.2020.35.20224.

Hinojosa, D. C. M., Cabrera, A. R. A., Procel, V. M. & Jarrin, I. A. D. (2022). El estrés laboral ¿es posible que genera condiciones suicidas en los trabajadores? *Polo del conocimiento. 7* (7). 274-293. DOI: 10.23857/pc.v7i7.

Jerónimo, A. M., Pinar, S., González, M. A. Agnés, S. B. M., Xavier, A. L. J., Martina, M. L., Acena, R., Pérez, V. & Córcoles, D. (2021). Intentos e ideas de suicidio durante la pandemia

por COVID-19 en comparación con los años previos. *Revista de psiquiatría y salud mental. 11* (4). 1-7. Doi: 10.1016/j.rpsm.2021.11.004.

Kaplan, C. V., Szapu, E. (2019). Jóvenes y subjetividad negada: Apuntes para pensar la intervención socioeducativa sobre prácticas autolesivas y suicidio. *Psicoperspectivas. 18* (1). 1-11. http://dx.doi.org/10.5027/psicoperspectivas-vol18-issue1-fulltext-1485.

Luna, C. M. y Cervantes, D. C. A. (2018). Adolescentes en riesgo: factores asociados con el intento de suicidio en México. *Revista Gerencia y Políticas de Salud. 17* (34). 606-616. http://dx.doi.org/10.32641/rchped.v90i6.1012.

O'Connor, R. C. & Nock, K. M. (2014). The Psychology of suicidal behavior. *The Lancet Pshychiatry. 1*. 73-85. Doi:10.1016/S2215-0366(14)70222-6.

Miranda, I. C., Cubillas, R. J., Roman, P. R. & Valdez, E. A. (2009). Ideación suicida en población escolarizada infantil: factores psicológicos asociados. *Salud mental. 35*. 495-502. ISSN: 0185-3325.

Muñetón, B. M. J. & Vasquez, A. Y. (2018). Diseño y validación de una escala para evaluar el riesgo suicida (ERS) en adolescentes colombianos. *Universitas Psychologica. 17* (4). DOI: https://doi.org/10.11144/Javeriana.upsy17-4.dvee.

Pabón, Y. A (2021). Intentos de suicidio y trastornos mentales. *Revista Habana de Ciencias Médicas. 20* (4). 1-9. ISSN 1729-519X.

Quevedo, B. R. & Palomares, R. J. (2023). Evolución histórica del perfil profesional de las víctimas de suicidio en España (1943-2005). *Revista Internacional y Comparada de relaciones laborales y derecho del empleo. 11* (3). 289-315. ISSN 2282-2313.

Quintero, A., Sánchez, L. Agrego, J & Oviedo, L. (2021). Parasuicidio en púberes y adolescentes escolarizados. *Revista Espacios. 42* (07). 1-21. Doi: 10.48082/espacios-a21v42n07p01

Rangel, V. J. N. & Jurado, C. S. (2020). Definición de suicidio y de los pensamientos y conductas relacionadas con el mismo: una revisión. *Psicología y Salud. 32* (1). 39-48. https://doi.org/10.25009/pys.v32i1.2709.

Rodríguez, F. M. V. (2019). Control de sesgos: piedra angular de la validez interna en la investigación para la salud. *Revista científica del instituto nacional de salud. 2* (2). 181-187. DOI: 10.5377/alerta.v2i2.7904.

Soto, S. A., Villaro, G. P. & Véliz, B. A. (2020). Factores que intervienen en riesgo suicida y parasuicida en jóvenes Chilenos. *Propósitos y representación. 8* (3). http://dx.doi.org/10.20511/pyr2020.v8n3.672.

Valdevila, F. J. A., Ruiz, B. O., Orellana, R. C., Valdevila, S. R., Fabelo, R. J. e Iglesias, M. S. (2021). Conducta suicida y dualidad en trastornos por consumo de sustancia en drogodependientes ecuatorianos. *Revista información científica. 100* (5). 1-15, https://www.redalyc.org/articulo.oa?id=551769293007.

COVID-19, EDUCACIÓN Y SALUD: REPRESENTACIONES SOCIALES DEL REGRESO PRESENCIAL A CLASES DE NIVEL MEDIO SUPERIOR EN POSTPANDEMIA

COVID-19, EDUCAÇÃO E SAÚDE: REPRESENTAÇÕES SOCIAIS DO RETORNO PRESENCIAL ÀS AULAS DO ENSINO MÉDIO NO PÓS-PANDEMIA

COVID-19, EDUCATION AND HEALTH: SOCIAL REPRESENTATIONS OF THE IN-PERSON RETURN TO HIGH SCHOOL CLASSES IN THE POST-PANDEMIC

Luis Enrique Hernández Gamundi
Angélica Heredia Sánchez
Donovan Casas Patiño
Alejandra Rodríguez Torres
Adán Flores Garnica[1]

Resumen

Las medidas preventivas establecidas para enfrentar la elevada tasa de contagios y mortalidad por COVID-19 (CV19), llevó al sistema educativo a adoptar estrategias que garantizaran continuar las clases sin exponer la salud de la comunidad escolar. Fue el caso del uso masivo de la educación virtual (EV); cabe mencionar que, tras el regreso presencial a clases, se percibieron cambios significativos en el rendimiento académico y conductual de los discentes, situación aparentemente resultada del Choque Cultural (CC) por los efectos de la CV19. Objetivo**:** Analizar desde las Representaciones Sociales (RS) el impacto del choque cultural generado en alumnos de 6º semestre de bachillerato tras el regreso presencial a clases. Método**:** Se realizó un estudio cualitativo, transversal en una preparatoria ubicada en la Región Oriente del Estado de México, participando 33 alumnos de 6º semestre, mayores de edad, quienes respondieron una entrevista semiestructurada (ESE). Resultados: Los discursos reflejan las RS de los alumnos tras el regreso presencial a clases y su capacidad de resiliencia para enfrentar las incidencias en salud y educación

[1] Todos de la Universidad Autónoma del Estado de México.

causadas por la CV19: *"… una enfermedad que no llegaría a México…. preocupación por lo que pudiera pasar con mi familia…" "… me negaba a vacunarme por lo que escuchaba… esperanza para proteger la vida…" "… la adaptación fue complicada… significa mucho vivir una pandemia y regresar a la escuela… afortunado de tener vida … alegría al volver a la escuela…" "… en presencial todo se tornó nuevo era como pisar una escuela por primera vez…".* Conclusiones**:** Conocer a través de las RS los factores negativos que desde el choque cultural afectan el rendimiento académico y conductual de los alumnos permite establecer protocolos para su manejo desde la resiliencia.

***Palabras clave*:** Covid-19, Choque Cultural, Educación, Salud y Resiliencia.

Resumo

As medidas preventivas estabelecidas para enfrentar o alto índice de infecções e mortalidade por COVID-19 (CV19) levaram o sistema educacional a adotar estratégias que garantissem a continuidade das aulas sem expor a saúde da comunidade escolar, foi o caso do uso massivo da educação virtual (EV); Vale ressaltar que após o retorno às aulas presenciais foram percebidas mudanças significativas no desempenho acadêmico e comportamental dos alunos, situação aparentemente decorrente do Choque Cultural (CC) devido aos efeitos da CV19. **Objetivo:** Analisar a partir das Representações Sociais (RS) o impacto do choque cultural gerado em alunos do 6º semestre do Ensino Médio após o retorno às aulas presenciais. **Método:** Foi realizado um estudo qualitativo, transversal, em uma escola secundária localizada na Região Leste do Estado do México, envolvendo 33 alunos do 6º semestre, maiores de idade, que responderam a uma entrevista semiestruturada (ESE). **Resultados:** As falas refletem as RS dos alunos após o retorno às aulas presenciais e sua capacidade de resiliência para enfrentar os incidentes na saúde e na educação causados pela CV19: *"…uma doença que não chegaria ao México…. preocupação com o que poderia acontecer com minha família…" "…recusei-me a me vacinar pelo que ouvi… espero proteger a vida…" "…a adaptação foi complicada…significa muito viver uma pandemia e voltar à escola… sorte de ter vida… alegria ao voltar para a escola…" "… pessoalmente tudo se tornou novo, foi como entrar pela primeira vez numa escola…".* **Conclusões:** Conhecer através das RS os fatores negativos que afetam o desempenho acadêmico e comportamental dos alunos

provenientes do choque cultural permite-nos estabelecer protocolos para a sua gestão a partir da resiliência.

***Palavras-chave*:** Covid-19, Choque Cultural, Educação, Saúde e Resiliência.

Summary

The preventive measures established to face the high rate of infections and mortality from COVID-19 (CV19) led the educational system to adopt strategies that guaranteed continuing classes without exposing the health of the school community, this was the case of the massive use of virtual education (EV); It is worth mentioning that after returning to classes in person, significant changes were perceived in the academic and behavioral performance of the students, a situation apparently resulting from Culture Shock (CC) due to the effects of CV19. **Objective:** To analyze from Social Representations (SR) the impact of the cultural shock generated in 6th semester high school students after returning to classes in person. Method: A qualitative, cross-sectional study was carried out in a high school located in the Eastern Region of the State of Mexico, involving 33 6th semester students, of legal age, who responded to a semi-structured interview (ESE). **Results:** The speeches reflect the SR of the students after returning to classes in person and their capacity for resilience to face the incidents in health and education caused by CV19: *"... a disease that would not reach Mexico…. concern about what could happen to my family…" "... I refused to get vaccinated because of what I heard… hope to protect life…" "... the adaptation was complicated… it means a lot to live through a pandemic and return to school… lucky to have life… joy when returning to school…" "... in person everything became new, it was like stepping into a school for the first time…".* **Conclusions:** Knowing through the SR the negative factors that affect the academic and behavioral performance of students from cultural shock allows us to establish protocols for their management from resilience.

***Keywords*:** Covid-19, Culture Shock, Education, Health and Resilience.

Introducción

En los albores de la presente década, la humanidad enfrentó de manera abrupta una enfermedad mortal generada por una cepa de coronavirus, el SARS-CoV-2, causante de la enfermedad hoy

ampliamente diagnosticada como COVID-19, la cual en breve tiempo le llevó a adquirir la categoría de Pandemia.

La (Organización Mundial de la Salud [OMS], 2020, párr. 20) se indicó: "profundamente preocupada por los alarmantes niveles de propagación de la enfermedad, por su gravedad y por los niveles también alarmantes de inacción, la OMS determina en su evaluación que la COVID-19 puede caracterizarse como una pandemia". Ante tal emergencia sanitaria individuos y colectivos sociales a fin de preservar su salud, reestructuraron sus actividades, medios y formas de vida. Adaptándose abruptamente a una "nueva normalidad". En este sentido tanto las complicaciones como las implicaciones propias de la CV19 representaron enormes retos para las disímiles estructuras y sistemas de gobierno como son educación y salud entre otras.

Al respecto la (Comisión Económica para América Latina y el Caribe [CEPAL], 2020, párr.1) señaló que "la pandemia de enfermedad por coronavirus (COVID-19) ha provocado una crisis sin precedentes en todos los ámbitos. En la esfera de la educación, esta emergencia ha dado lugar al cierre masivo de las actividades presenciales de instituciones educativas" alterando así la praxis en el ejercicio del sistema educativo en todo nivel incluida la Educación Media Superior (EMS) situación que llevó a implementar estrategias de educación que, simultáneamente garantizaran la salud de la comunidad escolar y el cumplimiento adecuado de los planes y programas educativos en vigor. De esta manera, la Educación a Distancia o Virtual (EV) se convirtió en una valiosa metodología de enseñanza-aprendizaje, pues cubría los parámetros solicitados; sin embargo, aun cuando la EV otorgó las debidas garantías para continuar con la educación, reduciendo la probabilidad de contagios por CV19, las características propias de esta metodología, generaron un cambio en la interacción social de los estudiantes.

El referido cambio en la interacción social, podría estar relacionado con la disminución en el rendimiento académico, cambios en la conducta e incremento en crisis de ansiedad, percibidos tras el regreso presencial a clases en alumnos que cursan la EMS en el Oriente del Estado de México; desde luego, cabe establecer que las condiciones de bajo rendimiento académico y conductual, existían previo a la pandemia, sin embargo, en la población que integra esta muestra, su prevalencia antes de la pandemia, no reflejaba una considerable problemática educativa; por lo antes mencionado resulta idóneo realizar un acercamiento al objeto de estudio desde el paradigma del choque cultural y la resiliencia en tiempos de pospandemia.

Metodología

Se realizó un estudio de enfoque cualitativo y corte transversal mediante la aplicación de una entrevista semiestructurada en una escuela preparatoria pública ubicada en el Oriente del Estado de México. Se utilizó un muestreo intencional determinándose el tamaño de la muestra por saturación del discurso, resultando 33 alumnos de 6º semestre, de ambos sexos, mayores de edad; cuya participación fue voluntaria previo otorgamiento de su consentimiento informado. La investigación fue estructurada a partir de tres fases, mismas que abarcaron una revisión sistemática de la bibliografía relacionada con la temática en estudio, un acercamiento desde la fenomenología hermenéutica a la comunidad estudiantil, así como el diseño del instrumento de investigación y su validación por juicio de expertos.

Respecto a lo anterior, es imprescindible citar que el instrumento de investigación final fue integrado a partir de 3 categorías de análisis: el análisis de los datos cualitativos obtenidos fueron codificados mediante el *software* Iramuteq v 0.7 Alpha 2; cabe señalar que realizar la investigación en el referido plantel, responde a la accesibilidad de su ubicación geográfica, la cual permite que se incorporen alumnos provenientes de los diversos municipios del oriente mexiquense, por otro lado, a su vasta matrícula escolar, actualmente conformada por 1904 alumnos de los diversos semestres, 95 académicos y 24 administrativos lo que ofrece un extenso universo de estudio.

Teorización de la problemática

El sustento teórico de la presente investigación reside en la peana de las Representaciones Sociales, teoría postulada por Serge Moscovici, siendo él quien las definiera como esa

> modalidad particular del conocimiento, cuya función es la elaboración de los comportamientos y la comunicación entre los individuos (…) por lo que cada sociedad e individuo entiende y comprende algo y lo explica conforme a su propia percepción y experiencia. (Moscovici, 1961; Gutiérrez. 2016; Flores et al., 2019, p.77).

Desde esta panóptica, una representación puede ser entendida como “un corpus organizado de conocimientos y una de las actividades psíquicas gracias a las cuales los hombres hacen inteligible

la realidad física y social" (Mora, 2002, p.7), a partir de la propuesta teórica de Moscovici, las RS han sido estudiadas por diversos autores, encontrando entre sus principales exponentes a Denise Jodelet y Jean Abric.

Contexto y Conceptos: Acercamiento al objeto de estudio

Analizar los efectos del CC en el rendimiento académico, conducta y estado de salud de los alumnos de 6 º semestre, así como las medidas resilientes adoptadas para hacerles frente, se debió a que fueron precisamente estos alumnos quienes en el mes de marzo de 2020 cursaban las primeras semanas del 2º semestre, cuando a causa de la pandemia por CV19 la (Secretaría de Educación Pública [SEP]) a fin de preservar la salud de la comunidad escolar, instituyó inéditas medidas preventivas como fue el cierre temporal de escuelas y espacios académicos; al respecto el Rojas (2021, p. 104) menciona lo siguiente: "En México, la suspensión oficial de clases ocurrió el 23 de marzo de 2020 sustentado en el Acuerdo número 02/03/2020. Se manifestó la necesidad de esta acción derivado de la covid-19, pero las recomendaciones pedagógicas fueron insuficientes". Situación por la cual se redireccionó la enseñanza tradicional en formato presencial hacia la EV, lo que, sin duda, resultó benéfico para la preservación y mantenimiento del estado de salud ante los estragos causados por la CV19.

Sin embargo, la EV restringió el contacto e interacción personal; por lo que, al retornar a los entornos de la educación presencial, tras una ausencia de casi dos años; estos jóvenes ahora estudiantes de 6º semestre, se enfrentaron a un nuevo escenario cuyas implicaciones requirieron de un proceso de reintegración a la presencialidad, mismo que les permitiese afrontar los efectos del choque cultural en tiempos y espacios de pospandemia, donde, además de existir la latencia de un contagio por CV19, imperó la necesidad de readaptarse a aquellos aspectos metodológicos y operacionales del Proceso Enseñanza-Aprendizaje como lo son: los horarios escolares, trabajo colaborativo, aumento del rigor académico e, incluso, la necesidad de fortalecer vínculos de socialización. Esto no solo representó un cambio marcado en la rutina de la vida diaria de los discentes, sino que también, en su conducta y rendimiento académico.

Por lo anterior, y con el fin de otorgar un mayor acercamiento al contexto de esta investigación, se presenta en orden alfabético el marco conceptual que le sustenta.

En este sentido se inicia con el Choque Cultural, el cual es un paradigma acuñado a finales de los años 50's del pasado siglo, la cual hace alusión a la ansiedad generada al perder el sentido de qué hacer, cuándo hacer, o cómo hacer algo tras la incorporación a un ambiente o entorno nuevo; el CC puede describirse como esa incomodidad física y emocional que se hace presente cuando se llega a vivir o se integra a un lugar diferente al de origen; esto puede llevar a experimentar tristeza, soledad, melancolía, preocupación por la salud, dolores, insomnio, cambios de temperamento, depresión, irritabilidad (Deza, 2001; Gutiérrez, 2016, p. 255).

Ahora bien, por lo que respecta a la Educación, esta es definida por el modelo educativo para la educación obligatoria de la SEP (2017a, p. 205) como "un proceso social por medio del cual se transmiten la cultura, valores, conocimientos, prácticas, creencias, tecnologías, usos o costumbres que, en una sociedad se consideran valiosos".

Con relación a la Educación Media Superior la SEP (2017b, p.45) indica la EMS es aquella donde se forman personas con conocimientos y habilidades que les permite desarrollarse en sus estudios superiores o en el trabajo y de forma amplia en la vida, donde los jóvenes adquieren actitudes y valores que tienen un impacto positivo en su comunidad y en la sociedad. De igual forma, es importante señalar que la EMS está integrada por el "bachillerato, sus equivalentes y la educación profesional que no solicita bachillerato. Las modalidades de este tipo incluyen bachillerato general, bachillerato tecnológico y profesional técnico" (Subsecretaría de Educación Media Superior [SEMS], 2019; Arellano y Ortiz, 2022).

Por lo que atañe a la Educación Virtual (Lara, 2002; Martínez, 2008) refieren que la EV "eleva la calidad de la enseñanza-aprendizaje respeta su flexibilidad o disponibilidad se puede canalizar para tiempos y espacios variables (…) logra mayor reconocimiento con la tecnología a través de los métodos asincrónico, sincrónico y autoformación" así también Silva (2010) señala que en México, la EV es nueva e innovadora y generó un rompimiento con el esquema tradicional de la educación contemporánea, donde la familia es elemento fundamental para educar, mientras que las instituciones, deben enseñar y ofertar infraestructura tecnológica-científica para el aprendizaje.

Con relación al término Escuela, Gutiérrez (2016) señala que esta representa un espacio social y cultural donde confluyen diversos colectivos sociales permeado cada uno por sus características y saberes culturales, resultando la multiculturalidad esencial para comprender la ambivalencia de la educación formal, la cual en ocasiones consigue insertarse en los colectivos,

mientras en otras es refutada principalmente, cuando existe la posibilidad de una interacción entre los grupos sociales en ámbitos y espacios geográficos diferentes al de origen. Cabe mencionar que esta investigación, abordó la forma de enfrentar la problemática referida por los alumnos a través de la resiliencia misma que es definida por la (Real Academia Española [RAE], 2022, párr.2) como la "capacidad de un ser vivo frente a un agente perturbador o un estado o situación adversos".

Otra interpretación de la resiliencia la detalla como "un proceso dinámico que tiene como resultado la adaptación positiva en contextos de gran adversidad" (Luthar y Cushing, 1999, p.543; García y Domínguez, 2013, p.66), en este mismo sentido García y Domínguez (2013, p.66) concluyen que en su conjunto las definiciones de resiliencia "enfatizan en características de los sujetos tales como: habilidad, adaptabilidad, baja susceptibilidad, enfrentamiento efectivo, capacidad, competencia, resistencia a la destrucción, conductas vitales positivas, temperamento especial y habilidades cognitivas, todas desplegadas frente a situaciones estresantes que les permiten superarlas". De igual forma Ortunio y Guevara (2016, p. 97) refieren que la resiliencia en educación:

> Abre un nuevo campo creativo en donde se entregan las herramientas necesarias para que el estudiante mantenga la curiosidad, use la lógica, sea responsable de sí mismo, se adapte con tolerancia a la frustración, maneje cierto control para afrontar la adversidad, recupere el equilibrio, la armonía personal y de esta manera logre salir fortalecido.

Resultados y Discusión

Los datos cualitativos permiten conocer el sentir, pensar y saberes a través de los discursos emanados de propia voz de los actores sociales, en este orden de ideas, Bustos (2021, p.2) señala, de manera enfática, que la nueva educación será aquella que "rescatemos de la experiencia que nos dejó el 2020, a partir de la recuperación de los aprendizajes propuestos a lo largo de la pandemia; lo vivido debe ser una valiosa herramienta de la cual valernos" por lo antes mencionado, el análisis e interpretación de los resultados, fue generado a partir de las RS percibidas por el impacto de la CV19 en la educación y salud de los discentes de 6º semestre de EMS oriundos de la Región Oriente del Estado de México, tras el regreso presencial a clases, lo que implicó una revisión de las categorías analíticas COVID-19: Educación, Salud y Enfermedad, Educación

Media Superior: Choque Cultural y Regreso presencial a clases en pospandemia: Resiliencia (Ver Cuadro 1).

Cuadro 1. Discursos representativos de las Categorías de Análisis.

COVID-19: Educación, Salud y Enfermedad.	
"*… sería una enfermedad más, algo pasajero que no llegaría a México… que pasaría pronto el tema de los contagios… que no iba a ser grave…*" [A3, A4, A6, A7, A10, A11, A14, A15, A17, A18, A19, A24, A26, A27, A29, A32], "*… íbamos a enfermar la mayoría… que tarde o temprano llegaría a nuestro país…*" [A1, A2, A8, A12, A13, A16, A21, A22, A23, A25, A28, A30, A33], "*… miedo y angustia… preocupación por lo que pudiera pasar con mi familia…*" [A5, A7, A16, A20, A21, A23, A24, A30, A32]. "*… vivo con ansiedad… entre en pánico me daba miedo salir…*" [A3, A4, A5, A7, A9, A12, A15, A19, A20, A21, A24, A25, A28, A29, A31], "*… había esperanza para prevenir el contagio… alivio… felicidad para proteger la vida… íbamos a estar protegidos… un avance en la preservación de la vida…*" [A1, A3, A4, A6, A7, A8, A9, A10, A12, A14, A17, A27, A28, A30, A31], "*… al vacunarme me sentí muy mal… miedo que pudiese traer consecuencias… provoca alteraciones a nuestro organismo… no sirve yo me enfermé más… me negaba a vacunarme por lo que escuchaba…*" [A5, A16, A19, A20, A21, A22, A23, A24, A25, A29, A32, A33].	
Educación Media Superior: Choque Cultural.	**Regreso presencial a clases en pospandemia: Resiliencia.**
"*… la adaptación fue bastante complicada… un gran reto… difícil por las medidas que se tienen… estresante… me siento nervioso y empecé con ataques de miedo y ansiedad…*" [A4, A5, A6, A7, A8, 9, A10, A11, A12, A14, A17, A25, A32, A33], "*… me siento afortunado de tener vida… significa mucho el vivir una pandemia y regresar a la escuela…*" [A18, A19, A23, A28, A30], [A1, A2, A13, A16, A20, A21, A22, A24, A26, A27, A29, A31], "*… ha sido bueno, me adapte a lo presencial… mejor que en línea, aprendo más en presencial… bueno, aunque ha sido difícil…*" [A1, A3, A4, A9, A11, A12, A14, A15, A19, A20, A22, A23, A24, A25, A27], "*… en presencial me cuesta trabajo… han bajado mis calificaciones… regular, me sigo adaptando a lo presencial… mi rendimiento no es bueno…*" [A5, A6, A7, A8, A10, A17, A18, A21, A26, A28, A29, A30, A31, A33], "*… soy más sociable… soy más responsable… soy más maduro como persona… me volví menos sociable por la sana distancia… me volví más inseguro …*" [A1, A3,	"*… me gustó la idea ya que no conocía a mis compañeros… emoción pues el contacto humano hacía falta… recuperar mi vida… mucha alegría al volver a la escuela…*" [A1, A2, A3, A4, A5, A6, A7, A9, A14, A17, A18, A21, A22, A27, A28, A31], "*… miedo por la posibilidad de contagio, por el rezago que se tenía en línea…*" [A8, A10, A11, A12, A13, A16, A20, A23, A25, A26, A30, A32, A33], "*… al contrario me gusto regresar a clases… fue complicado, pero no me quito las ganas de estudiar… me dan más ganas de estudiar después de tanto tiempo de encierro…*" [A4, A5, A7, A8, A9, A11, A14, A15, A17, A18, A21, A22, A23, A24, A26, A27, A28, A29, A30, A32, A33], "*… me siento más seguro… me pareció muy responsable por parte de la escuela… sentí alivio al saber que la escuela nos estaba cuidando… sí, cumplen con todas las medidas…*" [A1, A3, A4, A6, A8, A9, A10, A14, A15, A16, A18, A21, A26, A28, A29, A31, A32, A33], "*… inseguridad y ansiedad… ansiedad hacia las tareas… cuando regresé a clases asistí al*

A4, A7, A8, A9, A11, A19, A21, A25, A33], *"… todo se tornó nuevo por dicho regreso… era como pisar una escuela por primera vez…"* [A14, A16, A18, A21, A24].	*psicólogo…"* [A2, A3, A12, A21, A24, A25], *"… mis amigos son mi tratamiento emocional… ninguna, pude llevar mis emociones controladas…"* [A4, A5, A6, A7, A8, A9, A10, A11, A13, A14, A15, A16, A22, A26, A27, A28, A30, A31, A32, A33].

Fuente: Elaboración propia.

a) COVID-19: Educación, Salud y Enfermedad.

Esta categoría aborda las RS de los discentes, en torno a esa ignota patología que causaba estragos en la salud de las personas en otros países, así como sus efectos colaterales, vistos desde la panóptica de la educación respecto a la salud y la enfermedad; ahora bien, respecto a lo que pensaron la primera ocasión que escucharon sobre esa extraña enfermedad, indicaron que "*… sería una enfermedad más, algo pasajero que no llegaría a México… que pasaría pronto el tema de los contagios… que no iba a ser grave..."* [A3, A4, A6, A7, A10, A11, A14, A15, A17, A18, A19, A24, A26, A27, A29, A32], sin embargo, hubo quien señaló que sería un grave problema de salud pues *"… íbamos a enfermar la mayoría… que tarde o temprano llegaría a nuestro país… que podría ser una enfermedad mortal…"* [A1, A2, A8, A12, A13, A16, A21, A22, A23, A25, A28, A30, A33]. Maguiña, et al., (2020) indicaron que la pandemia fue causada por una cepa mutante de coronavirus, surgida a fines del año 2019 en la República Popular de China (Wuhan, provincia de Hubei) donde fueron reportados 27 casos de neumonía atípica, situación que llevó a la humanidad a enfrentar una inexorable crisis en diferentes ámbitos; siendo el 7 de enero 2020 cuando el Ministerio de sanidad de China identificó un nuevo coronavirus.

Con relación a su experiencia al ser diagnosticados con CV19 quienes padecieron esta enfermedad comentaron *"… miedo y angustia… preocupación por lo que pudiera pasar con mi familia…"* [A5, A7, A16, A20, A21, A23, A24, A30, A32]. *"… vivo con ansiedad… entre en pánico me daba miedo salir…"* [A3, A4, A5, A7, A9, A12, A15, A19, A20, A21, A24, A25, A28, A29, A31]. Así mismo respecto a lo que pensaron los alumnos cuando les dijeron que deberían vacunarse contra la CV19 señalaron que *"… había esperanza para prevenir el contagio… alivio… felicidad para proteger la vida… íbamos a estar protegidos… un avance en la preservación de la vida..."* [A1, A3, A4, A6, A7, A8, A9, A10, A12, A14, A17, A27, A28, A30, A31], mientras otras voces señalaron *"… al vacunarme me sentí muy mal… miedo que pudiese traer consecuencias… provoca alteraciones a nuestro organismo… no sirve yo me*

enfermé más… me negaba a vacunarme por lo que escuchaba…" [A5, A16, A19, A20, A21, A22, A23, A24, A25, A29, A32, A33].

En este sentido, Bustos (2021, p.2) plantea la probabilidad de un escenario intermitente, donde asistan a clases quienes puedan o así lo deseen, señalando la necesidad de esbozar modelos flexibles que faciliten una alternancia entre la modalidad virtual y presencial, a fin de cursar las diversas etapas de vacunación hasta que sea posible adquirir la inmunidad de rebaño; por otra parte, la ambivalencia registrada a partir de los discursos, devela la importancia que atañe al ámbito educativo, pues aun cuando la CV19 es un problema de vertiente biológica la educación es capaz de generar una respuesta social adaptativa que permite adquirir, aplicar e incluso transmitir conocimientos relacionados con la salud y la enfermedad. Al respecto el modelo educativo salud en tu escuela, elaborado conjuntamente por las (Secretarías de Educación Pública y Salud [SEP–SS], (2017) establece el vital papel que, en torno al cuidado de la salud, corresponde a la educación pues "contribuye al desarrollo de habilidades cognitivas que influyen en la forma de pensar y de tomar decisiones (…) estas intervienen en la capacidad para obtener y comprender información médica y actuar con base en ella".

b) Educación Media Superior: Choque Cultural.

Esta categoría se enfoca en situaciones relacionadas con la conducta, el rendimiento académico y el aumento en padecimientos de índole físico y emocional, percibidos en algunos discentes de 6º semestre tras el regreso presencial a clases, para tal efecto se planteó el supuesto hipotético, donde tales situaciones podrían estar relacionadas con las diferentes medidas establecidas para preservar la salud durante la fase álgida de la pandemia, como fue el confinamiento social, la sana distancia y la implementación de la EV por citar algunas. En este orden de ideas, respecto al significado de ser estudiante en tiempos de pospandemia comentaron: *"… la adaptación fue bastante complicada… un gran reto… difícil por las medidas que se tienen… estresante… me siento nervioso y empecé con ataques de miedo y ansiedad…"* [A4, A5, A6, A7, A8, 9, A10, A11, A12, A14, A17, A25, A32, A33], *"… me siento afortunado de tener vida… significa mucho el vivir una pandemia y regresar a la escuela…"* [A18, A19, A23, A28, A30], [A1, A2, A13, A16, A20, A21, A22, A24, A26, A27, A29, A31].

Los discursos observados en el párrafo anterior, hacen alusión a la identidad del discente y su educación en tiempos de pospandemia, al respecto Flores et al., (2019, p.82) refieren que la

identidad es entendida como "la construcción que define la relación desde el colectivo en relación al otro o algo, exclusivamente por sus características vividas (atribuidas por la interacción cultural) como el género, raza, religión, etc." interacción a la que se pueden aunar categorías como educación, salud y enfermedad. Asimismo, esta categoría planteó interrogantes relacionados con aquellos cambios percibidos tanto en el rendimiento académico, como en la conducta de los discentes tras el retorno a las aulas; cuyas respuestas oscilaron en dos sentidos permitiendo conocer que, para algunos jóvenes el cambio de modalidad *"… ha sido bueno, me adapté a lo presencial… mejor que en línea, aprendo más en presencial… bueno, aunque ha sido difícil…"* [A1, A3, A4, A9, A11, A12, A14, A15, A19, A20, A22, A23, A24, A25, A27], otros comentaron *"… en presencial me cuesta trabajo… han bajado mis calificaciones… regular, me sigo adaptando a lo presencial… mi rendimiento no es bueno…"* [A5, A6, A7, A8, A10, A17, A18, A21, A26, A28, A29, A30, A31, A33].

Por otra parte, los discursos de quienes percibieron cambios en su conducta, denotan cierta polarización pues manifestaron: *"… soy más sociable… soy más responsable… soy más maduro como persona… me volví menos sociable por la sana distancia… me volví más inseguro …"* [A1, A3, A4, A7, A8, A9, A11, A19, A21, A25, A33], algunos incluso, comentaron que su comportamiento se vio afectado debido a que *"… todo se tornó nuevo por dicho regreso… era como pisar una escuela por primera vez…"* [A14, A16, A18, A21, A24]. Por lo anterior, resulta imperativo retomar las aportaciones de Troncoso (2022. p. 95) quien señaló que, el regreso presencial a clases no cubrió las expectativas generadas en torno a ello, por el contrario, develó tanto las brechas socio-educacionales ligadas al retraso del currículo educativo y los aprendizajes como aquellas fisuras a nivel de desarrollo socioemocional, evidenciadas entre otros factores por: falta de empatía, ansiedad, síntomas depresivos y dificultades para resolución de conflictos

c) Regreso Presencial a Clases en Pospandemia: Resiliencia.

En esta categoría analítica se ahondó sobre el sentir de los alumnos al informarles que debían presentarse a clases presenciales, con un aforo del 100% de la comunidad escolar en tiempos de pospandemia; los discursos mostraron una dicotomía, señalando *"… me gustó la idea ya que no conocía a mis compañeros… emoción pues el contacto humano hacía falta… emoción debido a recuperar mi vida… mucha alegría al volver a la escuela…"* [A1, A2, A3, A4, A5, A6, A7, A9, A14, A17, A18, A21, A22, A27, A28, A31], sin embargo, también dijeron sentir *"… miedo por la posibilidad de contagio…*

miedo por el rezago que se tenía en línea…" [A8, A10, A11, A12, A13, A16, A20, A23, A25, A26, A30, A32, A33]. Cabe mencionar que, a causa de la CV19, se dejó de asistir a la escuela e interaccionar, quedando de lado, diversas actividades propias de la modalidad presencial; al respecto (Abufhele & Jeanneret, 2020; Troncoso 2022, p. 96) señalaron que los alumnos "estuvieron sin asistir a la escuela, a ver a sus amigos o realizar actividades de interés, frustrados por la limitación de liberación de energía o tensión, y con dificultades para la adaptación escolar".

Sobre la posibilidad de renunciar a sus estudios, los participantes de manera enfática mencionaron *"… al contrario, me gusto regresar a clases… jamás de hecho me gusta más presencial… fue complicado, pero no me quito las ganas de estudiar… me dan más ganas de estudiar después de tanto tiempo de encierro…"* [A4, A5, A7, A8, A9, A11, A14, A15, A17, A18, A21, A22, A23, A24, A26, A27, A28, A29, A30, A32, A33]. Respecto a la anterior discursiva (Caldera et al., 2016; Peñaloza et al., 2023, p. 2148) refieren que "frente a este fenómeno, la resiliencia es la capacidad de resistir la adversidad y recuperarse de sucesos vitales difíciles. Es el resultado de un equilibrio entre factores de riesgo, factores protectores y personalidad de cada individuo, funcionalidad y estructura familiar". Otro apartado de esta categoría abordó las medidas sanitarias implementadas en su escuela para reducir el riesgo de contagio señalando *"… me siento más seguro… me pareció muy responsable por parte de la escuela… sentí alivio al saber que la escuela nos estaba cuidando… sí, cumplen con todas las medidas…"* [A1, A3, A4, A6, A8, A9, A10, A15, A16, A18, A21, A26, A28, A29, A31, A32, A33].

A lo cual Bustos (2021, p. 1) indicó: "la pandemia nos obligó a poner en marcha nuevas formas de relacionarnos, de trabajar, de entretenernos y de aprender. Así, este ciclo lectivo (…) además de útiles escolares, también necesitará máscaras, tapabocas, alcohol en gel, sanitizantes, control de temperatura". El último aspecto explorado cuestionó si regresar a clases presenciales en pospandemia causó alguna condición que requiriese ayuda profesional; algunos respondiendo: *"… inseguridad… ansiedad hacia las tareas… cuando regresé a clases asistí al psicólogo…"* [A2, A3, A21, A24, A25], un número mayor refirió:*"… no, mis amigos son mi tratamiento emocional… ninguna, pude llevar mis emociones controladas…"* [A4, A5, A6, A7, A8, A9, A11, A14, A15, A16, A17, A18, A22, A26, A28, A29, A30, A31, A32, A33]. El confinamiento social, implementado a raíz de la CV19, generó que los estudiantes atraviesen "cuadros de estrés crónico con probables episodios sobrevenidos de estrés agudo, siendo un potencial factor dañino para el equilibrio emocional" (Ocampo & Correa, 2023; Quiroz et al., 2021; Varela et al., 2022; Peñaloza et al., 2023, p. 2147).

d) Asociación discursiva.

Los discursos obtenidos fueron codificados con el *software* Iramuteq v 0.7 Alpha2, esto permitió estructurar un esquema de interpretación asociativa, en cuyo núcleo encontramos la palabra Presencial observándose hacia la periferia una asociación directa con: Miedo, Enfermedad, Vida, Escuela, Emoción, Importante, Vacuna, Alivio, Regresar, Ayuda, Mucho, Conocer, Estudiantes, Adaptación, Difícil, Universidad, Medidas, Salud, Pandemia, Covid19, Contagios, Ansiedad, Cambio, Siento, Felicidad, Mejor, Nivel y Calificaciones; este dinamismo núcleo-periferia otorga, desde la subjetividad de los participantes, un vasto corpus de conocimiento el cual establece que tras el regreso presencial a clases, la COVID-19 generó en los discentes ansiedad ante la posibilidad de contagio, constituyéndose en una difícil adaptación. Sin embargo, volver a la escuela protegidos por las medidas de salud y la vacuna, les hizo sentir emoción y felicidad, por conocer a sus compañeros, asimismo los llevó a alcanzar un mejor nivel de calificaciones. (Ver figura 1 Esquema general de interpretación).

Figura 1. Esquema general de interpretación.

Fuente: Elaboración propia.

Cierre

Como se ha establecido, para el colectivo social en estudio, regresar a clases presenciales en tiempos de pospandemia por COVID-19 representó desde diversas aristas una complicada situación, debido a que el mencionado retorno, se vio subsumido por los efectos negativos del choque cultural en la salud y educación de los discentes; en este orden de ideas, resulta imperativo considerar el carácter biopsicosocial del ser humano el cual le ubica, dentro del marco de un complejo constructo, donde todo acto de su vida ha de regirse en un contexto de relaciones tanto objetivas como subjetivas; a decir de López et al. (2008, p.324) la dimensión material se manifiesta orgánica, biológica y socialmente, mientras la dimensión subjetiva contempla la interpretación de la forma en la cual viven y enfrentan los procesos vitales a partir de la intersubjetividad colectiva.

A pesar de los embates del CC, la resiliencia permitió a estos jóvenes estudiantes mantener la férrea convicción para continuar a paso firme con su educación a nivel medio superior; de esta forma las RS permitieron conocer aquellos factores propios del choque cultural, que afectaron el rendimiento académico y conductual de los alumnos de 6º semestre, lo cual pudiese otorgar la oportunidad de establecer protocolos adecuados para el manejo del CC desde la resiliencia; por lo tanto, adoptar una visión dilatada, permitirá a quienes toman las decisiones entender y atender los efectos colaterales de la COVID-19 desde una perspectiva amplia, integral e integradora; basada directamente en las necesidades específicas, manifiestas y vívidas de los discentes de nivel medio superior tras el regreso presencial a clases en tiempos de pospandemia.

Referencias

Arellano-Esparza, C. A., & Ortiz-Espinoza, Á. (2022). Educación media superior en México: abandono escolar y políticas públicas durante la covid-19. *Íconos. Revista de Ciencias Sociales*, (74), 33-52.

Bustos, M. V. (2021). Pospandemia: ¿educación nueva o nueva educación? -*SIGNOS EAD*, *(5)*.

Comisión Económica para América Latina y el Caribe (2020). *La educación en tiempos de la pandemia de COVID-19.*

Flores, A. Casas, O. Maya, M., y Rodríguez, A. (2019). Representaciones Sociales de la dieta en pacientes con diabetes mellitus tipo 2 con o sin control, en el Estado de México – Oriente. [Tesis de maestría UAEMÉX].

García, M. y Domínguez, O. (2013). Desarrollo teórico de la Resiliencia y su aplicación en situaciones adversas: Una revisión analítica. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud. vol. 11*(1). p. 63-77.

Gomáriz, M. (2020). Retorno escolar presencial pospandemia en Iberoamérica: avances, reflexiones y recomendaciones. Organización de Estados Iberoamericanos, 1-22.

Gutiérrez, R. (2016) La atención a la diversidad en la orientación educativa de la UAEM: hacia una cultura de paz. En Rojas, A., Villalobos, G., Brunett, K., y Martínez, J. (Eds.), *Por una cultura de la paz: Una mirada desde las ciencias de la conducta* (pp.255 – 264) UAEM.

López, O., Escudero J. y Carmona, L. (2008). Los determinantes sociales de la salud. Una perspectiva desde el Taller Latinoamericano de Determinantes Sociales de la Salud, ALAMES. *Medicina Social, 3*(4), 323-335.

Maguiña, C., Gastelo, R., & Tequen, A. (2020). El nuevo Coronavirus y la pandemia del Covid-19. *Revista Médica Herediana*, *31*(2), 125-131.

Martínez, C. (2008). La educación a distancia: sus características y necesidad en la educación actual. *Educación. 17*(33). 7-27.

Montalván, J. Viteri, M. Silva, F. y Pilco, M. (2022). Impacto del confinamiento pandémico en estudiantes universitarios retornado a clases presenciales: Una revisión bibliográfica de problemas latentes a partir de la psicología social y la pedagogía. *Encuentros.* (18). 204-225.

Mora, M. (2002) La teoría de las representaciones sociales de Serge Moscovici. *Athenea Digital, 2*. 1-25.

Organización Mundial de la Salud. (2020). *COVID-19: cronología de la actuación de la OMS.*

Ortunio, C. y Guevara, R. (2016). Aproximación teórica al constructo resiliencia. *Comunidad y Salud. 14*(33). 96-105.

Peñaloza, A. P. P., Cuevas, S. C. C., García, A. J. M., & Miranda, A. V. (2023). Resiliencia de estudiantes universitarios postpandemia de la COVID-19. *Polo del Conocimiento*, *8*(3), 2145-2154.

Real Academia Española (2022). *Diccionario de la lengua española.* Edición del Tricentenario.

Rojas, L. P., & Quezada, A. N. (2021). Educación a distancia y pandemia. Experiencias en educación básica en zonas rurales y urbanas. *Reencuentro. Análisis de problemas universitarios*, *31*(78), 101-122.

Secretaría de Educación Pública. (2017a). *Modelo educativo para la educación obligatoria*.

Secretaría de Educación Pública. (2017b). *Planes de estudio de referencia del marco curricular común de la Educación Media Superior.*

Silva, J. (2010). La educación virtual en México. Repositorio UPIICSA, XVIII, VII, 53-54

Troncoso Araya, J. L. (2022). ¿De vuelta a la normalidad? análisis psicológico de la vuelta a clases en tiempos de postpandemia covid-19. *Cuadernos de neuropsicología*, *16*(1), 94-99.