ENFOQUES BASADOS EN CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

Editores:

Mireya Tovar Vidal
María Teresa Torrijos Muñoz
Carlos Leopoldo Carreón Díaz de León
Daniel Marcelo González Arriaga
Juan Carlos Conde Ramírez
Jaime Alejandro Romero Sierra

Enfoques basados en Ciencia de Datos e Inteligencia Artificial

Enfoques basados en Ciencia de Datos e Inteligencia Artificial

Mireya Tovar Vidal
María Teresa Torrijos Muñoz
Carlos Leopoldo Carreón Díaz de León
Daniel Marcelo González Arriaga
Juan Carlos Conde Ramírez
Jaime Alejandro Romero Sierra
Editores

BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

Rectora: María Lilia Cedillo Ramírez • Secretario General: Damián Hernández Méndez • Vicerrector de Extensión y Difusión de la Cultura: Luis Antonio Lucio Venegas • Director General de Publicaciones: Mary Kriss Parra Górriz

Primera edición **2025**
ISBN: 978-968-9744-02-3

4 Sur 104, Centro Histórico, Puebla, Pue., CP 72000
Teléfono: 222 229 55 00
www.buap.mx

Dirección General de Publicaciones
2 Norte 1404, Centro Histórico, Puebla, Pue., CP 72000
Teléfono: 222 246 85 59
libros.dgp@correo.buap.mx | www.publicaciones.buap.mx

Diseño de portada: Adobe illustrator

Hecho en México
Made in Mexico

Prólogo

En el presente libro titulado *"Enfoques basados en Ciencia de Datos e Inteligencia Artificial"* se abordan diferentes áreas de investigación en ciencia de datos, aprendizaje automático, procesamiento de lenguaje natural, entre otras. La obra incluye diez capítulos de investigación divididos en distintas áreas de la Inteligencia Artificial.

Los capítulos que forman parte de esta obra fueron revisados mediante el sistema de doble par ciego y aprobados para su publicación por expertos en el área de conocimiento. De la misma manera, la obra en su totalidad fue revisada por dos expertos en el área de conocimiento. A continuación se menciona la aportación de cada uno de estos capítulos.

En el Capítulo 1 se presenta un enfoque de clasificación multiclase para la detección de enfermedades tiroideas (hipotiroidismo, hipertiroidismo y función tiroidea normal) utilizando modelos de bosque aleatorio y redes neuronales profundas (DNN). Para el estudio, se utilizó un conjunto de datos públicos del repositorio UCI Machine Learning que contenía variables clínicas estructuradas y un desequilibrio significativo entre las clases.

En el Capítulo 2 se ofrece una comparación sistemática de tres métodos explicables para la generación de datos sintéticos tabulares en el ámbito médico: árboles de decisión (CART), cópulas gaussianas y un enfoque combinado de ecuaciones estructurales con redes bayesianas (SEM+BN). También se presenta la capacidad de estas técnicas para preservar las relaciones estadísticas de conjuntos de datos clínicos reales, utilizando como métrica principal la distancia de Hamming para cuantificar la similitud entre registros originales y sintéticos.

En el Capítulo 3 se evalúan y comparan cuatro modelos Transformer preentrenados (BERT, RoBERTa, RoBERTuito y DistilBERT) para la clasificación de seis emociones básicas (Felicidad, Tristeza, Disgusto, Ira, Miedo y Sorpresa) en textos breves en español, utilizando un corpus propio de 341 respuestas de estudiantes universitarios. El estudio incluye un preprocesamiento detallado, ajuste fino de cada modelo y evaluación mediante métricas estándar (exactitud, precisión, recall, F1-score, AUC y matrices de confusión).

En el Capítulo 4 se propone un modelo híbrido para la clasificación de seis emociones básicas (felicidad, tristeza, disgusto, ira, miedo y sorpresa) en textos en español. Se construyó un corpus original de 4 200 textos etiquetados manualmente mediante encuestas. El enfoque combina embeddings de RoBERTa afinado con clasificadores tradicionales (regresión logística y SVM) a través de una etapa de meta-clasificación.

En el Capítulo 5 se presenta un estudio que combina el preprocesamiento CLAHE con un modelo de red neuronal convolucional llamado DenseNet-201 para mejorar la detección de COVID-19 en tomografías computarizadas. Se muestran mejores resultados para la combinación CLAHE + DenseNet201 sobre solo el uso de DenseNet-201.

En el Capítulo 6 se presenta la implementación de un sistema para identificar movimientos oculares mediante señales de electrooculografía (EOG) con la finalidad de controlar un prototipo para personas con discapacidad física o del habla. Los mejores resultados se obtienen al realizar un análisis temporal de las señales, empleando una red neuronal bilateral, con una exactitud del 97.8 % en la clasificación.

En el Capítulo 7 se aborda una propuesta para optimizar un algoritmo de marca de agua de imágenes digitales que combina la Transformada Wavelet Discreta (DWT) y la Descomposición en Valores Singulares (SVD). Los resultados muestran mejora en la calidad de la marca de agua extraída, reducción en el tiempo de ejecución respecto a otras metaheurísticas y buen desempeño general.

En el Capítulo 8 se propone un modelo de crecimiento económico basado en aprendizaje por refuerzo como solución numérica, procesos de Markov y Q-learning. A partir de los experimentos numéricos se muestra la trayectoria óptima del capital promedio y los resultados se interpretan en términos económicos.

En el Capítulo 9 se aborda un estudio descriptivo de las expectativas de 58 estudiantes que ingresan a la Ingeniería en Ciencia de Datos, como motivación, expectativas laborales, preocupaciones, impacto social, habilidades técnicas, aprendizaje automático, área de interés, experiencia previa, expectativas de la carrera.

Por último, en el Capítulo 10 se menciona el uso de las microcredenciales como herramienta innovadora que permite a los estudiantes adquirir habilidades específicas orientadas a resultados prácticos. Las microcredenciales que se indican pueden ser emitidas por universidades, plataformas de educación en línea o por actores de la industria tecnológica.

Finalmente, queremos agradecer a cada uno de los autores por su contribución, al comité revisor por su valiosa labor y cuidado en el proceso de revisión ciega, a la Facultad de Ciencias de la Computación de la Benemérita Universidad Autónoma de Puebla y a todos aquellos cuya participación contribuyó para la publicación de este libro.

Los editores,
Mireya Tovar Vidal
María Teresa Torrijos Muñoz
Carlos Leopoldo Carreón Díaz de León
Daniel Marcelo González Arriaga
Juan Carlos Conde Ramírez
Jaime Alejandro Romero Sierra

Índice general

Capítulo 1

Aplicación de técnicas de ciencia de datos con redes neuronales y random forest para la clasificación automatizada de trastornos tiroideos

Ana Laura Lezama Sánchez[1,2], Mireya Tovar Vidal[2]

[1] Secretaría de Ciencia, Humanidades, Tecnología e Innovación,
[2] Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla, México
{analaura.lezama,mireya.tovar}@correo.buap.mx

Resumen. En este trabajo se presenta un enfoque de clasificación multiclase para la detección de enfermedades tiroideas (hipotiroidismo, hipertiroidismo y función tiroidea normal) utilizando modelos de *random fores*t y redes neuronales profundas. Se empleó un conjunto de datos público del repositorio *UCI Machine Learning*, compuesto por variables clínicas estructuradas y una distribución altamente desbalanceada entre clases. A través de un análisis exploratorio, se identificó este desbalance, el cual influyó significativamente en el rendimiento de los modelos, especialmente en la red neuronal. Aunque ambos modelos alcanzaron un *accuracy* de 97.0% para *random forest* y 95.8% para la red neuronal, las matrices de confusión y las métricas por clase revelaron un sesgo hacia la clase mayoritaria ("normal"), con bajo *recall* y F_1 para las clases minoritarias. Los resultados evidencian las limitaciones de los modelos clásicos y de aprendizaje profundo en contextos clínicos desbalanceados, y subrayan la necesidad de estrategias complementarias como el balanceo de clases, la ingeniería de características y la optimización de arquitecturas de red para mejorar la detección de condiciones menos frecuentes.

Palabras Clave: redes neuronales, ciencia de datos, enfermedades autoinmunes

1 Introducción

Las enfermedades tiroideas, como el hipotiroidismo y el hipertiroidismo, afectan a millones de hombres y mujeres en el mundo. Su diagnóstico oportuno es clave para evitar complicaciones o mitigarlas como fatiga crónica, trastornos metabólicos, cardiovasculares, entre otros (Jameson y Weetman, 2021; ATA, 2022). Por lo que, el diagnóstico oportuno es difícil debido a la complejidad de los síntomas y la variedad de estos entre pacientes, así como su similitud con otros padecimientos.

En este escenario, la implementación de la inteligencia artificial en la creación de herramientas computacionales ha ganado una importancia cada vez mayor como respaldo en la toma de decisiones clínicas (Miotto et al., 2018).

Los avances recientes en ciencia de datos, aprendizaje automático y aprendizaje profundo han posibilitado la elaboración de modelos capaces de identificar patrones en datos clínicos estructurados, fortaleciendo de esta manera el apoyo al diagnóstico médico (Esteva et al., 2017).

Específicamente, las redes neuronales profundas (DNNs) han demostrado una habilidad excepcional para modelar relaciones complejas no lineales, incluso en situaciones de datos diversos o con valores ausentes (Goodfellow et al., 2016).

En este estudio se exponen dos modelos el primero *random forest* y el segundo una red neuronal profunda para la clasificación de enfermedades tiroideas. El conjunto de datos empleado proviene del repositorio *UCI Machine Learning* (Quinlan, 1986) y contiene miles de observaciones con variables continuas y booleanas, además de un alto porcentaje de datos faltantes, lo que refleja un escenario clínico realista. Por lo que, se implementó un pipeline integral que incluye preprocesamiento, codificación, entrenamiento y evaluación, con el objetivo de evaluar la capacidad de los modelos para clasificar tres condiciones clínicas: hipotiroidismo, hipertiroidismo y función tiroidea normal o negativa.

El presente trabajo se organiza de la siguiente manera: la Sección 2 presenta el marco teórico; la Sección 3 expone los trabajos relacionados; en la Sección 4 se describe la metodología propuesta; la Sección 5 muestra los resultados experimentales; y finalmente, la Sección 6 presenta las conclusiones y trabajos futuros, así como las referencias utilizadas en el desarrollo de este.

2 Marco teórico

Las enfermedades tiroideas, como el hipotiroidismo y el hipertiroidismo, se derivan de una alteración en la producción de hormonas tiroideas, lo que afecta funciones metabólicas clave del cuerpo importantes en la vida diaria del paciente. Estas condiciones son prevalentes a nivel mundial y pueden generar consecuencias graves si no se detectan a tiempo, como trastornos cardiovasculares, cognitivos, metabólicos, entro otros, así como problemas en la vida diaria del paciente (Jameson y Weetman, 2021).

El diagnóstico precoz es crucial, sin embargo, plantea retos clínicos significativos debido a la diversidad de los síntomas (como cansancio, incremento de peso, ansiedad, pérdida de cabello, sequedad en la piel) entre personas, su parecido con otras afecciones y la exigencia de analizar varias variables clínicas y bioquímicas (ATA, 2022). Esto ha promovido el avance de sistemas que asisten en la toma de decisiones médicas fundamentadas en datos.

Por lo que, en los últimos años el aprendizaje automático (ML) se ha posicionado como instrumento útil para clasificación y predicción de patrones en datos médicos. Modelos como *random forest*, árboles de decisión, máquinas de soporte vectorial (*SVM*), los *k*-vecinos más próximos (*k-NN*), entre otros han sido implementados en problemas clínicos generando resultados prometedores (Miotto et al., 2018).

Además, las redes neuronales profundas (DNNs) han permitido un manejo de datos complejos ofreciendo resultados con mayor precisión. Estas redes poseen la capacidad de modelar vínculos no lineales y obtener representaciones internas útiles para mejorar la precisión de los modelos generados, con mayor precisión que un clasificador tradicional (Goodfellow et al., 2016).

Algunos autores han abordado la aplicación de ML como auxiliar en la detección de enfermedades como en Quinlan (1986), donde desarrollaron uno de los primeros conjuntos de datos estructurados para este tipo de enfermedad, que incluye variables como niveles hormonales, antecedentes médicos y condiciones clínicas. Posteriormente, estos datos han sido empleados para probar en diferentes algoritmos, observando mejoras especialmente con modelos basados en redes neuronales (Kaya et al., 2019). No obstante, uno de los principales retos en estos conjuntos de datos es el desbalance de clases, ya que la mayoría de los registros corresponden a pacientes con función tiroidea normal. Este desequilibrio afecta el rendimiento de los modelos, especialmente en la detección de clases minoritarias como hipotiroidismo o hipertiroidismo, ya que el modelo se inclina hacia la clase con mayor número de elementos (Chawla et al., 2002).

3 Estado del arte

En esta sección se exponen algunos de los trabajos relacionados en el campo de estudio de este trabajo.

Diversos enfoques de aprendizaje automático han sido aplicados a esta tarea. Por ejemplo, (Reyes León et al., 2020) realizó un estudio documental y experimental evaluando el uso de modelos de cómputo inteligente en el pre-diagnóstico de enfermedades crónicas, aplicando nueve algoritmos de clasificación de patrones sobre bases de datos de cáncer de mama, hipertiroidismo e hipotiroidismo. Por otro lado (Kumari et al., 2024) realizaron un análisis exploratorio de algoritmos de aprendizaje automático para la clasificación de enfermedades tiroideas. Los autores emplearon un conjunto de datos que incluye variables como edad, género y niveles hormonales. Por lo que, implementaron algoritmos como *Support Vector Machines (SVM), random forest (RF), XGBoost* y un clasificador por *ensamblado (ensemble)*.

Además, la ausencia o variedad de atributos que incluyen variables booleanas, continuas y categóricas constituyen desafíos considerables para el desarrollo de modelos sólidos. Numerosos estudios han tratado este problema utilizando diferentes técnicas, las cuales se han catalogado como estrategias fundamentales para optimizar el rendimiento de los clasificadores (Chawla et al., 2002). Por otro lado, Chaganti et al. (2022) subrayan que la predicción de enfermedades tiroideas ha cobrado una relevancia creciente en el campo de la investigación; sin embargo, la mayoría de los trabajos previos se han centrado en tareas de clasificación binaria, basándose en conjuntos de datos pequeños y sin validaciones rigurosas. En su estudio, los autores se basan en la ingeniería de características para modelos de aprendizaje automático y profundo, utilizando métodos como la selección hacia adelante (*forward feature selection*), la eliminación hacia atrás (*backward feature elimination*), la eliminación bidireccional y

la selección basada en clasificadores *Extra Trees*. Sus resultados evidencian que emplear *Extra Trees* con un clasificador *Random Forest* alcanzó una precisión y una puntuación F_1 de 0.98, lo que demuestra un desempeño sobresaliente en la detección de diversas patologías tiroideas, incluidas la tiroiditis de Hashimoto y la tiroiditis autoinmune. Además, emplearon validación cruzada *k-fold* para fortalecer la evaluación del modelo y compararon sus resultados con otros algoritmos de *machine learning,* concluyendo que los métodos tradicionales ofrecen un balance favorable entre precisión y complejidad computacional.

En este contexto, el presente estudio expone como una contribución orientada a evaluar la eficacia de un modelo de red neuronal profunda y *random forest* en un entorno clínico simulado y la evaluación de métricas como precisión, *accuracy, F1, recall* y el rendimiento por clase.

4 Metodología

El presente estudio se desarrolló siguiendo un enfoque metodológico basado en Ciencia de Datos, el cual comprendió las siguientes etapas:

Adquisición de datos: Se utilizó un conjunto de datos público sobre diagnóstico de trastornos tiroideos, obtenido en formato CSV (hypothyroid.csv).

Análisis exploratorio y limpieza de datos: Se realizó un análisis exploratorio inicial (EDA) para identificar valores ausentes, inconsistencias y posibles errores de entrada. Además, de verificar la correspondencia entre variables y la presencia de valores faltantes. Con base en estos hallazgos: Se imputaron valores faltantes con la mediana (para variables numéricas) y la moda (para variables categóricas). Se eliminaron columnas con alta proporción de valores nulos o baja variabilidad y se descartaron entradas con datos anómalos.

Ingeniería de características: Dado que la variable objetivo original (target) contenía múltiples etiquetas clínicas (e.g. A, B, C...), se reclasificaron en tres categorías clínicas más generales: Hyperthyroid, Hypothyroid y Negative (sin trastornos). Esto permitió simplificar la tarea de clasificación multiclase. Asimismo, se transformaron variables categóricas mediante codificación binaria (LabelEncoder) y codificación one-hot (pd.get_dummies), según el modelo a emplear.

Visualización de datos: Se generaron gráficos de barras para analizar la distribución de clases en la variable objetivo (target), permitiendo observar un ligero desbalance entre clases.

Modelado predictivo: Se entrenaron y compararon dos enfoques de aprendizaje supervisado:

- *random forest*: Se aplicó escalamiento estándar a los datos.
  - Se dividió el conjunto en entrenamiento (80%) y prueba (20%).
  - Se evaluó el modelo utilizando métricas de clasificación multiclase: *accuracy, precision, recall y F1*.
- red neuronal profunda
  - Se construyó una red neuronal secuencial con dos capas ocultas y función de activación ReLU, utilizando Keras:
    - Primera capa con 64 neuronas y función de activación ReLU.
    - Segunda capa con 32 neuronas y activación ReLU.
    - Capa de regularización Dropout (0.3) después de la primera capa oculta.
  - La capa de salida tiene 3 neuronas (una por clase) con función de activación softmax.
  - Se empleó categorical_crossentropy como función de pérdida con optimizador Adam.
  - Las salidas se codificaron en formato one-hot.
  - Se evaluó el rendimiento sobre el conjunto de prueba usando las mismas métricas mencionadas anteriormente.

- Exportación de resultados: Las predicciones generadas por ambos modelos fueron exportadas a archivos .csv, incluyendo etiquetas verdaderas y predichas, con el fin de facilitar análisis posteriores.

5 Resultados y discusión

En esta sección se exponen los resultados obtenidos al aplicar la metodología propuesta sobre el conjunto de datos seleccionado.

5.1 Conjunto de datos

El conjunto de datos utilizado proviene del Garavan Institute en Sídney, Australia, y fue documentado por (Quinlan, 1986). Este repositorio incluye múltiples bases de datos relacionadas con enfermedades tiroideas, cada una con características específicas:

- Aproximadamente 2,800 instancias de entrenamiento y 972 de prueba por base de datos.
- Alrededor de 29 atributos, que pueden ser variables booleanas o valores continuos.
- Gran cantidad de datos faltantes en algunas bases.
- Existen bases adicionales, como *Hypothyroid.data* y *sick-euthyroid.data*, también atribuidas a Ross Quinlan, que presentan un formato similar, pero se consideran potencialmente corruptas.
- Otra base de datos con 9,172 instancias y 20 clases, junto con teoría de dominio relacionada.
- Una base específica para entrenamiento de redes neuronales artificiales con 3 clases, 3,772 instancias de entrenamiento y 3,428 de prueba, que incluye información adicional sobre costos, donada por Peter Turney.

En general, estas bases de datos presentan un conjunto diverso de ejemplos clínicos para el diagnóstico de condiciones tiroideas, con distintas clases y variables que permiten desarrollar modelos predictivos robustos.

5.2 Resultados

El modelo de *random forest* mostró un rendimiento robusto en la clasificación de las tres clases: Hipotiroidismo, Hipertiroidismo y Normal. En la Tabla 1, podemos observar que el modelo alcanzó un *accuracy* del 97.00%, destacando particularmente en la identificación de la clase Normal y con buen desempeño en la detección de Hipotiroidismo.

En contraste, el modelo de red neuronal profunda alcanzó un *accuracy* del 95.80%, lo que indica un buen desempeño general.

Tabla 1. Resultados obtenidos por cada modelo

Modelo	Precision	Recall	Accuracy	F_1
random forest	0.9653	0.9700	0.9700	0.9667
red neuronal profunda	0.9525	0.9580	0.9580	0.9496

Las métricas específicas por clase permiten un análisis más detallado del comportamiento de cada modelo, como se muestra en las Tablas 2 y 3. En el caso del modelo *random forest*, la clase "Normal" presentó los mejores resultados, con una precisión y *recall* superiores al 98%. La clase Hipotirodismo también fue clasificada con alta efectividad (F_1 de 0.96), mientras que la clase Hipertiroidismo mostró un desempeño más bajo, con un F_1 de 0.43. Este patrón también se observa en la red neuronal, aunque con un F_1 aún menor para Hipertiroidismo (0.31), lo que refuerza la necesidad de estrategias para abordar el desbalance de clases.

Tabla 2. Resultados obtenidos por cada clase con *random forest*

Clase	Precisión	Recall	F_1
Hipertiroidismo	0.58	0.34	0.43
Hipotiroidismo	0.94	0.98	0.96
Normal	0.98	0.99	0.98

Tabla 3. Resultados obtenidos por cada clase con la red neuronal

Clase	Precisión	Recall	F_1
Hipertiroidismo	0.65	0.21	0.31
Hipotiroidismo	0.80	0.87	0.84
Normal	0.97	0.98	0.98

Además de las métricas anteriores, se generaron gráficos de barras (ver Figura 1) para analizar la distribución de clases en la variable objetivo (target), lo cual permitió identificar un desbalance entre clases, especialmente en los casos de hipertiroidismo. Este desbalance fue considerado al momento de evaluar las métricas con promedios ponderados (average='weighted').

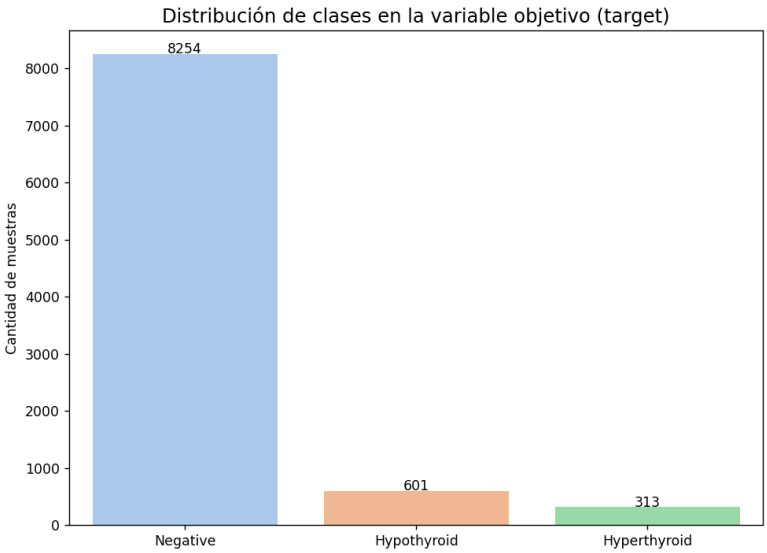


Figura 1. Distribución de clases

Para evaluar el rendimiento específico por clase, se analizaron las matrices de confusión generadas por ambos modelos (ver Figuras 2 y 3). Estas matrices permiten visualizar cómo fueron clasificadas las instancias reales de cada clase, y evidencian las fortalezas y debilidades de cada enfoque.

En el caso del modelo *random forest*, se observó un desempeño robusto en la clasificación de las tres clases, particularmente para la clase normal (Negativo), con una alta proporción de verdaderos positivos (1,660 de 1,678 instancias reales). La clase hipotiroidismo también fue bien clasificada, con 101 verdaderos positivos de un total de 103. Sin embargo, se evidenció una limitación en la detección de casos de hipertiroidismo, donde solo 18 de 53 casos fueron correctamente identificados, mientras que 33 fueron clasificados erróneamente como normal.

Por otro lado, la red neuronal profunda mostró un rendimiento general aceptable, pero con un mayor grado de confusión entre las clases. La clase normal nuevamente fue la mejor clasificada, con 1,652 aciertos, aunque se presentaron más falsos positivos hacia hiper e hipo en comparación con *random forest*. La clase hipotiroidismo fue clasificada correctamente en 90 de 103 casos, pero se detectó un descenso en el desempeño al clasificar hipertiroidismo, con solo 11 aciertos y 40 errores hacia la clase normal. Esto indica que el modelo de red neuronal tiende a sobreajustarse a la clase mayoritaria, afectando especialmente a las clases menos representadas como hipertiroidismo.

Estas observaciones refuerzan la importancia de considerar el desbalance de clases en el conjunto de datos, así como la necesidad de aplicar técnicas específicas de ajuste o rebalanceo para mejorar la sensibilidad en clases minoritarias.

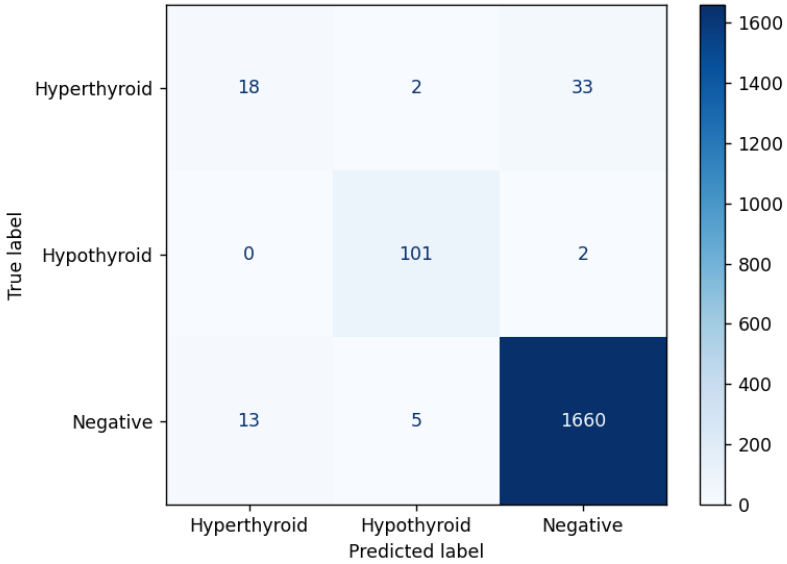


Figura 2. Matriz de confusión de random forest

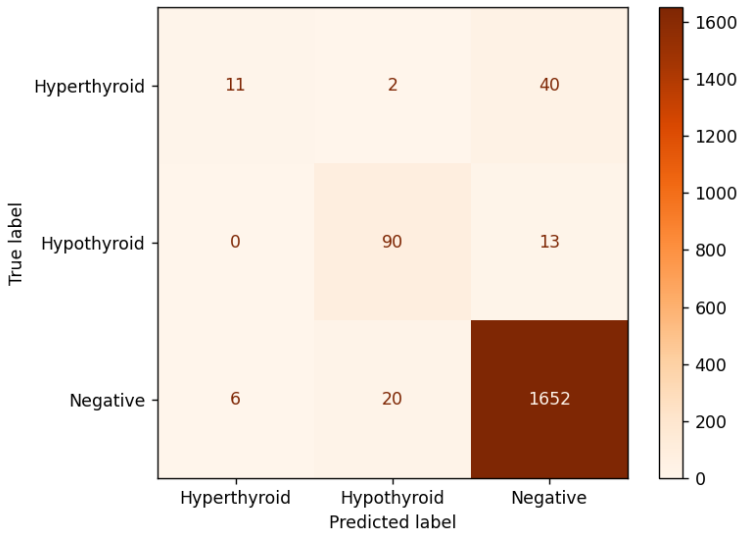


Figura 3 Matriz de confusión de la red neuronal

A manera de ejemplo, en la Tabla 4 se presenta un fragmento del archivo .csv generado con las predicciones de ambos modelos. En este se incluyen tanto la clase original, así como las predicciones obtenidas por cada modelo.

Tabla 3. Ejemplo de resultados del modelo: clase original y predicción por cada modelo

Clase original	random forest	red neuronal
Hypothyroid	Hypothyroid	Hypothyroid
Negative	Negative	Negative
Hyperthyroid	Negative	Negative
Negative	Negative	Negative
Hypothyroid	Hypothyroid	Hypothyroid
Negative	Negative	Negative
Negative	Negative	Negative

En contraste con nuestro estudio, donde se implementó una red neuronal profunda y *random forest* para la clasificación multiclase de enfermedades tiroideas, otros autores han abordado el problema con enfoques diferentes. Por ejemplo, un estudio reciente se (Chaganti et al., 2022) centró en la ingeniería de características para mejorar la predicción de enfermedades tiroideas, utilizando técnicas de selección de atributos como *forward feature selection, backward elimination* y clasificadores basados en árboles extra. Lo que llevó a los autores a reportar resultados superiores utilizando un clasificador *random forest*, alcanzando una precisión y un F1 de 0.98. Estos valores son significativamente más altos que los obtenidos en el presente estudio, donde el modelo de *random forest* alcanzó un *accuracy* de 0.9700 y la red neuronal profunda obtuvo un *accuracy* de 0.9580.

Aunque la red neuronal profunda tiene la habilidad de modelar relaciones complejas, su desempeño fue inferior en comparación con el modelo de *random forest*, en particular en la categorización de la clase minoritaria "Hipertiroidismo". Esto puede deberse a diversas causas. En primer lugar, las redes neuronales suelen ser más susceptibles al desequilibrio de clases, debido a que la función de pérdida (*categorical_crossentropy*) puede estar bajo el control de las clases dominantes, lo que complica el aprendizaje de patrones representativos para clases poco comunes. Además, la arquitectura utilizada con únicamente dos capas ocultas y sin métodos sofisticados de regularización más allá del *Dropout* podría haber restringido la habilidad del modelo para generalizar con corrección, especialmente en caso de ruido o desbalance.

Cabe resaltar que, en este trabajo, no se exploraron arquitecturas alternativas más profundas ni se aplicaron técnicas de mejora como el ajuste de pesos por clase, la generación sintética de muestras (SMOTE), ni la incorporación de penalizaciones en la función de pérdida. Además, aunque se identificó un desbalance significativo en la distribución de clases, no se aplicaron técnicas de rebalanceo durante este estudio con el objetivo de mantener la distribución original del conjunto de datos clínicos. Sin

embargo, se reconoce que el desbalance afectó especialmente el desempeño del modelo en la clasificación de la clase “Hipertiroidismo”.

6 Conclusiones y trabajos a futuro

En este estudio, se propuso utilizar una red neuronal profunda para clasificar las diferentes patologías tiroideas. El conjunto de datos utilizado en los experimentos proviene del repositorio de aprendizaje automático de la UCI. La metodología utilizada se basó en un enfoque de ciencia de datos, que incluyó el análisis exploratorio del conjunto de datos, identificar desbalances en las clases, selección de métricas de evaluación y la comparación de varios modelos. El modelo registró una precisión del 95.8%, mostrando un rendimiento global en la categorización de la categoría "normal". No obstante, detectó casos reales de hipotiroidismo (*recall* de 0.87 y F_1 de 0.84), pero evidenció restricciones significativas en la identificación de hipertiroidismo (*recall* de 0.21 y F_1 de 0.31).

Los resultados obtenidos nos permitió detectar un sesgo hacia la clase dominante, un fenómeno habitual en situaciones con un marcado desequilibrio de clases, tal como sucede en este conjunto de datos, donde la clase "normal" constituye más del 90% de los casos.

Como trabajo futuro, se propone implementar las siguientes acciones:

- Balanceo del conjunto de datos: aplicar técnicas como SMOTE o submuestreo.
- Optimización de la arquitectura: explorar diferentes arquitecturas.
- Regularización y ajuste fino del modelo para mejorar su generalización.
- Selección de características relevantes: mediante técnicas como PCA.

Referencias

Quinlan, J. R. (1986). *Thyroid Disease Data Set*. UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/datasets/Thyroid+Disease

Reyes León, Patricio, Salgado Ramírez, Julio César, y Velázquez Rodríguez, José Luis. (2020). *Pre-diagnóstico de enfermedades crónicas mediante la aplicación de modelos de cómputo inteligente. Computación y Sistemas*, 24(3), 1313-1325. Epub 09 de junio de 2021.https://doi.org/10.13053/cys-24-3-3492

American Thyroid Association. (2022). *Thyroid disease information*. https://www.thyroid.org

Chawla, N. V., Bowyer, K. W., Hall, L. O., y Kegelmeyer, W. P. (2002). *SMOTE: Synthetic Minority Over-sampling Technique*. Journal of Artificial Intelligence Research, 16, 321–357.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., y Thrun, S. (2017). *Dermatologist-level classification of skin cancer with deep neural networks*. Nature, 542 (7639), 115–118. https://doi.org/10.1038/nature21056

Goodfellow, I., Bengio, Y., y Courville, A. (2016). *Deep learning*. *MIT Press*.

Jameson, J. L., y Weetman, A. P. (2021). *Disorders of the thyroid gland*. En J. L. Jameson et al. (Eds.), *Harrison’s Principles of Internal Medicine* (20a ed.). McGraw-Hill.

Miotto, R., Wang, F., Wang, S., Jiang, X., y Dudley, J. T. (2018). *Deep learning for healthcare: review, opportunities and challenges*. *Briefings in Bioinformatics*, 19(6), 1236–1246. https://doi.org/10.1093/bib/bbx044

Kaya, Y., Pehlivan, M., y Yildirim, T. (2019). *Classification of thyroid diseases using machine learning algorithms. Procedia Computer Science*, 154, 376–382.

Kumari, P., Kaur, B., Rakhra, M., et al. (2024). *Explainable artificial intelligence and machine learning algorithms for classification of thyroid disease*. *Discover Applied Sciences*, 6(360), 1-17. https://doi.org/10.1007/s42452-024-06068-w

Chaganti, R., Rustam, F., De La Torre Díez, I., Mazón, J. L. V., Rodríguez, C. L., y Ashraf, I. (2022). Thyroid disease prediction using selective features and machine learning techniques. Cancers, 14(16), 3914.

Capítulo 2

Una comparativa de técnicas explicables de generación de datos sintéticos

Joab Atietl Aguilar Mendoza, Daniel Carreño Aguilar, Gustavo Emilio Mendoza Olguín, Luis Yael Méndez Sánchez

Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla, México

joab.aguilar@alumno.buap.mx, daniel.carrenoag@alumno.buap.mx, gustavo.mendozaol@correo.buap.mx, luis.mendezsanchez@correo.buap.mx

Resumen.

En el ámbito médico el uso de datos sintéticos es una necesidad debido a la dificultad de obtener datos reales por diversas causas como la protección de datos o la falta de representatividad de algunos casos. El objetivo de este estudio es evaluar la calidad de los datos sintéticos tabulares generados por 3 algoritmos seleccionados: árboles de decisión (CART), cópulas gaussianas y una combinación del modelo de ecuaciones estructurales con redes bayesianas (SEM+BN), a partir de su capacidad de generar valores que mantengan las relaciones estadísticas existentes en datos reales usando la distancia de Hamming como métrica de similaridad. Con base en los resultados se discuten las fortalezas y debilidades de cada algoritmo y se ofrecen recomendaciones para su utilización en tareas como simulaciones clínicas o entrenamiento de modelos de aprendizaje automático.

Palabras Clave: Datos sintéticos, generación de datos, CART, SEM, Cópulas Gaussianas, Redes Bayesianas.

1 Introducción

Los datos sintéticos se definen como registros generados artificialmente que reproducen propiedades estadísticas de conjuntos reales sin mostrar información sensible ni privada. (Reiter, 2003). Este tipo de datos cada vez son más importantes en el ámbito médico, donde el acceso a datos clínicos está restringido por razones éticas, legales o prácticas (López et al., 2020). Cuando los datos son generados de forma correcta, pueden preservar relaciones estructurales y distribuciones importantes de las variables,

lo que nos permite su uso en simulaciones, validación de modelos o entrenamiento de algoritmos de aprendizaje automático (Barbierato et al., 2022).

Entre los retos que la generación de datos sintéticos presenta, se encuentran la incapacidad de algunos métodos de generación para tratar variables y sus dependencias, lo que pudiera generar errores al momento de entrenarlos (Zhao, 2025). Además, existe el riesgo de que los registros sintéticos sean muy similares a los originales, lo que comprometería la privacidad de la información (Boneh et al., 2024). Para este tipo de problemas, se ha propuesto la distancia de Hamming como una métrica para comparar registros binarios y medir la similitud entre conjuntos reales y sintéticos. Esta métrica ha mostrado ser especialmente efectiva en tareas de clasificación, agrupamiento y detección de patrones en contextos sensibles como el médico (Preneel et al., 2024).

Este trabajo analiza cuán buenos son los datos sintéticos de tres métodos: árboles de decisión (CART), cópulas gaussianas y una combinación de ecuaciones estructurales con redes bayesianas (SEM+BN) se describe qué tan buenos son los datos sintéticos, cuando se hace explicables y se verifica si son similares a los datos reales mediante a la distancia de Hamming.

En el mundo médico, elegir métodos explicables es clave porque nos ayuda a comprender como se generan los datos, asegurándose de que todo esté claro y seguro al hacer tomas de decisiones clínicas. Por lo que es necesario asegurarnos de que estamos utilizando datos sintéticos para enseñar a nuestros modelos de predicción sin utilizar información delicada.

2 Marco teórico

La creación de datos sintéticos se ha convertido en una alternativa para resolver problemas en temas relacionados con la privacidad y calidad de los datos auténticos, enfocados en el sector médico. La literatura contemporánea sugiere diversas técnicas explicables, que incluyen modelos fundamentados en árboles de decisión, métodos probabilísticos como las cópulas gaussianas, y métodos mixtos que fusionan la causalidad estructural con el aprendizaje bayesiano.

Estas técnicas son consideradas explicables ya que facilitan una comprensión clara de cómo se crean los datos sintéticos. Papadaki et al. (2024) proporciona un análisis general de los métodos de generación tabular, de no solo preservar el valor analítico y asegurar la privacidad de los registros artificiales.

2.1 Árboles de decisión (CART)

El algoritmo CART (Classification and Regression Trees) forman árboles jerárquicos que representan las relaciones entre variables a través de reglas condicionales binarias, según Drechsler (2011). Reiter (2003) evidenció su capacidad para generar datos sintéticos con alta capacidad de preservación estadística, preservando así correlaciones cruzadas y distribuciones marginales. Pinheiro y Ribeiro (2025) modificaron la aplicación de CART para la producción de datos en contextos desbalanceados, demostrando beneficios al identificar patrones de la clase minoritaria. Las principales ventajas de CART incluyen su facilidad de implementación y capacidad para manejar datos mixtos (numéricos y categóricos). Sin embargo, tiene una inestabilidad ante los

pequeños cambios en los datos y puede generar registros de forma redundante si no se controla la poda del árbol.

Sea Cm el valor constante asignado a la región Rm (por ejemplo, la media en un problema de regresión o la moda en clasificación), $I(x \in Rm)$ la función indicadora que toma el valor 1 si la observación x pertenece a la región Rm y 0 en caso contrario, y m el índice que recorre las M regiones generadas por el árbol (es decir, los nodos terminales u "hojas"). Entonces, la predicción $y^{\wedge}$ para una nueva instancia x se define como:

$$\hat{y} = \sum_{\{m=1\}} Cm \cdot I(x \in R_m)$$

2.2 Cópulas Gaussianas

Las cópulas gaussianas son modelos probabilísticos que permiten separar las dependencias entre variables de sus distribuciones marginales (Shaposhnyk et al., 2025). En términos matemáticos, se basan en la transformación de variables mediante funciones inversas de distribución acumulativa.

Sea ΦR la función de distribución conjunta normal estándar con matriz de correlación R, y sea $\Phi{-}1$ la inversa de la función de distribución acumulada (CDF) univariada normal estándar. Sean $u1$, $u2$,...,$un \in [0,1]$ valores transformados de las variables originales mediante sus distribuciones marginales empíricas. Entonces, la cópula gaussiana multivariada se define como:

$$C(u_1, u_2, \ldots, u_n) = \Phi_{R(\Phi^{\{-1\}}(u_1),\ldots,\Phi^{\{-1\}}(u_n))}$$

López et al. (2020) mostraron su aplicación en geociencias para generar series sintéticas multivariadas. Las ventajas principales son su robustez para modelar dependencia no lineal y su adecuación para datos continuos. No obstante, requieren preprocesamiento cuidadoso y pueden tener limitaciones al modelar estructuras de dependencia complejas en datos mixtos.

2.3 SEM + Redes Bayesianas (SEM+BN)

Este método combina la lógica causal de las ecuaciones estructurales (SEM) con el aprendizaje de grafos probabilísticos dirigido (redes bayesianas). SEM permite definir describir cómo se relacionan entre sí variables que no siempre pueden observarse directamente, mientras que las redes bayesianas ayudan a representar esas relaciones en forma de un grafo dirigido, permitiendo hacer inferencias sobre los datos. La combinación de ambas técnicas busca generar datos sintéticos que conserven no solo patrones estadísticos, sino también la lógica causal subyacente. Barbierato et al. (2022) desarrollaron un generador que preserva tanto la causalidad como la diversidad estadística en datos sintéticos. Combrink et al. (2021) aplicaron SEM+BN en contextos educativos para generar datos sintéticos que conservaran estructuras de dependencia entre competencias y desempeño. Este método es ideal para mantener coherencia estructural en contextos sensibles, pero su desventaja es que requiere demasiado tiempo

de aprendizaje, poder de cómputo y conocimientos técnicos avanzados para definir correctamente las relaciones entre variables.

Sea η el vector de variables endógenas, ξ el vector de variables exógenas, B la matriz de coeficientes de regresión entre variables endógenas, Γ la matriz de coeficientes que conecta las variables exógenas con las endógenas, y ζ el vector de errores estructurales no correlacionados. Entonces, un modelo de ecuaciones estructurales (SEM) se expresa mediante:

$$\eta = B\eta + \Gamma\xi + \zeta$$

Por otro lado, una red bayesiana representa un conjunto de variables aleatorias $X1$,$X2$,...,Xn y sus relaciones condicionales mediante un grafo dirigido acíclico (DAG). Su distribución conjunta se descompone como:

$$P(X_1, \dots, X_n) = \prod_{\{i=1\}}^{n} P(Xi \mid Pa(Xi))$$

donde Pa(Xi) denota el conjunto de padres de en el grafo.

2.4 Métrica de evaluación: Distancia de Hamming

Para evaluar la similitud entre los datos reales y los sintéticos generados por cada técnica, se utiliza la distancia de Hamming, una métrica que cuantifica el número de diferencias entre dos vectores de igual longitud. Es particularmente útil cuando los datos se han transformado en representaciones binarias. Zhao (2025) y Boneh et al. (2024) demostraron que su cálculo puede realizarse de forma eficiente e incluso preservar la privacidad de los datos mediante técnicas criptográficas.

Sea x=($x1$,$x2$,...,xn) y y=($y1$,$y2$,...,yn) dos vectores de longitud n, con componentes binarias o categóricas. La distancia de Hamming entre x e y, denotada como $H(x,y)$, se define como: $H(x,y)$=$\sum i$=1$nI(xi\neq yi)$

$$H(x, y) = \sum_{\{i=1\}}^{n} I(x_i \neq y_i)$$

donde $I(xi\neq yi)$ es una función indicadora que toma el valor 1 si $xi\neq yi$(es decir, si hay una discrepancia en la posición i), y 0 en caso contrario.

3 Estado del arte

El ámbito médico exige soluciones que permitan desarrollar modelos predictivos sin comprometer la información sensible de los pacientes. En respuesta, la literatura ha identificado y evaluado diversos métodos para la generación de datos sintéticos tabulares, entre las que destacan los árboles de decisión (CART), las cópulas gaussianas y los modelos estructurales con redes bayesianas (SEM+BN) (Papadaki et al., 2024).

El método CART, continúa vigente como técnica base para la generación de microdatos sintéticos, especialmente en contextos donde la claridad del modelo es prioritaria. El modelo mantiene competitividad en dominios con estructuras de datos heterogéneas debido a su facilidad de interpretación e implementación, incluso contra

los modelos basados en aprendizaje profundo (Pinheiro y Riberido, 2025). Drechsler (2011), la utilizó en contextos institucionales a través de su integración con técnicas bayesianas, incrementando de esta manera su habilidad para producir diversas variantes sintéticas legítimas. Sin embargo, investigaciones actuales alertan acerca de sus restricciones al modelar relaciones complejas entre variables, particularmente en dominios de alta dimensionalidad (Pinheiro y Ribeiro, 2025). En el campo de la medicina, Drechsler (2011) utilizó el modelo CART para la creación de microdatos artificiales en cuestionarios de salud, manteniendo estructuras de jerarquía sin poner en riesgo la privacidad de los pacientes.

Las cópulas gaussianas, por otro lado, facilitan el modelado de dependencias lineales en grupos de datos continuos. Papadaki, (2024), destaca su solidez en campos donde la correlación entre variables es crucial para la validez del modelo, como en los ámbitos ambientales y económicos. No obstante, tal como señalan Shaposhnyk et al. (2025), su rendimiento disminuye frente a estructuras no lineales o variables categóricas mal codificadas, lo que restringe su utilidad en dominios de gran complejidad semántica. Sin embargo, su eficacia y escaso requisito computacional la establecen como una alternativa significativa en investigaciones comparativas y en situaciones donde la interpretación es más importante que la sofisticación técnica.

Papadaki et al. (2024) utilizaron cópulas gaussianas para analizar la producción de datos artificiales en grupos clínicos con variables continuas, evidenciando su eficacia en la conservación de correlaciones fisiológicas, como la tensión arterial o los niveles de glucosa.

Por otro lado, los modelos SEM+BN constituyen una de las iniciativas más audaces dentro de los métodos explicables. Esta combinación de técnicas posibilita la reproducción de patrones estadísticos y la modelación de vínculos causales entre variables. Barbierato et al. (2022) utilizaron este método para mantener la equidad en datos sintéticos delicados, mientras que Combrink et al. (2021) lo utilizaron en el ámbito educativo para simular vínculos entre habilidades académicas y desempeño. Más recientemente, la investigación de Shaposhnyk et al. (2025) sitúa a SEM+BN como uno de los procedimientos con el mejor equilibrio entre la fidelidad estructural y la capacidad de explicación, sobrepasando incluso a técnicas detalladas en contextos de evaluación regulatoria. No obstante, su aplicación demanda conocimiento especializado y procedimientos de discretización apropiados, lo que limita su adopción generalizada en ambientes operativos.

En el campo médico, Barbierato et al. (2022) emplearon SEM+BN para simular variables delicadas en registros médicos con el objetivo de equidad algorítmica, lo que posibilitó mantener estructuras causales sin duplicar datos individuales de manera directa.

4 Técnicas y herramientas

Se compararon tres métodos generativos explicables aplicados a datos tabulares médicos: árboles de decisión (CART), cópulas gaussianas y un modelo estructural combinado con redes bayesianas (SEM+BN). Todos los algoritmos fueron implementados en Python.

La comparación se realizó sobre tres conjuntos de datos públicos del ámbito clínico, correspondientes a enfermedades, accidente cerebrovascular (stroke) y cáncer de pulmón, todos obtenidos de la plataforma Kaggle. Estos datasets presentan características mixtas (ordinales y nominales) y fueron adaptados a un esquema de clasificación binaria. Para homogeneizar el análisis, se aplicaron estrategias de balanceo y una codificación binaria sobre todas las variables categóricas.

Como métrica principal de evaluación se utilizó la distancia de Hamming promedio, que permite medir la similitud entre registros binarios reales y sintéticos. Esta métrica fue seleccionada para la evaluación estructural de los datos generados.

5 Métodos

Para garantizar la calidad de la comparación, todos los métodos fueron aplicados a un mismo conjunto de datos preprocesado de forma idéntica. Las variables categóricas fueron codificadas binariamente y las numéricas normalizadas en el intervalo [0,1]. Se generaron registros sintéticos para cada método, replicando las estructuras de los conjuntos originales que son Brain Stroke, Heart Disease y Lung Cancer.

En el caso de CART, se emplearon árboles de clasificación y regresión con criterio Gini y profundidad máxima de 6. Se utilizó muestreo bayesiano dentro de las hojas para generar los valores sintéticos, incorporando DecisionTreeClassifier y DecisionTreeRegressor desde scikit-learn, siguiendo la metodología descrita por Reiter (2005) y Drechsler & Reiter (2011).

Para el caso de Cópula Gaussiana, se implementó el método de simulación condicional basado en cópulas bivariadas gaussianas empíricamente ajustadas, mediante la biblioteca pyvinecopulib (Nagler & Vatter, 2020). Se utilizaron transformaciones inversas de márgenes empíricos para mapear los valores generados al espacio original.

En SEM+BN, se ajustó un modelo de ecuaciones estructurales (semopy) seguido del aprendizaje de una red bayesiana con estructura dirigida (pgmpy).

6 Resultados

La evaluación de los datos sintéticos generados por cada método se basó en la distancia de Hamming promedio entre los registros sintéticos y los reales. Esta métrica permite cuantificar la similitud binaria entre los registros, siendo un indicador útil para evaluar la preservación estructural de los datos. A menor distancia, mayor fidelidad estadística entre ambos conjuntos. La Tabla 1 resume los valores obtenidos para cada técnica en los tres datasets utilizados (Heart Disease, Brain Stroke, y Lung Cancer).

Tabla 1 Comparación de la distancia de Hamming promedio por técnica generativa en tres datasets clínicos

Dataset	CART	Cópula	SEM+BN
Brain Stroke	0.200	0.277	0.314
Heart Disease	0.244	0.380	0.347
Lung Cancer	0.304	0.315	0.277

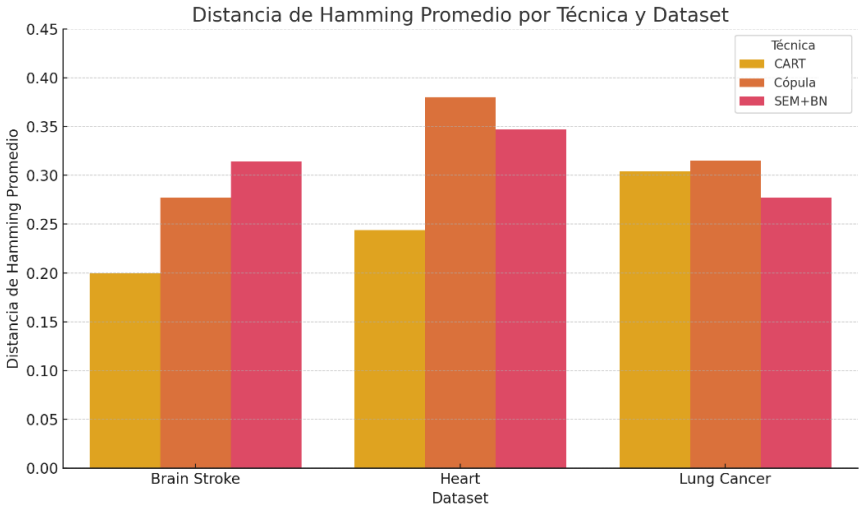


Figura 1 Distancia de Hamming promedio por técnica y dataset

En la Figura 1. se observa la distancia de Hamming promedio alcanzada por cada técnica generativa en los tres conjuntos de datos. Se aprecia cómo el rendimiento varía según el dataset, siendo CART la técnica más consistente y con menor distancia en general.

7 Discusión

Los resultados obtenidos revelan diferencias en la capacidad de cada técnica generativa para preservar la estructura semántica y estadística de los datos clínicos. El método basado en árboles de decisión (CART) demostró ser el más efectivo al generar registros sintéticos con menor distancia de Hamming promedio en dos de los tres datasets evaluados.

El modelo SEM+BN, aunque no logró superar a CART, destacó en representar relaciones causales entre variables clínicas, obteniendo su mejor rendimiento en el

conjunto de Lung Cancer. Su valor reside en la capacidad de incorporar conocimiento experto, lo que puede resultar clave en aplicaciones médicas donde la explicabilidad es crítica.

Por otro lado, la técnica de Cópula Gaussiana presentó el mayor valor de distancia de Hamming en dos de los tres datasets, lo que sugiere dificultades en la generación de datos cuando se trabaja con variables categóricas codificadas binariamente. Aun así, su capacidad para modelar dependencias continuas la mantiene como una opción viable en otros dominios, aunque menos efectiva en este contexto clínico.

Cabe destacar que los resultados reportados corresponden a evaluaciones estadísticas estructurales. Aún no se ha entrenado ningún modelo predictivo con los datos sintéticos generados para contrastar su desempeño frente a los datos reales, por lo que los valores obtenidos deben considerarse como estimaciones iniciales de fidelidad estructural.

8 Conclusiones

Se realizó una comparación entre tres técnicas explicables para la generación de datos sintéticos tabulares en contextos clínicos: árboles de decisión (CART), cópulas Vine gaussianas y una combinación de modelos estructurales con redes bayesianas (SEM+BN). A partir de un conjunto de datos médicos y utilizando un preprocesamiento estandarizado, se evaluó la fidelidad estructural de los datos generados mediante la distancia de Hamming promedio, como una métrica preliminar de similitud entre registros sintéticos y reales.

Los resultados revelaron que el método basado en árboles CART resultó ser el más eficaz en comparación con otras técnicas, produciendo datos a la menor distancia de Hamming y debido a su facilidad de implementación. Esta habilidad para representar de forma explícita las relaciones jerárquicas lo hace una alternativa factible para contextos clínicos. En cuanto a SEM+BN, logró un desempeño intermedio, que puede aplicarse en situaciones donde se requiera entender las conexiones subyacentes entre variables médicas. Por otro lado, el modelo fundamentado en Cópulas Gaussianas experimentó más obstáculos al manejar variables discretas y categóricas, lo cual se manifiesta en sus distancias de Hamming más amplias.

No obstante, hay que resaltar que este análisis se enfocó exclusivamente en la comparación estructural de los datos producidos. Como trabajo a futuro se establecerán modelos de aprendizaje supervisado que permita verificar la eficacia predictiva de los datos artificiales en comparación con los datos reales.

Referencias

Reiter, J.P. (2003). “Using CART to Generate Partially Synthetic, Public Use Microdata”, Journal of Official Statistics, vol. 19(3), pp. 441–456.

Papadaki, E., Vrahatis, A.G., & Kotsiantis, S. (2024). “Exploring Innovative Approaches to Synthetic Tabular Data Generation”, Electronics, vol. 13(10), pp. 1965.

López, C., Smirnov, E., & König, M. (2020). “Copula-based synthetic data generation for machine learning”, Geoscientific Model Development, vol. 13, pp. 427–443.

Barbierato, E., Della Vedova, M.L., Tessera, D., Toti, D., & Vanoli, N. (2022). “A Methodology for Controlling Bias and Fairness in Synthetic Data Generation”, Applied Sciences, vol. 12(9), pp. 4619.

Barbierato, F., Trubiani, C., & Tamburrelli, G. (2022). “Explaining Fairness in Synthetic Tabular Data”, Proceedings of the 44th International Conference on Software Engineering (ICSE '22), pp. 1352–1364. https://doi.org/10.1145/3510003.3510191

Combrink, H.M., De Waal, A., & Van den Berg, G.J.(2021). “Generating Educational Data Using Structural Equation Modeling and Bayesian Networks”, Education and Information Technologies, vol. 26, pp. 4159–4178. https://doi.org/10.1007/s10639-021-10527-2

Scientific Reports (2024). “A novel and fully automated platform for synthetic tabular data

Pinheiro, C., & Ribeiro, R. (2025). “Comparativa de métodos generativos explicables en datos desbalanceados”, Proceedings del Congreso Iberoamericano de Datos Sintéticos 2025, pp. 85–97.

Drechsler, J. (2011). Synthetic Datasets for Statistical Disclosure Control: Theory and Implementation. Springer. https://doi.org/10.1007/978-1-4419-7802-8

Pinheiro, C., & Ribeiro, R. (2025). “CART-based Synthetic Tabular Data Generation for Imbalanced Regression”, arXiv preprint arXiv:2506.02811. https://arxiv.org/abs/2506.02811

Combrink, H.M., Marivate, V., & Rosman, B.(2021). “Comparing Synthetic Tabular Data Generation Between a Probabilistic Model and a Deep Learning Model for Education Use Cases”, arXiv preprint arXiv:2210.08528. https://arxiv.org/abs/2210.08528

Shaposhnyk, S., Machanavajjhala, A., & Salamatian, S. (2025). “TabSyn: Benchmarking Explainable Synthetic Data Generators”, arXiv preprint arXiv:2504.11547. https://arxiv.org/abs/2504.11547

Zhao, D. (2025). “Privacy-Preserving Hamming Distance Computation with Property-Preserving Hashing”, arXiv preprint arXiv:2503.17844. https://arxiv.org/abs/2503.17844

Boneh, D., Kim, S., & Lin, S. (2024). “Improving Hamming-Distance Computation for Adaptive Matching”, arXiv preprint arXiv:2407.05430. https://arxiv.org/abs/2407.05430

Preneel, B., et al. (2024). “Improving Hamming-Distance Computation for Adaptive Similarity Search”, ResearchGate. https://www.researchgate.net/publication/357619499_Improving_Hamming-Distance_Computation_for_Adaptive_Similarity_Search_Approach

GitLab Blog (2024). Synthesizing tabular data with Gaussian Copulas. Recuperado de: https://about.gitlab.com/blog/2024/03/05/synthesizing-tabular-data-with-gaussian-copulas/

Fedesoriano. (2021). Conjunto de datos de predicción de insuficiencia cardíaca. Kaggle. Recuperado de https://www.kaggle.com/fedesoriano/heart-failure-prediction.

Drechsler, J., & Reiter, J.P. (2011). “An empirical evaluation of easily implemented, non-parametric methods for generating synthetic datasets”, Computational Statistics & Data Analysis, vol. 55(12), pp. 3232–3243.

Nagler, T., & Vatter, T. (2020). Vinecopulib. Zenodo. https://doi.org/10.5281/zenodo.4287554.

Nagler, T., Schellhase, C., & Czado, C. (2017). “Nonparametric estimation of simplified vine copula models: Comparison of methods”, Dependence Modeling, vol. 5(1), pp. 99–120.

Kaur, D., Sobiesk, M., Patil, S., Liu, J., Bhagat, P., Gupta, A., & Markuzon, N. (2021). “Application of Bayesian Networks to Generate Synthetic Health Data”, Journal of the American Medical Informatics Association, vol. 28(4), pp. 801–811.

Capítulo 3

Evaluación comparativa de modelos Transformer preentrenados (BERT, RoBERTa, RoBERTuito y DistilBERT) para la clasificación de emociones en textos breves en español

Ángel de Jesús Elvis Gutiérrez Pacheco, Keira Marisa Garrido Aguirre, Iván Pérez Sánchez, Michell León Paredes, Luis Yael Méndez Sánchez

Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla, Avenida San Claudio y 14 Sur, Ciudad Universitaria, Puebla, México
{angel.gutierrezpa, keira.garridoa, ivan.perezsa, michell.leon}@alumno.buap.mx, luis.mendezsanchez@correo.buap.mx

Resumen. El análisis de emociones en textos es un área de investigación importante dentro del Procesamiento del Lenguaje Natural. Este estudio aborda la clasificación de emociones en textos escritos en español mediante el ajuste fino de cuatro modelos: BERT, RoBERTa, RoBERTuito y DistilBERT. Para ello, se ha conformado un corpus de 4,092 textos generados por estudiantes de la Facultad de Ciencias de la Computación, de la Benemérita Universidad Autónoma de Puebla, a través de un cuestionario para obtener patrones psicológicos. El cuestionario induce los textos intencionadamente a una de las seis emociones básicas (Felicidad, Tristeza, Disgusto, Ira, Miedo y Sorpresa) propuestas por Paul Ekman, psicólogo pionero en el estudio de las emociones. Tras una fase de preprocesamiento, los datos fueron divididos en 80 % entrenamiento y 20 % prueba. La evaluación se ha basado en métricas estándar de clasificación, así como el análisis de las matrices de confusión para identificar las principales confusiones en la predicción. Los resultados obtenidos muestran que RoBERTa es el modelo más eficaz, alcanzando una exactitud de 0.9028, un AUC de 0.9849 y un F1-score de 0.9006, con una confusión más notable entre Disgusto y Tristeza del 13 %. Le siguen DistilBERT (F1 = 0.8871, AUC = 0.9809), RoBERTuito (F1 = 0.7173, AUC = 0.9379) y BERT (F1 = 0.6266, AUC = 0.8896).

Palabras Clave: Clasificación de Emociones, Procesamiento de Lenguaje Natural, Transformer, Aprendizaje Profundo, Aprendizaje Automático.

1 Introducción

Hoy en día la generación de texto es una actividad ubicua: las personas producen texto de manera masiva por distintos medios como redes sociales, blogs, correos electrónicos, etc. A partir de ello, el Procesamiento del Lenguaje Natural surge como una rama de la

Inteligencia Artificial que tiene como objetivo implementar algoritmos en sistemas computacionales para que estos sean capaces de comprender holísticamente el lenguaje humano (García, 2021), de tal manera que realicen diferentes actividades como Traducción Automática, Síntesis de Textos o Generación de Texto (Gillioz et al., 2020).

Por otro lado, las emociones son procesos psicológicos complejos que sirven como formas de adaptación y respuesta a un entorno. La comprensión de las emociones explica qué pasa cuando ocurre una reacción ante determinados estímulos, ya sean internos o externos; así como por qué sucede (Chóliz, 2005).

En este contexto, el Análisis de Emociones es un campo de investigación dentro del Procesamiento del Lenguaje Natural que comprende a la lingüística, la psicología y a las ciencias computacionales. Esta área abarca diferentes objetivos, sin embargo, el presente artículo se enfoca en la clasificación de sentimientos, cuyo fin es determinar el sentimiento que trae consigo un texto dado; para ello se necesita de un marco que determine las emociones en las que el texto puede recaer (Wankhade, 2022).

Además, en los últimos años, el desarrollo de la arquitectura Transformer, propuesta por Vaswani et al. (2017), marcó un hito en el estado del arte y catalizó la creación de nuevos modelos del lenguaje capaces de tener una mayor comprensión de un idioma, aspecto fundamental para alcanzar el objetivo del Análisis de Emociones (Rojas, 2024). Uno de estos modelos es BERT, presentado por Devlin et al. (2019), caracterizado por una arquitectura bidireccional, mediante la cual se permite una comprensión más completa y precisa del texto.

Así pues, este artículo presenta el desarrollo de cuatro modelos basados en BERT para la clasificación de emociones a respuestas textuales de estudiantes universitarios de Puebla, México; con el objetivo de identificar cuál es el que tiene un mejor desempeño en esta tarea, para ello, se toma como métrica principal el F1-Score debido a que representa la media armónica entre precisión y recall, proporcionando una evaluación equilibrada de cuántas etiquetas positivas fueron detectadas correctamente, sin embargo, también se analiza la precisión, la exactitud, el área bajo la curva ROC-AUC y la matriz de confusión, todo esto con el fin de contar con un criterio de comparación bien estructurado. Esta investigación forma parte de un proyecto más grande enfocado a la detección de patrones psicológicos en estudiantes universitarios a partir de texto escrito por ellos.

El artículo se organiza de la siguiente manera: la Sección 2 detalla los fundamentos teóricos en los que se sustenta esta investigación. La Sección 3 presenta una revisión de trabajos relacionados, contextualizando el estado del arte. La Sección 4 describe la metodología de trabajo seguida. La Sección 5 expone los resultados obtenidos de la evaluación de los modelos. Finalmente, la Sección 6 discute los resultados para generar conclusiones, en conjunto con un cuestionamiento sobre las futuras líneas de trabajo que podrían expandir esta investigación.

2 Fundamentos Teóricos

El núcleo del artículo se sustenta en la Minería de Datos (MD), disciplina encargada de aplicar técnicas estadísticas a conjuntos de datos con el fin de identificar patrones en

ellos. Esta área se encuentra relacionada íntimamente con la Inteligencia Artificial (IA), rama de las Ciencias Computacionales que ha ganado popularidad en los últimos. Sin embargo, desde hace décadas, los datos disponibles se han cargado de variedad, veracidad y volumen; características que hacen su análisis una actividad imprescindible (Rajput, 2019).

Bajo este contexto, el Procesamiento del Lenguaje Natural (PLN) surge como área de investigación dentro de la IA para crear técnicas que hagan a una máquina capaz de procesar el lenguaje humano y traducirlo a un formato que pueda entender (Rajput, 2019). Esto genera diversas aplicaciones, pero una de ellas es el Análisis de Emociones, el cual ha tenido un crecimiento notorio a partir de 2018 (Miriam, 2024).

El Análisis de Emociones (AE), también conocido como Análisis de Opiniones o Minería de Opiniones, se encarga en la identificación automática y la categorización de emociones expresadas dentro de un texto. Este proceso es aplicable a diferentes sectores de la sociedad como las redes sociales, el ámbito empresarial, la medicina, etc. (Barrón, 2024).

3 Estado del Arte

Como se menciona, el AE ha obtenido gran interés en ámbitos académicos y empresariales desde 2018, sin embargo, es un campo que ha existido desde los sesenta (Robles, 2024). Revisar la literatura existente ayuda a tener una comprensión del contexto de esta investigación y conocer los enfoques más recientes en esta área.

Los modelos más tradicionales y mayormente usados son aquellos que utilizan técnicas estadísticas de Machine Learning (ML). Estos modelos están basados usualmente en las características superficiales del texto tal como la bolsa de palabras o los n-gramas; e incluyen modelos como Naive Bayes (NB), Regresión Logística (RL), Máquinas de Vectores de Soporte (SVM), etc. (Zhang, 2024).

Con el crecimiento del Deep Learning (DL) se hizo un progreso considerable en los modelos de AE basados en redes neuronales profundas. A diferencia de los modelos estadísticos convencionales, los modelos basados en redes neuronales pueden aprender automáticamente representaciones semánticas de alto nivel en el texto, sin requerir implementaciones complejas (Zhang, 2024).

Las Redes Neuronales Recurrentes (RNNs) y las Redes Neuronales Convolucionales (CNNs) son las estructuras más usadas; para la primera, además, existen dos variantes mejoradas: las Redes LSTM (Long and Short-Term Memory) y las Redes GRU (Gated Cycle Unit), las cuales manejan de manera efectiva la información semántica del texto y capturan patrones complejos y dependencias a largo plazo en él, aumentando su desempeño en tareas tales como el AE (Zhang, 2024; Robles, 2024).

No obstante, Vaswani et al. (2017) introdujeron la arquitectura Transformer, revolucionando el campo de la IA. Esta arquitectura aplica un conjunto en evolución de técnicas matemáticas llamadas atención o atención propia, para detectar formas sutiles en que los elementos de datos en una serie se influencian y dependen entre sí (Rojas, 2024); aplicado al PLN, esta arquitectura consigue un entendimiento más profundo de la gramática del lenguaje y no solo de la semántica de las palabras

concretas. Casi todos los modelos actuales para el PLN están inspirados, de una u otra forma en esta arquitectura (García, 2021).

Con base en el Transformer, se han creado miles de modelos inspirados en él. Uno de los más conocidos y que aún sigue siendo usado es el modelo propuesto por Devlin et al. (2019) conocido como BERT (Bidirectional Encoder Representations from Transformers).

BERT es un modelo que usa solo la parte del codificador y con capas bidireccionales. Su mayor novedad es el modelo del lenguaje enmascarado (MLM); este es una forma de entrenar un modelo autosupervisado que enmascara aleatoriamente un porcentaje de las palabras en el texto. Esto ayuda al modelo a aprender, no solo lo que precede a una palabra, sino lo que le antecede, permitiéndole un mayor entendimiento general del lenguaje (García, 2021).

A partir de BERT, han surgido versiones mejoradas de él, que varían en aspectos como el número de parámetros, hiperparámetros o, incluso, el idioma sobre el que son entrenados. En este sentido, los modelos DistilBERT, RoBERTa y RoBERTuito, presentados por Sanh et al. (2020), Liu et al. (2019) y Pérez et al. (2022), respectivamente; representan modelos clave para experimentar con ellos, pues son algunos de los más conocidos, a diferencia de RoBERTuito, sin embargo, este es un modelo enfocado en español, por lo que resulta interesante comparar su rendimiento con los otros modelos.

Como producto de estos modelos, se han derivado investigaciones en el área del AE enfocado en el idioma español. Una de las más relevantes es el trabajo presentado por García (2021), quien realiza el proceso completo para generar varios modelos basados en BERT y capaces de clasificar texto: desde la elección de los conjuntos de datos, su limpieza, preprocesamiento y la forma en la que fueron entrenados dichos modelos; además, detalla algunos algoritmos utilizados para generar más datos, técnica conocida como oversampling. Este trabajo, aparte de servir como referente para comprender el proceso de creación de modelos del lenguaje basados en Transformers, funge como base teórica sólida para comprender la forma en la que operan los modelos realizados a partir de BERT.

Por su parte, Rojas (2024) realiza el proceso completo para crear un modelo basado en RoBERTa y orienta el AE hacia quejas en atención al cliente. Esta investigación actúa como ejemplo para la aplicación de estos modelos en la industria. También describe en qué consisten las métricas de evaluación usadas en la presente investigación, realiza modelos de ML tradicional y obtiene como resultado que los Transformers rebasan a los modelos clásicos, sin embargo, resalta el alto costo computacional en la implementación de un modelo Transformer, pero concluye que el retorno de inversión es mayor en entornos productivos a diferencia de uno tradicional.

De manera similar, Barrón et al. (2024) crearon un corpus de 18,000 muestras no etiquetadas y 853 muestras etiquetadas manualmente. Este corpus se enfoca en un contexto estudiantil mexicano, por lo que captura la diversidad y complejidad del español hablado en México, así como de la cultura en este contexto. Aunado a lo anterior, desarrollaron tres modelos tipo BERT para detectar toxicidad en textos, tarea que recae dentro del AE.

Asimismo, Robles et al. (2024) comparan el desempeño de los modelos BERT, RoBERTa, RoBERTuito, DistilBERT y GPT-4 en el AE en español. En sus resultados mencionan que DistilBERT y BERT son los modelos con el rendimiento más bajo, mientras que RoBERTuito es el que tiene el mejor performance. Además, concluyen que los modelos tipo BERT son mejores que los tipo GPT, pues estos últimos no están orientados a la clasificación de texto, sin embargo, declaran que aun así son capaces de realizar la tarea.

Finalmente, Merchán (2024) explica el funcionamiento interno de modelos como BERT, RoBERTa y DistilBERT. Además, describe en qué consisten las métricas de evaluación usada en la presente investigación y qué factores pueden influir en un bajo rendimiento de los modelos, haciendo hincapié en el lenguaje base del modelo y el tamaño del conjunto de datos.

4 Metodología

En esta sección se presenta la metodología propuesta para llevar a cabo este trabajo de investigación y lograr el desarrollo del modelo de clasificación de emociones.

4.1 Origen del Conjunto de Datos

El conjunto de datos utilizado fue generado a través de un cuestionario diseñado con el objetivo principal de evocar emociones concretas mediante preguntas cuidadosamente planteadas. Este instrumento fue aplicado mayormente a estudiantes universitarios de diversas áreas y rangos de edad.

Cada pregunta fue orientada a inducir una emoción en específico: Felicidad (preguntas 1 y 2), Tristeza (3 y 4), Disgusto (5 y 6), Ira (7 y 8), Miedo (9 y 10), y Sorpresa (11 y 12). A continuación, se presenta el conjunto de preguntas utilizadas en el cuestionario:

1. Recuerda un momento en el que lograste algo que realmente deseabas. ¿Cómo fue la experiencia y qué impacto tuvo en tu vida?
2. Describe una ocasión en la que alguien hizo algo especial o inesperado por ti y te hizo sentir bien. ¿Cómo reaccionaste?
3. Piensa en una ocasión en la que perdiste algo o a alguien importante para ti. ¿Cómo viviste ese momento y qué cambió después de ello?
4. Recuerda un día en el que sentiste que todo te salía mal. ¿Qué sucedió y cómo te sentiste al respecto?
5. Describe una situación en la que experimentaste algo que te resultó desagradable y quisiste evitar. ¿Cómo fue ese momento?
6. Recuerda una ocasión en la que viste o viviste algo que te pareció totalmente inaceptable. ¿Cómo reaccionaste y qué pensaste al respecto?
7. Piensa en un momento en el que sentiste que alguien fue injusto contigo. ¿Qué ocurrió y cómo reaccionaste?

8. Describe una ocasión en la que te sentiste frustrado(a) porque no tomaron en cuenta tu opinión o esfuerzo. ¿Cómo reaccionaste ante esta situación y cómo la manejaste?

9. Recuerda una ocasión en la que tuviste que enfrentarte a algo incierto o desconocido. ¿Cómo fue la experiencia y qué sentiste en ese momento?

10. Describe un evento en el que sentiste que algo estaba fuera de tu control y no sabías qué hacer. ¿Cómo reaccionaste y qué pasó después?

11. Piensa en un momento en el que ocurrió algo totalmente inesperado en tu vida. ¿Cómo fue y qué pasó después?

12. Recuerda una ocasión en la que recibiste una noticia o viviste un evento que nunca imaginaste. ¿Cómo reaccionaste y qué impacto tuvo en ti?

En total, se recolectaron las respuestas de 341 participantes únicos, 339 de ellos identificados mediante un código alfanumérico (ID_Unico) asignado para preservar el anonimato de los participantes, obteniendo de esta manera 4,092 textos únicos.

4.2 Estructura del Conjunto de Datos

El conjunto de datos consta de 342 filas (una por participante más el encabezado) y 26 columnas, de estas, 12 corresponden a variables demográficas y otras 12 a las preguntas diseñadas para inducir emociones. De dichas 26 columnas, 24 son de tipo de dato object y 2 de tipo float64.

Las columnas demográficas incluyen información como edad, género, nivel socioeconómico, grado de estudios, carrera, situación académica y laboral, al igual que lugar de origen. Mientras que las variables emocionales corresponden a respuestas abiertas frente a estímulos narrativos diseñados para evocar una de las 6 emociones predominantes.

La completitud de las variables demográficas es alta, en su mayoría superan el 90%. A su vez, de las 12 preguntas sobre emociones, 11 recibieron una tasa de respuesta de 100%, siendo la pregunta 5 la única con un porcentaje menor quedando en 99.71% con 340 respuestas en lugar de 341.

4.3 Limpieza y Preprocesamiento

En esta fase, el corpus original, obtenido a partir de cuestionarios y registros de usuarios, se redujo eliminando aquellas columnas irrelevantes para el análisis, de modo que cada registro quedara representado únicamente por un texto y su etiqueta emocional asociada. A continuación, los datos se reestructuraron para disponer de una única columna de enunciados y otra con la emoción correspondiente, lo que facilitó su tratamiento como pares texto-etiqueta.

Una vez organizado el conjunto, se aplicaron filtros de longitud y complejidad: se descartaron los enunciados con menos de diez caracteres o con menos de cuatro palabras, de forma que se preservara únicamente el material textual con suficiente contenido informativo. Posteriormente, se procede a la normalización léxica, suprimiendo abreviaturas y expresiones informales mediante un diccionario de mapeo

("k" por "que", "xq" por "porque", etc.) y eliminando espacios o caracteres redundantes. El resultado fue un corpus homogéneo y libre de ruidos que pudiera optimizar la capacidad de aprendizaje de los modelos de Transformers seleccionados.

Para preparar los datos con vistas al entrenamiento, las etiquetas textuales de las emociones fueron codificadas numéricamente en orden alfabético, garantizando la reproducibilidad del mapeo. A continuación, el conjunto se dividió de manera estratificada en subconjuntos de entrenamiento (80 %) y prueba (20 %), conservando la proporción de cada categoría emocional en ambos conjuntos de datos. Finalmente, cada enunciado se tokenizó con el vocabulario específico del transformer empleado (p. ej., RoBERTa, DistilBERT, BERT o RoBERTuito), convirtiendo las secuencias de texto en tensores de identificadores de tokens y máscaras de atención de longitud fija.

Esos tensores quedaron encapsulados en DataLoaders de PyTorch, capaces de entregar lotes balanceados durante el ajuste fino, lo que permitió entrenar de manera eficiente y consistente los distintos modelos evaluados.

4.4 Entrenamiento

Durante la fase de ajuste fino se cargaron los modelos preentrenados RoBERTa, DistilBERT, BERT y RoBERTuito con sus cabezas de clasificación configuradas para seis etiquetas emocionales; el procedimiento fue idéntico salvo por el identificador de cada variante, primero se transfirieron los pesos a la GPU y luego se congelaron las seis capas inferiores de la arquitectura base para mitigar el riesgo de sobreajuste en un corpus de tamaño moderado, a continuación se definió la función de pérdida de entropía cruzada y se empleó el optimizador AdamW con tasa de aprendizaje de 2×10^{-5} y épsilon de 1×10^{-8}, valores alineados con las recomendaciones originales de BERT para permitir ajustes suaves de los pesos sin desestabilizar la función de pérdida. La semilla aleatoria se fijó en 42 para garantizar la reproductibilidad de los experimentos. El tamaño de lote de 16 muestras responde a limitaciones de memoria de GPU y el máximo de 256 tokens por secuencia preserva el contexto necesario para capturar expresiones emocionales.

La decisión de congelar las seis capas iniciales se fundamenta en la modularidad de los bloques de atención de BERT y sus variantes. Cada modelo cuenta con doce capas de codificadores apilados, donde las capas inferiores capturan patrones lingüísticos genéricos (morfología, sintaxis, semántica de bajo nivel) y las superiores refinan representaciones orientadas a tareas específicas. Al bloquear el ajuste de los parámetros de las seis capas iniciales, se preserva el conocimiento lingüístico profundo adquirido durante el preentrenamiento, reduciendo el número de parámetros entrenables y, por ende, la complejidad del espacio de búsqueda. De este modo, el modelo concentra su capacidad de adaptación en las capas superiores, que son precisamente las que modelan la relación entre la entrada y las etiquetas emocionales, lo que suele traducirse en mayor estabilidad y mejores resultados cuando el corpus disponible para el fine tuning es de tamaño limitado.

Para controlar dinámicamente el proceso de aprendizaje se incorporó un scheduler lineal sin pasos de calentamiento y con decaimiento proporcional al número total de iteraciones (batches × épocas), el entrenamiento se extendió a cinco épocas iterando

lote a lote mediante un DataLoader que extrae secuencias previamente tokenizadas y sus máscaras de atención, en cada lote se calcularon predicciones, se computó la pérdida, se propagaron gradientes y se actualizaron los pesos ajustando simultáneamente la tasa de aprendizaje según la política establecida por el scheduler, tras cada época se evaluaron las métricas de pérdida y exactitud en los conjuntos de entrenamiento y validación para monitorizar la convergencia y detectar posibles signos de sobreajuste, finalmente se almacenó en disco el estado del modelo ajustado para su posterior evaluación.

5 Resultados

A continuación, se muestran los resultados de la evaluación de rendimiento de los modelos previstos. La evaluación se basa en métricas estándar de clasificación: Precisión (Precisión), Sensibilidad (Recall), Puntuación F1 (F1 Score), Exactitud (Accuracy), Matriz de Confusión y Área Bajo la Curva ROC (AUC). La Tabla 1 presenta los resultados de la evaluación realizada a cada modelo.

Tabla 1. Resultados de evaluación de cada modelo

Modelo	Exactitud	Precisión	Sensibilidad	AUC	F1-Score
RoBERTa	0.9028	0.9056	0.9011	0.9849	0.9006
DistilBERT	0.8880	0.8894	0.8878	0.9809	0.8871
BERT	0.6276	0.6485	0.6275	0.8896	0.6266
RoBERTuito	0.7193	0.7196	0.7193	0.9379	0.7173

Por otro lado, la Tabla 2 muestra los resultados obtenidos de las matrices de confusión de cada modelo, además, en las siguientes subsecciones se discuten estos resultados.

Tabla 2. Resultados por emoción y principales confusiones por modelo

Modelo	Disgusto	Felicidad	Ira	Miedo	Sorpresa	Tristeza	Confusión Principal
RoBERTa	0.73	0.96	0.91	0.93	0.94	0.94	Disgusto–Tristeza (13 %)
DistilBERT	0.87	0.96	0.91	0.87	0.78	0.94	Sorpresa–Felicidad (9 %)
BERT	0.56	0.78	0.51	0.75	0.47	0.69	Disgusto–Miedo (20 %), Sorpresa–Miedo (20 %)
RoBERTuito	0.70	0.86	0.65	0.70	0.56	0.84	Ira–Disgusto (20 %), Sorpresa–Felicidad (18 %)

5.1 RoBERTa

Los resultados obtenidos con este modelo indican un buen desempeño, aceptable y equilibrado, con elevados niveles de precisión en las clases de Felicidad (0.96), Miedo (0.93), Ira (0.91), Tristeza (0.94) y Sorpresa (0.94). Sin embargo, se evidencian confusiones en ciertas categorías (como se muestra en la Figura 1), entre Disgusto e Ira (0.13), lo que indica que, aunque el modelo es capaz de identificar patrones aun presenta dificultades para distinguir con mayor precisión.

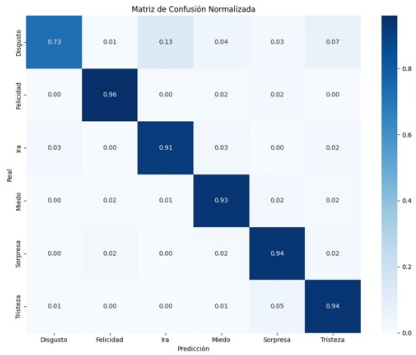


Figura 1. Matriz de confusión del modelo RoBERTa

5.2 DistilBERT

El modelo identifica correctamente emociones como Felicidad (0.96), Ira (0.91) y Tristeza (0.94) con alta precisión. No obstante, las categorías con mayor confusión son Sorpresa y Felicidad (0.9), lo que puede deberse a la cercanía semántica entre ambas emociones. En conjunto el modelo demuestra una buena capacidad de predicción entre emociones similares, siendo especialmente preciso en las emociones más marcadas como Felicidad y Tristeza como se muestra en la Figura 2.

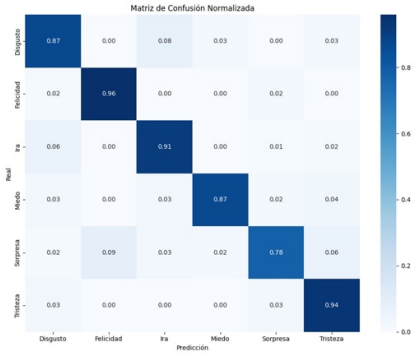


Figura 2. Confusiones obtenidas del modelo DistilBERT.

5.3 BERT

La matriz de confusión (Figura 3) muestra un leve desempeño de precisión en las categorías mostrando como las más destacadas Felicidad (0.78) y Miedo (0.75) lo cual indica que el modelo no logra identificar satisfactoriamente los patrones emocionales. No obstante, también se observan confusiones en las categorías emocionalmente similares, como Disgusto y Miedo (0.20) y entre Sorpresa y Miedo (0.20), lo que indica una leve dificultad para distinguir entre estados emocionales. En general este modelo cumple con un bajo estándar para desempeñar un rendimiento balanceado.

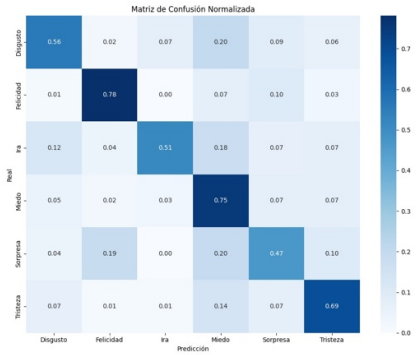


Figura 3. Matriz de confusión obtenida de la evaluación del modelo BERT.

5.4 RoBERTuito

La matriz de confusión de la Figura 4, refleja un buen desempeño para identificar con precisión las emociones como Felicidad (0.86) y Tristeza (0.84), lo que demuestra su capacidad para reconocer patrones emocionales. No obstante, se observa confusión en la categoría emocionalmente similar, como Ira y Disgusto (0.20) y entre Sorpresa y Felicidad (0.18). En general el modelo muestra un rendimiento equilibrado en algunas categorías.

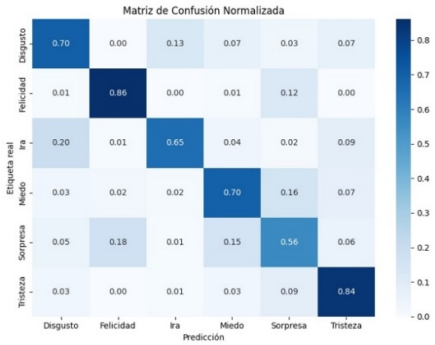


Figura 4. Matriz de confusión del modelo RoBERTuito.

6 Conclusiones

Este documento aborda el problema de la clasificación de emociones en texto mediante la integración de técnicas de PLN junto con modelos de aprendizaje supervisado basados en Transformers. La metodología incluyó la evaluación cuantitativa del rendimiento de cuatro modelos (BERT, RoBERTa, DistilBERT y RoBERTuito), con el propósito de comparar su precisión y eficacia.

Los resultados evidencian que el modelo RoBERTa alcanzó el mejor desempeño con un F1-Score de 0.9006, superando a sus contrapartes. Esto permite afirmar que el uso de modelos preentrenados, especialmente en textos en español, ofrece alternativas más eficaces. Además, la matriz de confusión muestra que emociones como Felicidad, Ira, Miedo, Sorpresa y Tristeza fueron clasificadas con alta precisión, mientras que categorías como Disgusto e Ira presentaron niveles notorios de confusión.

En comparación con trabajos similares, Robles et al. (2024) señalan que la implementación de modelos tipo BERT resulta de gran ayuda, ya que se adaptan al idioma objetivo. La literatura también destaca que RoBERTa, al ser un modelo preentrenado con textos informales de redes sociales, es más sensible a expresiones emocionales de carácter coloquial, similares a las utilizadas por los estudiantes en este corpus. Aunque RoBERTa fue originalmente diseñado para datos en inglés, existen versiones entrenadas específicamente para el español.

Las fortalezas de este estudio radican en el uso de metodologías sólidas, el balance adecuado de categorías en los conjuntos de entrenamiento y prueba, así como la aplicación de diversas métricas que permiten un análisis detallado del rendimiento. No obstante, los datos empleados provienen únicamente de estudiantes universitarios, lo que ocasiona que los modelos aprendan patrones específicos de este grupo y no generalicen correctamente hacia otros contextos.

Como trabajo futuro, se aplicará validación cruzada (K-Fold) con valores de k entre 5 y 9, con el objetivo de obtener una estimación más precisa de los modelos. Finalmente, es importante señalar que esta investigación forma parte de un proyecto más amplio enfocado en la detección de patrones psicológicos en estudiantes universitarios a partir de textos escritos por ellos, lo cual abre la posibilidad de generar aportes valiosos para los campos de la psicología y la educación.

Referencias

Barrón, M., Zatarain, R., Camacho, R. et al. (2024). PLN con Transformers para detección de toxicidad: construcción y evaluación de corpus para la plataforma MisProfesores.com. In Research in Computing Science (pp. 137–145).

Chóliz, M. (2005). Psicología de la emoción. Recuperado de https://www.uv.es/~choliz/Proceso%20emocional.pdf.

Devlin, J., Chang, M., Lee, K. et al. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019) (pp. 4171–4186).

García, G. (2021). Modelos de Transformers para Clasificación de Texto. [Trabajo fin de máster]. Universidad Politécnica de Madrid.

Gillioz, A., Casas, J., Mugellini, E. et al. (2020). Overview of the Transformer-based Models for NLP Tasks. In Federated Conference on Computer Science and Information Systems. (pp. 179–183).

Liu, Y., Ott, M., Goyal, N. et al. (2019), RoBERTa: A Robustly Optimized BERT Pretraining Approach. [Preprint]. arXiv. https://arxiv.org/abs/1907.11692

Merchán, E. (2024). Aplicación de Modelos Transformers para Clasificar Textos en Idioma Español. [Trabajo fin de grado]. Universidad Estatal Península de Santa Elena.

Miriam, F., Curry, A., Cercas, A. et al. (2024). Emotion Analysis in NLP: Trends, Gaps and Roadmap for Future Directions. [Preprint]. arXiv. https://arxiv.org/abs/2403.01222

Pérez, J., Furman, D., Alonso, L. et al. (2022). RoBERTuito: a pre-trained language model for social media text in Spanish. [Preprint]. arXiv. https://arxiv.org/abs/2111.09453

Rajput, A. (2019). Natural Language Processing, Sentiment Analysis and Clinical Analytics. [Preprint]. arXiv. https://arxiv.org/abs/1902.00679

Robles, C., Carrillo, M., Meza, I. (2024). Detección de emociones en texto en español utilizando transformers. In Abstraction & Application. (pp. 87–97).

Rojas, E. (2024). Aplicación de Modelos de Transformer para la Clasificación y Análisis de Quejas en Atención al Cliente. [Trabajo fin de grado]. Instituto Tecnológico y de Estudios Superiores de Occidente.

Sanh, V., Debut, L., Chaumond, J. et al. (2020). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. [Preprint]. arXiv. https://arxiv.org/abs/1910.01108

Vaswani, A., Shazeer, N., Parmar, N. et al. (2017). Attention is All You Need. In Advances in Neural Information Processing Systems (pp. 5998–6008).

Wankhade, M., Sekhara, A., Kulkarni, C. (2022). A survey on sentiment analysis methods, applications and challenges. In Artificial Intelligence Review (pp. 5731–5780).

Zhang, B., Xiao, J., Yan, H. et al. (2024). Review of NLP Applications in the Field of Text Sentiment Analysis. In Journal of Industrial Engineering and Applied Science (pp. 28–34).

Capítulo 4

Clasificación de emociones en textos en español mediante un enfoque híbrido de Transformers y Aprendizaje Automático

Jorge León Vivanco, Luis Yael Méndez Sánchez, María Josefa Somodevilla García, Darnes Vilariño Ayala, Gustavo Emilio Mendoza Olguín

Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla, Avenida San Claudio y 14 Sur, Ciudad Universitaria, Puebla, México
jorge.leonvi@alumno.buap.mx, {luis.mendezsanchez, maria.somodevilla, darnes.vilarino, gustavo.mendozaol}@correo.buap.mx

Resumen. En este trabajo se desarrolla un modelo híbrido para la clasificación de seis emociones básicas (felicidad, tristeza, disgusto, ira, miedo y sorpresa) en textos en español, combinando RoBERTa afinado con clasificadores tradicionales (regresión logística y SVM) mediante meta-clasificación. Para ello construimos un corpus propio de 4 200 textos obtenidos a partir de 350 participantes universitarios que respondieron 12 preguntas abiertas (dos por emoción), con etiquetado manual. En la evaluación sobre el 20 % del corpus (840 textos; 140 por clase), el sistema híbrido alcanzó Accuracy = 0.91 y F1 macro = 0.91, superando a los modelos individuales; por clase, se observaron F1 altos en miedo (0.98) y disgusto (0.98), con menor recall en felicidad (0.76). En comparación, las líneas base tradicionales se situaron alrededor de 0.78 de accuracy. Estos resultados muestran que la combinación de representaciones contextuales y clasificadores clásicos mejora la detección de emociones en español y sientan bases para aplicaciones educativas y de bienestar.

Palabras clave: Clasificación de emociones, Procesamiento de Lenguaje Natural, Transformers, Aprendizaje Automático, modelo híbrido.

1 Introducción

En esta sección se presenta el trasfondo y la motivación del estudio realizado: se inicia con un análisis del estado actual de la detección automática de emociones en textos en español, apoyándose en hallazgos recientes como los de Maria et al. (2024), para luego plantear el problema central y los desafíos metodológicos a los que se enfrenta el modelo híbrido propuesto.

1.1 Contexto general

El crecimiento de textos en español en medios digitales demanda sistemas automáticos que reconozcan emociones con precisión. Esta capacidad es crucial en la atención al

cliente, la educación y la salud mental, pero su desarrollo se complica por variaciones dialectales y la limitada disponibilidad de datos anotados. Este trabajo propone un enfoque híbrido para abordar dichas limitaciones y mejorar la clasificación multicategoría de emociones (Chauhan et al., 2022).

La cantidad de datos presenta nuevos retos como oportunidades para áreas clave de nuestra vida diaria como la educación, la psicología y la atención al cliente. La detección de emociones en textos escritos adquiere especial relevancia en ámbitos como la atención al cliente, la educación y la salud pública, ya que permite identificar el estado afectivo de las personas y orientar intervenciones, optimizar respuestas automatizadas y mejorar la toma de decisiones operativas.

1.2 Justificación

Los métodos tradicionales basados en diccionarios léxicos o reglas simples presentan muchas dificultades para identificar la vaguedad y las dependencias contextuales en español, sobre todo en casos que incluyen negaciones o intensificadores (Thakur et al., 2024). Por otra parte, modelos clásicos de Machine Learning como Naive Bayes o SVM ofrecen legibilidad y eficiencia, pero carecen de la riqueza semántica necesaria para especificar matices emocionales. En contraste, Vaswani et al. (2017) introdujeron la arquitectura Transformer y demostraron que su mecanismo de atención bidireccional es capaz de poder interpretar toda la cadena de palabras de una vez, a diferencia de los modelos direccionales que procesan en secuencia, capturando relaciones a largo plazo. RoBERTa, a su vez, optimiza BERT mediante un preentrenamiento ampliado y la eliminación de la tarea de predicción de la siguiente oración (Liu et al., 2019), pero ambos enfoques demandan recursos computacionales elevados.

1.3 Planteamiento del problema

A pesar de los avances en modelos de lenguaje preentrenados, la mayoría de las investigaciones se han centrado en el inglés o en tareas binarias de sentimiento, dejando un vacío metodológico en la clasificación multicategoría de emociones específicas en español (Sindhu et al., 2023). Asimismo, los estudios que combinan Transformers con algoritmos clásicos han mostrado mejoras en F1-score y AUC-ROC, pero rara vez abordan el caso de seis emociones básicas (felicidad, tristeza, disgusto, ira, miedo y sorpresa) de forma exhaustiva (Srivastava et al., 2023; Singh y Srivastava, 2023). Por ello, se propone el diseño de un flujo de procesamiento híbrido que combine las ventajas de las representaciones contextuales profundas y de los clasificadores tradicionales, y cuya eficacia se evalúe sobre un corpus balanceado y anotado manualmente en español.

1.4 Importancia y aportaciones del estudio

En esta investigación se conformó un corpus propio de 4 200 textos (350 respuestas por cada uno de los 12 ítems), obtenido a partir de encuestas aplicadas a estudiantes de la Facultad de Ciencias de la Computación de la Benemérita Universidad Autónoma de Puebla. Las instancias fueron anotadas manualmente según las seis emociones básicas,

lo que proporciona una base equilibrada y adecuada para los experimentos de clasificación. Sobre dicho recurso se diseñó un modelo híbrido que integra el fine-tuning de RoBERTa, cuyas mejoras frente a BERT se han documentado previamente (Liu et al., 2019) con clasificadores tradicionales (regresión logística y máquinas de vectores de soporte). Las salidas de estos componentes se combinan mediante una capa de meta-clasificación para aprovechar simultáneamente la riqueza semántica de las representaciones profundas y la solidez de los algoritmos clásicos (Srivastava et al., 2023).

2 Fundamentos teóricos

En esta sección se exponen los principios y mecanismos que sustentan el enfoque híbrido; en primer lugar, se revisa la arquitectura Transformer, responsable de la capacidad para modelar dependencias a largo plazo en secuencias textuales (Vaswani et al., 2017). Luego, se analiza el modelo RoBERTa y sus optimizaciones sobre BERT (Liu et al., 2019). Finalmente, se describen los clasificadores clásicos regresión logística y SVM y la técnica TF–IDF, resaltando sus ventajas en interpretabilidad y eficiencia para tareas de texto.

2.1 Arquitectura Transformer y self-attention

La arquitectura Transformer fue introducida por Vaswani et al. (2017) en el artículo "Attention Is All You Need". A diferencia de modelos recurrentes o convolucionales, el Transformer se basa exclusivamente en mecanismos de self-attention, que permiten ponderar la influencia de cada token de la secuencia sobre todos los demás de forma simultánea. Este enfoque bidireccional facilita la captura de dependencias a largo plazo y la desambiguación contextual de palabras polisémicas (Vaswani et al., 2017).

2.2 Modelo RoBERTa

RoBERTa (Robustly optimized BERT approach), propuesto por Liu et al. (2019), optimiza el pre-entrenamiento de BERT mediante tres cambios clave: mayor volumen de datos de preentrenamiento, eliminación de la tarea de predicción de siguiente oración y ajuste de hiperparámetros como el tamaño del batch y la longitud de secuencia. Estas modificaciones resultan en embeddings más ricos y una mejora consistente en benchmarks de comprensión y clasificación de texto (Liu et al., 2019).

2.3 Regresión logística

La regresión logística es un clasificador lineal que estima la probabilidad de cada clase mediante la función sigmoide aplicada a una combinación lineal de características. En tareas de análisis de sentimiento, suele emplearse sobre representaciones tradicionales como Bag-of-Words o TF–IDF (Chauhan et al., 2022). Su principal ventaja radica en la interpretabilidad de los coeficientes y la eficiencia computacional, aunque su

capacidad de captura semántica es limitada frente a modelos basados en embeddings profundos.

2.4 Máquinas de vectores de soporte (SVM)

Las Máquinas de vectores de soporte buscan que en un hiperplano se maximice el margen de separación entre clases en el espacio de características. Dado el conjunto de vectores de características (por ejemplo, TF–IDF o embeddings), el algoritmo selecciona los vectores soporte más cercanos al hiperplano para definir la frontera de decisión. Este método ofrece resistencia al sobreajuste y buen rendimiento en espacios de alta dimensión, aunque su escalabilidad puede verse afectada por grandes volúmenes de datos (Kumar et al., 2022).

3 Trabajos Relacionados

En esta sección se realiza una revisión de la literatura sobre los métodos empleados en análisis de emociones y sentimientos. Se cubren desde los primeros enfoques basados en diccionarios y reglas léxicas, pasando por los modelos estadísticos y de Machine Learning tradicional (Naive Bayes, SVM, Regresión Logística), hasta la revolución de los Transformers (BERT y RoBERTa).

3.1 Métodos basados en diccionarios y reglas léxicas

Los primeros intentos de análisis de emociones se fundamentaron en la creación de diccionarios de sentimientos y reglas basadas en palabras, asignando un valor positivo o negativo a cada término en función de su significado. Con este enfoque, el análisis consistía en contar las ocurrencias de palabras positivas y negativas en un texto. Estas técnicas fueron muy populares en el comercio electrónico entre 2000 y 2010, pues permitían realizar análisis de opinión de productos de forma rápida. Sin embargo, mostraban limitaciones para manejar sentimientos contradictorios dentro de un mismo texto y la ambigüedad lingüística inherente al lenguaje natural, sobre todo en casos con negaciones o intensificadores (Thakur et al., 2024).

3.2 Modelos estadísticos y algoritmos clásicos de aprendizaje automático

A partir de la década de 2010, surgieron métodos estadísticos y de aprendizaje automático tradicionales que incorporaron características como TF–IDF (frecuencia de término – frecuencia inversa de documento) para mejorar la precisión. Algoritmos como Naive Bayes, máquinas de vectores de soporte (SVM) y regresión logística comenzaron a dominar el campo, demostrando mejor desempeño que los métodos basados en diccionarios. Chauhan et al. (2022) compararon Naive Bayes, SVM y regresión logística en el IMDB Movie Reviews Dataset, hallando que SVM ofrece un equilibrio óptimo entre precisión y F1-score. De forma similar, Kavitha et al. (2022) aplicaron TF–IDF junto a Random Forest y regresión logística en datos de Twitter,

subrayando la importancia de un preprocesamiento cuidadoso para alcanzar métricas competitivas.

3.3 Advenimiento de los Transformers

El hito decisivo se produjo en 2017 con la propuesta de la arquitectura Transformer por Vaswani et al. (2017). Mediante su mecanismo de atención (self-attention), los Transformers procesan secuencias de forma paralela y modelan dependencias a largo alcance entre tokens, superando así las limitaciones inherentes a enfoques recurrentes o basados en reglas. Este avance transformó el campo del procesamiento del lenguaje, y en particular las tareas de análisis de sentimiento y detección de emociones, al mejorar la gestión de la ambigüedad y las contradicciones presentes en el texto.

3.4 Modelos preentrenados y data augmentation

Los Transformers preentrenados (BERT, RoBERTa, DistilBERT) han elevado el estándar en PLN. Comparaciones recientes muestran que RoBERTa suele alcanzar F1 superiores frente a BERT y a versiones ligeras (Sindhu et al., 2023; Thakur et al., 2024), con diferencias reportadas de hasta 0,05 puntos en escenarios de reseñas de productos. Asimismo, Wang et al. (2023) evidencian que técnicas de data augmentation como la sustitución de sinónimos y la traducción inversa aumentan la robustez de RoBERTa. En consecuencia, en este trabajo priorizamos una variante Transformer robusta y exploramos augmentación de datos para mejorar la generalización en español.

3.5 Enfoques híbridos

Para combinar las virtudes de los Transformers y los clasificadores convencionales, se emplean esquemas híbridos de integración. Tanwar et al. (2023) integraron BERT con análisis por aspectos, subrayando la necesidad de segmentar primero los aspectos del texto antes de la clasificación. Kumar et al. (2022) combinaron regresión logística y Random Forest sobre embeddings de BERT en reseñas de Amazon Alexa, alcanzando una precisión del 93.7 %. Srivastava et al. (2023) propusieron un enfoque híbrido que combina máquinas de vectores de soporte (SVM) y redes recurrentes (RNN); los autores reportaron una precisión del 93.6 % y mostraron que la meta-clasificación basada en probabilidades mejora la generalización respecto de los modelos. Singh y Srivastava (2023) evaluaron técnicas de ensamblado (stacking) de múltiples clasificadores, llegando a incrementar el AUC-ROC en 0.03 puntos respecto a modelos aislados.

3.6 Métricas de evaluación y desafíos en español

La elección de métricas adecuadas es crucial para valorar de forma completa el desempeño de los modelos. Maria et al. (2024) destacaron la importancia de utilizar Exact Match, F1-score y AUC-ROC en tareas multiclasificador y de preguntas y respuestas. No obstante, persisten retos en la clasificación de emociones en español: la ambigüedad semántica, la limitada disponibilidad de corpora anotados y la necesidad

de adaptar los modelos al contexto sociocultural. Singh y Srivastava (2023) indican que el fine-tuning de modelos preentrenados sobre corpora específicos en español puede mejorar de manera notable su rendimiento. Asimismo, la clasificación por aspectos (ABSA) presenta retos adicionales, como el manejo de opiniones contradictorias en un mismo texto (Tanwar et al., 2023; Wang et al., 2023).

4 Metodología de trabajo

En este capítulo se habla del desarrollo del modelo híbrido de clasificación de emociones en textos en español, se propone un enfoque en ocho fases que combina técnicas avanzadas de Procesamiento de Lenguaje Natural (PLN) y Aprendizaje Automático. La figura 1 presenta el diagrama de la metodología propuesta en este trabajo de investigación.

Figura 1. Metodología propuesta para este trabajo de investigación

4.1 Fase 1: Construcción del corpus

Se ha diseñado un formulario con 12 preguntas abiertas dos orientadas a evocar cada emoción básica (felicidad, tristeza, disgusto, ira, miedo y sorpresa) y se aplicó a la comunidad universitaria. El resultado fue un corpus original de 4200 textos etiquetados manualmente (350 estudiantes x 12 preguntas), que garantiza representatividad y contexto real para la tarea de clasificación emocional.

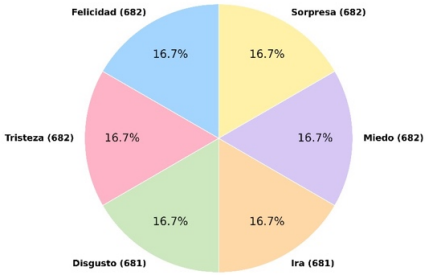


Figura 2. Distribución de respuestas por emoción.

4.2 Fase 2: Análisis Exploratorio de Datos (EDA)

Caracterizamos el corpus antes del preprocesamiento revisando distribución por emoción, posibles sesgos/outliers y longitud de las narrativas. Se muestra alta presencia de verbos en primera persona y conectores temporales (“cuando”, “sentí”, “después”), junto con vocabulario emocional (“miedo”, “triste”).

La figura 3 resume las estadísticas descriptivas sobre la longitud de las narrativas (en número de palabras). Se observa una longitud media de 25 palabras, una moda aproximadamente de 41 palabras, una desviación estándar de 26.1 palabras y una longitud mínima igual a 0 palabras. Estos valores indican que, si bien la longitud típica de respuesta es moderada, existe una amplia heterogeneidad, es decir, coexisten textos muy breves y textos considerablemente más largos.

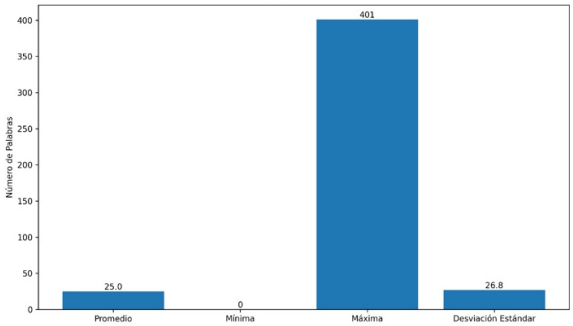


Figura 3. Caracterización de la longitud de los textos del corpus.

Así mismo, los bigramas más frecuentes, entre ellos: “hizo sentir”, “primera vez”, “sentí feliz/triste”, “tenía miedo”, “cómo reaccionar”, “seguir adelante”, confirman el carácter narrativo y afectivo de las respuestas. Esto guía el pipeline: lematización, revisión de stopwords y conservación de bigramas informativos.

La Figura 4 muestra la elevada frecuencia de conectores temporales y verbos en primera persona, lo cual indica que los textos tienen una estructura narrativa centrada en experiencias personales; esta característica favorece la extracción de características sintáctico-temporales para la clasificación emocional.

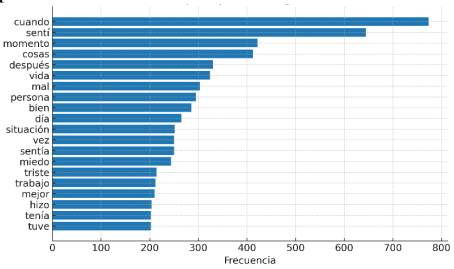


Figura 4. Top-20 conectores temporales y verbos en primera persona con carga afectiva.

La Figura 5 presenta los bigramas globales más frecuentes; expresiones como ‘hizo sentir’ y ‘tenía miedo’ funcionan como marcadores afectivos robustos que conviene preservar durante la tokenización para mejorar la señal discriminativa del clasificador.

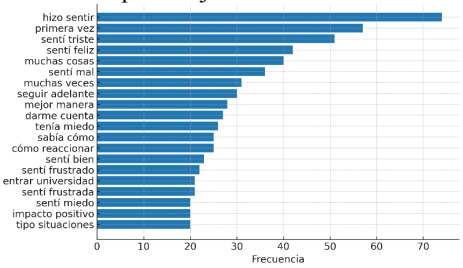


Figura 5. Top-20 bigramas. Co-ocurrencias narrativas/emocionales útiles para la clasificación.

4.3 Fase 3: Preprocesamiento de los textos

Normalización: eliminación de URLs, caracteres especiales, tildes y conversión a minúsculas.
Tokenización: uso de RobertaTokenizer.from_pretrained("roberta-base") max_length=128, truncamiento y padding dinámico.
Lematización: aplicación de SpaCy en español.
Eliminación de stopwords y marcaje de negaciones: descartando palabras vacías e indicando el alcance de negaciones, según prácticas recomendadas en PLN.

4.4 Fase 4: Extracción de características

Embeddings contextuales: vectores de 768 dimensiones extraídos del pooler de RoBERTa, aprovechando su capacidad para modelar dependencias a largo plazo (Liu et al., 2019).
TF–IDF: vectores dispersos de unigrama y bigrama para complementar la representación semántica clásica (Chauhan et al., 2022).

4.5 Fase 5: Fine-tuning de RoBERTa

Se entrena RobertaForSequenceClassification sobre el corpus preprocesado con parámetros: learning rate = 2×10^{-5}, batch size = 8, hasta 10 épocas y early stopping tras 2 épocas sin mejora en F1-score. El modelo obtuvo las probabilidades de cada emoción para cada texto.

4.6 Fase 6: Entrenamiento de clasificadores tradicionales

Se entrenaron dos modelos supervisados sobre los embeddings de RoBERTa:

1. Regresión logística, aprovechando su interpretabilidad y eficiencia en espacios de alta dimensión (Chauhan et al., 2022).
2. Support Vector Machines (SVM), explorando kernels lineal y RBF y ajustando el parámetro de regularización C por validación cruzada (Kumar et al., 2022).

4.7 Fase 7: Integración híbrida mediante meta-clasificación

Las probabilidades generadas por los tres modelos (RoBERTa, regresión logística y SVM) se concatenaron en un nuevo vector de características. Este vector alimenta un clasificador que aprende a combinar las fortalezas de cada enfoque, replicando estrategias de ensamblado que han demostrado mejoras en AUC-ROC y F1-score.

4.8 Fase 8: Evaluación del modelo

El desempeño del modelo híbrido se mide sobre el 20 % del corpus (840 textos; 140 por clase) , usando métricas multiclasificador estándar:

- Precisión (Accuracy)
- Recall macro
- F1-score macro
- AUC-ROC

Los resultados se compararon frente a los obtenidos por cada modelo individual, validando la superioridad del enfoque híbrido en la clasificación de emociones en textos en español.

5 Resultados

En esta sección se presentan los resultados cuantitativos obtenidos para el modelo híbrido y los clasificadores tradicionales,

5.1 Desempeño por clase del modelo híbrido

Se muestra el rendimiento detallado del modelo híbrido sobre las seis emociones analizadas: Disgusto (Precision 0.96, Recall 1.00, F1-score 0.98, Support 140), Felicidad (0.90, 0.76, 0.83, 140), Ira (0.73, 0.88, 0.80, 140), Miedo (1.00, 0.96, 0.98, 140), Sorpresa (0.96, 0.88, 0.92, 140) y Tristeza (0.92, 0.96, 0.94, 140). En la tabla 1 se presentan los resultados; precision mide la proporción de predicciones correctas por emoción, recall la capacidad de detección de instancias reales y F1-score la media armónica entre ambas. La fila de accuracy refleja un acierto global del 91 % sobre 840 muestras, support 840 presenta las medias simples de Precision, Recall y F1-score sin ponderar por frecuencia. El modelo híbrido alcanza así un 91 % de exactitud global, muy superior a los enfoques individuales. La emoción Miedo resulta la mejor detectada (F1-score 0.98), seguida de Disgusto y Sorpresa, mientras que Felicidad, con un Recall de 0.76, requeriría un mayor enriquecimiento de ejemplos para mejorar su captación.

Tabla 1. Resultados de la evaluación del modelo híbrido.

Clase	Precision	Recall	F1-score	Support
Disgusto	0.96	1.00	0.98	140
Felicidad	0.90	0.76	0.83	140
Ira	0.73	0.88	0.80	140
Miedo	1.00	0.96	0.98	140
Sorpresa	0.96	0.88	0.92	140
Tristeza	0.92	0.96	0.94	140
Accuracy			0.91	840

La matriz de confusión presentada en la figura 6, muestra que el modelo clasifica correctamente la mayoría de las instancias en cada categoría, lo que se refleja en los valores más altos concentrados en la diagonal principal. Por ejemplo, las clases Preocupado, Miedo y Tristeza presentan un buen nivel de acierto, con pocas confusiones hacia otras categorías. Sin embargo, también se observan errores relevantes: algunas muestras de Feliz y Enojado se confunden entre sí, lo cual indica que el modelo tiene dificultades para distinguir emociones con matices similares.

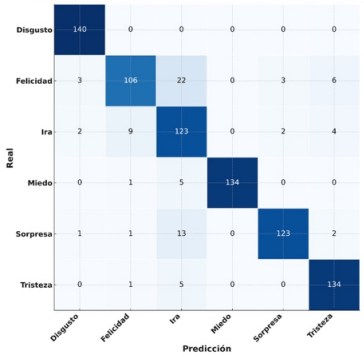


Figura 6. Matriz de confusión del modelo de clasificación de emociones.

5.2 Rendimiento de los clasificadores tradicionales

Los clasificadores tradicionales regresión logística y SVM se quedan en torno al 78% de accuracy, claramente por debajo del híbrido. Aunque el SVM obtiene un rendimiento ligeramente superior al de regresión logística, especialmente en la métrica AUC-ROC, ninguno de los dos iguala la robustez que aporta la etapa de meta-clasificación sobre las representaciones de RoBERTa como se muestra en la tabla 2.

Tabla 2. Resultados de la evaluación de los modelos tradicionales.

Modelo	Accuracy	Recall	F1-score	AUC-ROC
Regresión logística	0.778	0.765	0.770	0.821
SVM (lineal / RBF)	0.789	0.772	0.780	0.834

6 Conclusiones

En este estudio se ha propuesto un modelo híbrido que combina un Transformer RoBERTa fine-tuneado con clasificadores clásicos mediante meta-clasificación. El enfoque alcanza un 91 % de exactitud y un F1-score ponderado de 0.9127, superando ampliamente a la regresión logística (77.8 % / F1 0.770) y al SVM (78.9 % / F1 0.780), y el experimento de ablación confirma una ganancia de 3 pp frente al RoBERTa solo. Además, las emociones Miedo y Disgusto muestran la mayor precisión, mientras que Felicidad requiere un mayor número de muestras. Como líneas futuras se sugieren ampliar el corpus o aplicar data augmentation, explorar variantes ligeras de Transformers (DistilRoBERTa, TinyBERT) y mecanismos de atención por aspecto, y validar la generalización del método en dominios como redes sociales o reseñas de productos. En conjunto, estos resultados demuestran el valor de fusionar representaciones profundas y aprendizaje clásico para un análisis más preciso de emociones en texto. Finalmente, y con el objetivo de obtener estimaciones de desempeño más robustas y reducir la varianza de las métricas, la siguiente etapa de investigación consistirá en implementar validación cruzada estratificada (K-Fold), reportando medias y desviaciones estándar por clase. Esta práctica permitirá una evaluación más fiable del modelo y orientará decisiones de diseño y selección de hiperparámetros para mejorar su poder predictivo y su capacidad de generalización.

7 Referencias

Chauhan, H., Rathore, K. S., & Rajan, V. G. (2022). Sentiment analysis using supervised machine learning. In Proceedings of the 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 1276–1278).

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019) (pp. 4171–4186).

Jurafsky, D., & Martin, J. H. (2021). Speech and language processing (3.ª ed.; borrador). Recuperado de https://web.stanford.edu/~jurafsky/slp3/.

Kavitha, M., Naib, B. B., Mallikarjuna, B., Kavitha, R., & Srinivasan, R. (2022). Sentiment analysis using NLP and machine learning techniques on social media data. In Proceedings of the 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 112–114).

Kumar, S., Sharma, V., Kanshal, S., & Talwar, R. (2022). Sentiment analysis on Amazon Alexa reviews using NLP and classification. In Proceedings of the 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) (pp. 353–357).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach [Preprint]. arXiv. https://arxiv.org/abs/1907.11692.

Maria, M. N., Akter, S., Kabir, T., & Khan, R. (2024). Sentiment analysis using NLP libraries and machine learning. In Proceedings of the 2024 8th International Conference on ISMAC (IoT in Social, Mobile, Analytics and Cloud) (pp. 911–914).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

Rawat, A., & Samant, S. S. (2022). Comparative analysis of transformer-based models for question answering. In Proceedings of the 2022 2nd International Conference on Innovative Sustainable Computational Technologies (CISCT).

Sindhu, S., Kumar, S., & Noliya, A. (2023). A review on sentiment analysis using machine learning. In Proceedings of the International Conference on Innovative Data Communication Technologies and Application (ICIDCA 2023) (pp. 137–140).

Singh, H., & Srivastava, D. (2023). Sentiment analysis: Quantitative evaluation of machine learning algorithms. In Proceedings of the 5th International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 946–950).

Srivastava, A., Srivastava, V., Kumar, K., Srivastava, S., & Garg, N. (2023). Hybrid machine learning method for sentiment analysis. In Proceedings of the 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA 2023) (pp. 646–650).

Tanwar, G., Franklin, C., & Deepak, G. (2023). A case study on BERT models for aspect-based sentiment analysis. In Proceedings of the 2023 International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT).

Thakur, S., Mahadule, S., Chaple, S., Agrawal, P., Badhiye, S. S., & Rakesh, N. (2024). Comparative analysis of consumer sentiments using NLP and pre-trained transformer-based models. In Proceedings of the 2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS) (pp. 821–826).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (Vol. 30, pp. 5998–6008).

Wang, X., Xue, S., Liu, J., Zhang, J., Wang, J., & Zhou, J. (2023). Sentiment classification based on RoBERTa and data augmentation. In Proceedings of CCIS 2023.

Capítulo 5

Mejora de la detección de COVID-19 en tomografías computarizadas de tórax mediante el filtrado CLAHE

Victor Daniel Machorro García

Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla, México

victor.machorro@correo.buap.mx

Resumen Este estudio tiene como objetivo verificar cuan efectivo es realizar el preprocesamiento de ecualización adaptativa de histograma con limitación de contraste (CLAHE por sus siglas en idioma inglés) en tomografías computarizadas (TC) de tórax junto con el modelo DenseNet-201 para la detección de COVID-19. Los hallazgos muestran que CLAHE mejora las características de la imagen, lo que lleva a una mejora de las métricas de rendimiento, en particular en la recuperación y la puntuación F1. El modelo DenseNet201 que ha sido entrenado con imágenes preprocesadas con CLAHE logra una mayor recuperación y puntuación F1 en comparación con el modelo entrenado sin preprocesamiento. Esto destaca la importancia de las técnicas de mejora de imágenes para incrementar la precisión y fiabilidad de los modelos de aprendizaje profundo en el ámbito del análisis de imágenes médicas. Estos hallazgos sugieren que la combinación de estos dos enfoques puede proporcionar una herramienta de diagnóstico más fiable y precisa para los profesionales sanitarios en la gestión de COVID-19 (Adak et al., 2021; Hasan et al., 2021; Tjoa et al., 2022).

Palabras Clave: CLAHE, preprocesamiento de imagen, Deep Learning.

1 Introducción

La identificación temprana y precisa de las infecciones por COVID-19 mediante tomografía computarizada de tórax se ha convertido en una necesidad urgente ante la actual pandemia de SARS-CoV-2 (Dawod et al., 2021). Las técnicas de inteligencia artificial aplicadas a imágenes médicas, como lo son las radiografías de tórax y las tomografías computarizadas, han generado una gran atención para la detección automatizada de COVID-19, lo que ha dado lugar al desarrollo de numerosos sistemas (Ahmed et al., 2022). Las TC de tórax son particularmente valiosas para diagnosticar el COVID-19, y los investigadores de todo el mundo realizan esfuerzos para descubrir enfoques de tratamientos efectivos y rápidos para esta pandemia (Saad et al., 2021; Ukwuoma et al., 2023). Sin embargo, el rendimiento de los modelos de aprendizaje profundo en este contexto se encuentra intrínsecamente relacionado con calidad y las características de los datos de entrenamiento, así como la disponibilidad de conjuntos

de datos de alta calidad y correctamente anotados lo cual significo un reto, especialmente en las primeras etapas de una pandemia (Swayamsiddha et al., 2021). Las transformaciones de datos son cruciales para lograr una mejora en el rendimiento en los modelos de IA que utilizan imágenes y características clínicas, pero estas transformaciones a veces pueden llevar a resultados inesperados o inexactos (Giuste et al., 2022). Además, las variaciones en los protocolos de adquisición de imágenes, la demografía de los pacientes y la prevalencia de la enfermedad pueden introducir sesgos que impidan la capacidad de generalización de estos modelos (Giuste et al., 2022). El desarrollo de herramientas de diagnóstico por medio de la IA tiene el potencial de revolucionar las pruebas en el punto de atención (Hajij et al., 2021).

2 Trabajos relacionados

Los datos de imagenología médicas y los datos se han propuesto como suplementos potenciales de las pruebas moleculares de cribado de pacientes sospechosos de estar infectados por COVID-19 (Giuste et al., 2022). De hecho, se han desarrollado y evaluado modelos de IA para el diagnóstico de COVID-19 utilizando imágenes médicas clínicas y, en algunas ocasiones, generadas virtualmente (Tushar et al., 2023).

La IA ha mostrado su capacidad para mejorar la precisión y la eficacia de la interpretación de imágenes médicas como radiografías, resonancias magnéticas y tomografías computarizadas (Khalifa y Albadawy, 2024). Esto es particularmente útil para detectar patrones complejos que pueden ser complicados de distinguir a simple vista (Coelho, 2023). Al acelerar el procesamiento de imágenes médicas, los modelos de IA pueden detectar anomalías, células malignas y otras estructuras sospechosas, lo que permite que los pacientes sean priorizados (Wenderott et al., 2024). El papel de los métodos basados en IA tiene un rol cada vez más importante por su capacidad para analizar rápidamente imágenes de TC, disminuyendo el tiempo requerido para el diagnóstico en comparación con los enfoques tradicionales (Maghdid et al., 2021Esto es especialmente valioso en situaciones de emergencia y en entornos con recursos limitados, en los que una toma de decisiones rápida es esencial para una atención eficaz del paciente. Las técnicas de aprendizaje profundo se consideran herramientas fundamentales para que radiólogos y neumólogos distingan a los pacientes con COVID-19 de otros tipos de neumonía y de casos sanos (Beghoura et al., 2024).

3 Conjunto de datos

El conjunto de datos utilizado para realizar el entrenamiento consiste en tomografías computarizadas de tórax de pacientes sanos e infectados. Los datos proceden del *National Lung Screening Trial (NLST)* (National Lung Screening Trial Research Team, 2013) para los datos de pacientes sanos y de *The Cancer Imaging Archive (TCIA)* en

específico de la colección *CT Images in COVID-19* (An et al., 2020). para los datos de pacientes infectados.

El conjunto de datos CT *Images in COVID-19* consiste en conjuntos de datos de imágenes NIfTI de TC torácicas sin realce:

- Primer conjunto de datos: de 632 pacientes con infecciones por COVID-19 en el punto de atención inicial, y
- El segundo conjunto de datos comprende un segundo conjunto de 121 TC de 29 pacientes con infecciones por COVID-19, incluidas TC secuenciales seriadas.

De cada paciente se obtuvieron un aproximado de 70 imágenes del plano axial, aplicamos un ventaneo para la conversión de unidades Hounsfield(HU) a escala de grises en un rango de -1000 a +200 (HU) y se almacenaron en formato PNG, de esta forma obtuvimos 46,897 imágenes de pacientes infectados con neumonía causada por COVID-19.

Para el conjunto de individuos no infectados, se utilizó el conjunto de datos de tomografías computarizadas de tórax de acceso público del *National Lung Screening Trial (NLST)*. Este conjunto de datos es uno de los más grandes en su tipo con un total de 26,254 sujetos de estudio del cual seleccionamos sólo las TC de tórax del Estudio de cribado pulmonar de pacientes sanos, que tenía 17,000 pacientes y se seleccionaron aleatoriamente 50,000 imágenes. Estas imágenes ya se encuentran desde su origen en formato PNG y en el eje sagital.

En total se logra tener un total de 96,897 imágenes en el conjunto de datos (sanos e infectados), el cual se redimensionó a una resolución de 512×512 píxeles utilizando un algoritmo de interpolación lineal, esto debido a que no todas las tomografías eran del mismo tamaño. El total de instancias se dividió en un 70% para el entrenamiento, un 15% para la validación y un 15% para las pruebas.

4 Metodología

Para abordar estas limitaciones y mejorar la fiabilidad de la detección de COVID-19 basada en IA, este artículo explora la aplicación de la ecualización de histograma adaptativa con contraste limitado como paso previo al procesamiento para mejorar el contraste y la visibilidad de características de diagnóstico cruciales en imágenes de TC torácicas. La urgencia mundial en torno a COVID-19 impulsó a los investigadores a desarrollar rápidamente herramientas de IA para radiólogos; sin embargo, muchos estudios iniciales informaron de resultados poco realistas, casi perfectos, que disminuyeron significativamente en pruebas externas (Tushar et al., 2023). Este trabajo postula que la integración del filtrado CLAHE conducirá a una clasificación más robusta y precisa de las tomografías computarizadas de tórax como infectadas o no infectadas, incluso al tratar con las complejidades y variabilidades inherentes presentes en los datos clínicos del mundo real. El filtrado CLAHE tiene el potencial de aumentar las sutiles señales visuales que indican COVID-19 al mejorar la calidad de la imagen y reducir el ruido, lo que permite una extracción y clasificación más exactas por medio

de algoritmos de aprendizaje profundo. Es fundamental implementar un triaje clínico rápido en los centros de salud para clasificar a los pacientes en distintas categorías de urgencia, sobre todo si el acceso a las pruebas biológicas es restringido (Chassagnon et al., 2020). En primer término, aplicamos el filtro CLAHE a todas las imágenes del conjunto de datos para el posterior entrenamiento del modelo de clasificación. Posteriormente, evaluamos el modelo entrenando una red neuronal de aprendizaje profundo denominada Densenet201. De esta forma, podemos confirmar si el filtro CLAHE mejora el rendimiento de la red.

5 Filtro CLAHE

Con el objetivo de mejorar la calidad de imagen de los datos de entrada, se utilizó un filtro de ecualización adaptativa del histograma con limitación de contraste (CLAHE).

El filtro CLAHE se emplea resaltar las características diagnósticas sutiles en imágenes de TC torácica. CLAHE opera en pequeñas regiones locales (mosaicos) de la imagen, ajustando adaptativamente el contraste para mejorar el discernimiento de los detalles (Avcı & Karakaya, 2023). El proceso se compone de dos pasos clave:

1. Ecualización adaptativa del histograma:
    - Para cada mosaico, se calcula el histograma de intensidades de píxeles.
    - Las intensidades de píxeles se redistribuyen para aproximarse a una distribución uniforme, mejorando así el contraste local.
2. Limitación del contraste:
    - Para prevenir la amplificación excesiva del ruido, CLAHE limita el realce del contraste recortando el histograma a un límite de corte predefinido (L).
    - El límite de corte (L) especifica la altura máxima permitida de cualquier contenedor (*bin*), estos contenedores representan rangos de intensidad de píxeles, en el histograma. Los píxeles que superasen este límite se redistribuyen a otros contenedores, evitando la amplificación excesiva del ruido.

La mejora puede representarse matemáticamente como sigue:
Sea $I(x, y)$ la intensidad del píxel en las

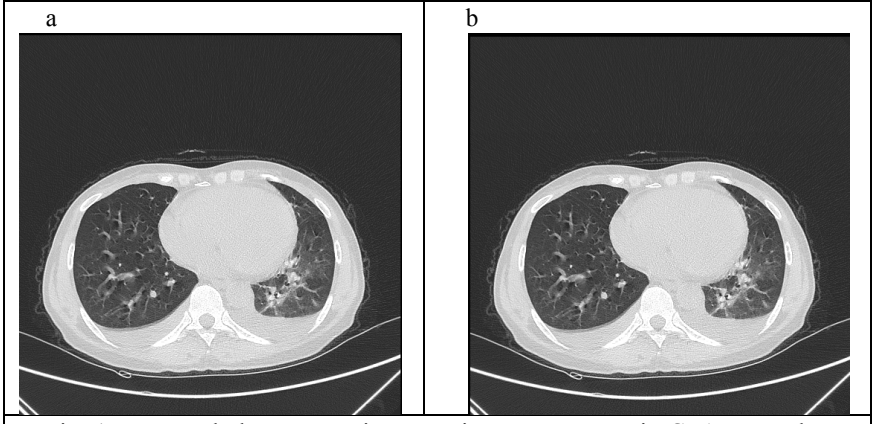
$$I'(x, y) = CLAHE(I(x, y), L, K)$$

Donde:

- *CLAHE* denota la operación CLAHE.

- *L* es el límite de corte.
- *K* es el tamaño del *kernel* que define la región local.

Se elige un límite de corte de 0,2 y un tamaño de *kernel* de 4x4 para optimizar la mejora del contraste y minimizar la amplificación del ruido, mejorando así la fiabilidad de los análisis de diagnóstico posteriores. En la Fig. 1 podemos observar la aplicación del filtro en una instancia del conjunto de datos.



Fig, 1 Muestra de la comparación entre imágenes con y sin CLAHE, podemos observar que las características de la imagen como texturas y bordes son más visibles en las imágenes procesadas por CLAHE imagen (b), a diferencia de la imagen (a) sin procesar. El valor BRISQUE de la imagen (a) fue de 45.98 mientras que el de la imagen (b) fue de 41.50 indicando una mejor calidad visual.

Con el fin de obtener una métrica que pudiese evaluar las imágenes se realizó el cálculo de BRISQUE (*Blind/Referenceless Image Spatial Quality Evaluator* por sus siglas en ingles) en algunas imágenes, este algoritmo analiza características estadísticas naturales en las imágenes sin distorsiones, como lo son patrones y texturas, cuando estas cambian es un indicativo de que la imagen esta dañada, borrosa o comprimida. BRISQUE mide cuando se desvían esas características de lo que se espera en una imagen natural y utiliza esos valores para dar un puntaje de calidad el cual va de 0 a 100 y mientras más bajo sea mejor será la calidad visual percibida. Al evaluar la métrica a todo el conjunto de datos se tuvo una mejora en promedio de 5.06 puntos.

6 DenseNet201

DenseNet201 es un tipo de red neuronal convolucional que se utiliza en el análisis de imágenes médicas, incluida la clasificación COVID-19 (Adak et al., 2021). La idea principal que subyace a DenseNet es que cada capa está conectada a todas las demás

de forma feed-forward (Hasan et al., 2021). Esto ayuda a aliviar el problema del gradiente de fuga, fomenta la reutilización de características y reduce el número de parámetros (Hasan et al., 2021).

En el contexto de la clasificación de COVID-19, DenseNet201 puede entrenarse para distinguir entre tomografías computarizadas de individuos sanos, pacientes con neumonía y aquellos con COVID-19 (Sanghvi et al., 2022). Al aprender de un gran conjunto de datos de imágenes etiquetadas, la red puede identificar patrones sutiles y características indicativas de la enfermedad (2020).

Modificamos el modelo base de Densenet201 para que se ajustara mejor al escenario de COVID-19, eliminamos la última capa de clasificación y la sustituimos por una personalizada con dos clases, infectados y no infectados

Se definió un total de 30 épocas para entrenar los dos modelos, uno utilizando las imágenes sin procesamiento y el otro utilizando las imágenes con el filtro CLAHE aplicado.

7 Resultados

Las métricas de evaluación incluyen la exactitud, la precisión, la recuperación y la puntuación F1, que se calculan a partir de la matriz de confusión. Esta matriz proporciona un desglose detallado de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos. Estas métricas proporcionan una evaluación exhaustiva de la capacidad del modelo de clasificación para distinguir entre casos infectados y no infectados. Las arquitecturas DenseNet han demostrado su eficacia en la identificación de COVID-19 mediante el análisis de imágenes de TC (Hasan et al., 2021).

Tras el entrenamiento, el modelo que utilizó el filtro CLAHE logró una mejora considerable en las métricas de detección de COVID-19, posiblemente porque el algoritmo CLAHE mejora la mayoría de las regiones, pero más notablemente en las métricas *Recall* y *F1-score*.

A continuación, se muestra una Tabla 1 con el informe de clasificación de la red entrenada con imágenes preprocesadas con CLAHE:

Class	**Precision**	**Recall**	**F1-Score**	**Support**
COVID	**0.86**	0.91	**0.88**	5008
NORMAL	0.96	0.94	0.95	12245
Accuracy			**0.93**	**17253**
Macro Avg	**0.91**	**0.92**	**0.91**	**17253**
Weighted Avg	**0.93**	**0.93**	**0.93**	**17253**

Tabla 1. informe de clasificación con imágenes procesadas con el algoritmo de CLAHE

A continuación, se muestra una table 2 con el informe de clasificación de la red entrenada sin imágenes preprocesadas:

Class	**Precision**	**Recall**	**F1-Score**	**Support**

COVID	0.83	0.90	0.86	5008
NORMAL	0.96	0.92	0.94	12245
Accuracy			**0.92**	**17253**
Macro Avg	**0.89**	**0.91**	**0.90**	**17253**
Weighted Avg	**0.92**	**0.92**	**0.92**	**17253**

Tabla 2. informe de clasificación con imágenes sin procesamiento.

Según las tablas de informes de clasificación, la mejora del rendimiento del modelo DenseNet201 utilizando imágenes preprocesadas con CLAHE es evidente en el aumento de los valores de las métricas de evaluación clave. Específicamente, el modelo entrenado con imágenes mejoradas con CLAHE muestra un mayor *recall* y *F1-score* para la detección de COVID-19.

- *Recall*: La métrica *recall*, que representa la capacidad del modelo para identificar correctamente todos los casos positivos reales (individuos infectados por COVID-19), mejoró de 0.90 a 0.91 al utilizar CLAHE. Esto indica que el modelo con preprocesamiento CLAHE es más efectivo para detectar un porcentaje más alto de casos reales de COVID-19.

- *F1-score*: La puntuación F1, que es la media armónica de la precisión y la recuperación, también aumentó de 0.86 a 0.88 con CLAHE. Esta métrica proporciona una medida equilibrada de la precisión del modelo, teniendo en cuenta tanto los falsos positivos como los falsos negativos.

- *Macro Avg:* macro promedio es un mecanismo de evaluación de métricas como recall, precisión o *F1-score*, calculando estas por clase y posteriormente promediando sin ponderar según el número de ejemplos en cada clase, los *Macro avg* expuestos en la Tabla 1 indican que se obtuvo un mayor rendimiento en cada una de las métricas evaluadas en comparación con los valores que se muestran en la Tabla 2.

- *Weighted Avg:* el promedio ponderado evalúa el desempeño del modelo de clasificación considerando la proporción de instancias de cada clase en el cálculo de las métricas como precisión, *recall* o F1-score, entre mas alto sea este valor podríamos decir que el modelo presenta un buen desempeño en las clases mayoritarias, en este caso también se obtiene una mayor puntuación en el modelo con las imágenes procesadas con el algoritmo CLAHE.

Estas mejoras sugieren que el preprocesamiento CLAHE mejora las características relevantes para la detección de COVID-19, facilitando al modelo DenseNet201 la distinción entre casos infectados y no infectados. El contraste mejorado y la visibilidad de detalles sutiles en las imágenes procesadas con CLAHE probablemente contribuyen a mejorar el rendimiento del modelo en términos de memoria y *F1-score*.

En la Fig. 2 podemos observar la matriz de confusión de los dos modelos, la matriz perteneciente al modelo DenseNet201 con imágenes preprocesadas es más precisa

porque demuestra un mejor rendimiento de clasificación a la hora de distinguir entre casos infectados y no infectados por COVID-19. He aquí por qué:

- Mayor tasa de verdaderos positivos: El modelo con preprocesamiento CLAHE tiene una mayor recuperación para la clase COVID (0.91 frente a 0.90). Esto indica que identifica correctamente una mayor proporción de casos COVID-19 reales. En la práctica, esto significa menos falsos negativos, lo que es fundamental en un entorno de diagnóstico.

- Rendimiento equilibrado (*F1-score*): es mayor para el modelo con CLAHE (0.88 frente a 0.86). Esto sugiere un mejor equilibrio entre identificar correctamente los casos positivos y evitar los falsos positivos.

- Precisión global: El modelo con CLAHE alcanza una mayor precisión global (93% frente a 92%). Esto significa que clasifica correctamente un mayor porcentaje de todos los casos, incluidos tanto los casos COVID-19 como los no COVID-19.

En resumen, una mejor matriz de confusión para el modelo DenseNet201 con preprocesamiento CLAHE indica que el modelo es más sensible a la hora de detectar casos de COVID-19, al tiempo que mantiene un buen equilibrio entre precisión y recuerdo. Esto conduce a una herramienta de diagnóstico más fiable y precisa.

(a) (b)

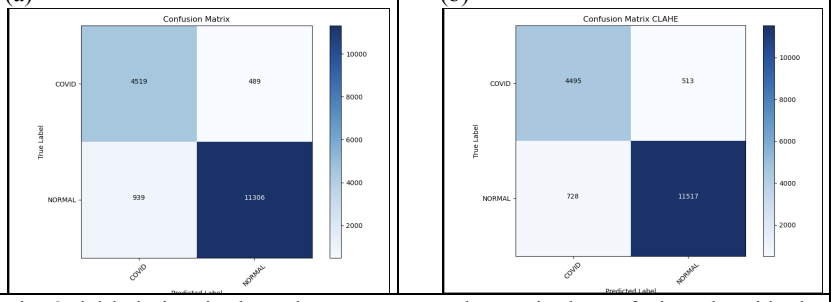


Fig. 2 del lado izquierdo podemos encontrar la matriz de confusión obtenida de la evaluación del modelo de red entrenado con las imágenes sin procesamiento, mientras que en la imagen (b) encontramos los resultados de la evaluación del modelo con las imágenes procesadas con el algoritmo CLAHE.

8 Conclusiones

En conclusión, nuestro estudio demuestra la eficacia de combinar el preprocesamiento CLAHE con un modelo DenseNet201 para la detección de COVID-19 en tomografías computarizadas. La aplicación de CLAHE mejora las características de la imagen, lo que conduce a una mejora de las métricas de rendimiento, en particular

en la recuperación y la puntuación F1. Esto sugiere que el contraste mejorado y la visibilidad de detalles sutiles en imágenes procesadas con CLAHE contribuyen a la capacidad del modelo para distinguir entre casos infectados y no infectados (Avcı & Karakaya, 2023).

Específicamente, nuestros resultados indican que el modelo DenseNet201 entrenado con imágenes preprocesadas con CLAHE alcanzan mayor recuperación (0,91) y puntuación F1 (0,88) en comparación con el modelo entrenado sin preprocesamiento. Esta mejora enfatiza la importancia de las técnicas de mejora de imágenes para aumentar la precisión y confiabilidad de los modelos de aprendizaje profundo para el análisis de imágenes médicas. Por medio del uso del modelo Densenet201 con imágenes mejoradas con el filtro CLAHE, somos capaces de proporcionar una herramienta de diagnóstico más confiable y precisa, con una precisión del 93%, que podría ayudar a los profesionales de la salud en una mejor gestión y clasificación de esta enfermedad.

Estos resultados concuerdan con los obtenidos en otros estudios que han evidenciado el potencial de los modelos de aprendizaje profundo, como DenseNet, para la detección de COVID-19 en imágenes médicas (Afifi et al., 2021; Hasan et al., 2021; Maharjan et al., 2021). Además, nuestros hallazgos enfatizan la importancia de las técnicas de preprocesamiento, como CLAHE, para mejorar el rendimiento de estos modelos. El uso de tomografías computarizadas de tórax puede detectar características anormales incluso antes de que aparezca la sintomatología clínica, lo que resulta útil para el diagnóstico temprano. La investigación futura podría centrarse en explorar diferentes técnicas de mejora de imágenes, así como, la combinación entre ellas y arquitecturas de aprendizaje profundo para incrementar aún más la precisión y la eficiencia de la detección de COVID-19.

Para futuros trabajos, se podría explorar varias alternativas para aprovechar los prometedores resultados obtenidos al combinar el preprocesamiento CLAHE con DenseNet201 para la detección de COVID-19 como, por ejemplo, investigar el potencial de otros métodos avanzados de mejora de imágenes, como *Pipeline for Advanced Contrast Enhancement* o la descomposición de modo empírico bidimensional, para mejorar aún más la calidad de las imágenes de TC y alcanzar un mayor el rendimiento de los modelos de aprendizaje profundo.

Agradecimientos

The Multi-national NIH Consortium for CT AI in COVID-19.

El autor agradece al Laboratorio Nacional de Supercómputo del Sureste de México (LNS), perteneciente al padrón de laboratorios nacionales SECIHTI, por los recursos computacionales, el apoyo y la asistencia técnica brindados, a través del proyecto No. 202101032C

Referencias

Aayush Jaiswal, Manipal University Jaipur, Jaipur, India, Neha Gianchandani, Manipal University Jaipur, Jaipur, India, Dilbag Singh, Manipal University Jaipur, Jaipur, India, Vijay Kumar, National Institute of Technology Hamirpur, Hamirpur, India, Manjit Kaur, Manipal University Jaipur, Jaipur, India, manjit.kaur@jaipur.manipal.edu. (2020). Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning. https://www.tandfonline.com/doi/full/10.1080/07391102.2020.1788642

Adak, C., Ghosh, D., Chowdhury, R. R., & Chattopadhyay, S. (2021). COVID-19–affected medical image analysis using DenserNet. In Elsevier eBooks (p. 213). Elsevier BV. https://doi.org/10.1016/b978-0-12-824536-1.00021-6

Afifi, A., Hafsa, N. E., Ali, M. A. S., Alhumam, A., & Alsalman, S. (2021). An Ensemble of Global and Local-Attention Based Convolutional Neural Networks for COVID-19 Diagnosis on Chest X-ray Images. Symmetry, 13(1), 113. https://doi.org/10.3390/sym13010113

Ahmed, S. A. A., Yavuz, M. C., Şen, M. U., Gülşen, F., Tutar, O., Korkmazer, B., Samancı, C., Şirolu, S., Hamid, R., Eryürekli, A. E., Mammadov, T., & Yanıkoğlu, B. (2022). Comparison and ensemble of 2D and 3D approaches for COVID-19 detection in CT images. Neurocomputing, 488, 457. https://doi.org/10.1016/j.neucom.2022.02.018

An, P., Xu, S., Harmon, S. A., Turkbey, E. B., Sanford, T. H., Amalou, A., Kassin, M., Varble, N., Blain, M., Anderson, V., Patella, F., Carrafiello, G., Turkbey, B. T., & Wood, B. J. (2020). CT Images in COVID-19 [Data set]. The Cancer Imaging Archive. https://doi.org/10.7937/TCIA.2020.GQRY-NC81

Avcı, H., & Karakaya, J. (2023). A Novel Medical Image Enhancement Algorithm for Breast Cancer Detection on Mammography Images Using Machine Learning. Diagnostics, 13(3), 348. https://doi.org/10.3390/diagnostics13030348

Beghoura, I., Benssalah, M., & Sbargoud, F. (2024). An Improved CovidConvLSTM model for pneumonia-COVID-19 detection and classification. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2408.11507

Chassagnon, G., Vakalopoulou, M., Battistella, E., Christodoulidis, S., Hoang-Thi, T.-N., Dangeard, S., Deutsch, É., André, F., Guillo, E., Halm, N., Hajj, S. E., Bompard, F., Neveu, S., Hani, C., Saab, I., Campredon, A., Koulakian, H., Bennani, S., Freche, G., … Paragios, N. (2020). AI-Driven CT-based quantification, staging and short-term outcome prediction of COVID-19 pneumonia. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2004.12852

Coelho, L. (2023). How Artificial Intelligence Is Shaping Medical Imaging Technology: A Survey of Innovations and Applications. Bioengineering, 10(12), 1435. https://doi.org/10.3390/bioengineering10121435

Dawod, E., Mahmoud, N., & El-Sisi, A. B. (2021). Hybrid approach for COVID-19 detection from chest radiography. IJCI International Journal of Computers and Information, 8(2), 71. https://doi.org/10.21608/ijci.2021.207754

Giuste, F., Shi, W., Zhu, Y., Naren, T., Isgut, M., Sha, Y., Li, T., Gupte, M., & Wang, M. D. (2022). Explainable Artificial Intelligence Methods in Combating Pandemics: A Systematic Review [Review of Explainable Artificial Intelligence Methods in Combating Pandemics: A Systematic Review]. IEEE Reviews in Biomedical Engineering, 16, 5. Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/rbme.2022.3185953

Hajij, M., Zamzmi, G., & Batayneh, F. (2021). TDA-Net: Fusion of Persistent Homology and Deep Learning Features for COVID-19 Detection in Chest X-Ray Images. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2101.08398

Hasan, N., Bao, Y., Shawon, A., & Huang, Y. (2021). DenseNet Convolutional Neural Networks Application for Predicting COVID-19 Using CT Image. SN Computer Science, 2(5). https://doi.org/10.1007/s42979-021-00782-7

Khalifa, M., & Albadawy, M. (2024). AI in diagnostic imaging: Revolutionising accuracy and efficiency. Computer Methods and Programs in Biomedicine Update, 5, 100146. https://doi.org/10.1016/j.cmpbup.2024.100146

Maghdid, H. S., Asaad, A., Ghafoor, K. Z., Sadiq, A. S., Mirjalili, S., & Khan, M. K. (2021). Diagnosing COVID-19 pneumonia from x-ray and CT images using deep learning and transfer learning algorithms. https://doi.org/10.1117/12.2588672

Maharjan, J., Calvert, J., Pellegrini, E., Green-Saxena, A., Hoffman, J., McCoy, A. J., Mao, Q., & Das, R. (2021). Application of deep learning to identify COVID-19 infection in posteroanterior chest X-rays. Clinical Imaging, 80, 268. https://doi.org/10.1016/j.clinimag.2021.07.004

National Lung Screening Trial Research Team. (2013). Data from the National Lung Screening Trial (NLST) [Data set]. The Cancer Imaging Archive. https://doi.org/10.7937/TCIA.HMQ8-J677

Saad, W., Shalaby, W. A., Shokair, M., El-Samie, F. A., Dessouky, M. I., & Abdellatef, E. (2021). COVID-19 classification using deep feature concatenation technique. Journal of Ambient Intelligence and Humanized Computing, 13(4), 2025. https://doi.org/10.1007/s12652-021-02967-7

Sanghvi, H. A., Patel, R., Agarwal, A., Gupta, S., Sawhney, V., & Pandya, A. S. (2022). A deep learning approach for classification of COVID and pneumonia using DenseNet-201. International Journal of Imaging Systems and Technology, 33(1), 18. https://doi.org/10.1002/ima.22812

Swayamsiddha, S., Kumar, P., Shaw, D., & Mohanty, C. (2021). The prospective of Artificial Intelligence in COVID-19 Pandemic [Review of The prospective of Artificial Intelligence in COVID-19 Pandemic]. Health and Technology, 11(6), 1311. Springer Science+Business Media. https://doi.org/10.1007/s12553-021-00601-2

Tjoa, E. A., Suparta, I. P. Y. N., Magdalena, R., & Kumalasari, N. (2022). The use of CLAHE for improving an accuracy of CNN architecture for detecting pneumonia. SHS Web of Conferences, 139, 3026. https://doi.org/10.1051/shsconf/202213903026

Tushar, F. I., Dahal, L., Sotoudeh-Paima, S., Abadi, E., Segars, W. P., Samei, E., & Lo, J. Y. (2023). Data diversity and virtual imaging in AI-based diagnosis: A case study based on COVID-19. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2308.09730

Ukwuoma, C. C., Cai, D., Heyat, M. B. B., Bamisile, O., Adun, H., Al-Huda, Z., & Al-antari, M. A. (2023). Deep learning framework for rapid and accurate respiratory COVID-19 prediction using chest X-ray images. Journal of King Saud University - Computer and Information Sciences, 35(7), 101596. https://doi.org/10.1016/j.jksuci.2023.101596

Wenderott, K., Krups, J., Zaruchas, F., & Weigl, M. (2024). Effects of artificial intelligence implementation on efficiency in medical imaging—a systematic literature review and meta-analysis [Review of Effects of artificial intelligence implementation on efficiency in medical imaging—a systematic literature review and meta-analysis]. Npj Digital Medicine, 7(1). Nature Portfolio. https://doi.org/10.1038/s41746-024-01248-9

Capítulo 6

Aprendizaje supervisado aplicado a señales de electrooculografía

Ana P. Rojas Martínez, Maya Carrillo Ruiz, Juan F. Leyva Bonilla, Rogelio González Velázquez

Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla, Av. San Claudio, Boulevard 14 Sur, Ciudad Universitaria, C.P. 72592. Heroica Puebla de Zaragoza, Puebla, México.
rm223470501@alm.buap.mx, maya.carrilloruiz@viep.com.mx, juan.leyvab@correo.buap.mx,rogelio.gonzalez@correo.buap.mx

Resumen. La electrooculografía es una prueba que recoge señales producidas por la diferencia de potencial corneo-retina cuando ocurre un cambio de estado del reposo ocular. Con el objetivo de implementar un prototipo domótico controlado por señales de electrooculografía, durante este trabajo nos enfocamos en la identificación de señales oculares. Para este propósito las señales se procesan mediante un análisis tiempo-frecuencia. Posteriormente, las representaciones de las señales se clasifican empleando redes neuronales bilaterales, trilaterales y la máquina de soporte vectorial con diferentes funciones kernel. Los mejores resultados se obtienen con un análisis en tiempo de las señales y la red neuronal bilateral con un 97.8% de exactitud en la clasificación de señales. El siguiente paso dentro de esta investigación será la construcción de un prototipo domótico para que usuarios, con habilidades físicas y del habla limitadas, puedan tener control de su entorno dentro de una habitación.

Palabras Clave: electrooculografía, bioseñal, domótica, clasificación.

1 Introducción

La automatización del hogar es un tema de interés desde el surgimiento de asistentes electrónicos y la posibilidad de hacer una conexión global de nuestros dispositivos para su control y gestión. Si bien los dispositivos se pueden controlar mediante la voz o movimiento, algunos estudios muestran el desarrollo de aplicaciones que permiten controlar los dispositivos eléctricos mediante señales biomédicas. Uno de los objetivos de explorar el uso de señales biomédicas con estos fines es proporcionar a personas, con habilidades físicas y del habla limitadas, una forma diferente de controlar sus dispositivos eléctricos dentro de su entorno. Según la Organización Mundial de la Salud (OMS) se calcula que 1300 millones de personas (1 de cada 6 personas) sufren una discapacidad importante (OMS, 2023) y en México, según datos de la Encuesta Nacional de la Dinámica Demográfica (ENADID) del 2014, 7.1 millones de habitantes del país no pueden o tienen mucha dificultad para hacer actividades como: caminar,

subir o bajar usando sus piernas; ver (aunque use lentes); mover o usar sus brazos o manos; aprender, recordar o concentrarse; escuchar (aunque use aparato auditivo); bañarse, vestirse o comer; hablar o comunicarse; y problemas emocionales o mentales (INEGI, 2017). Por ello, este trabajo busca emplear la señal biológica que surge del movimiento de los ojos conocida como señal de electrooculografía para controlar un prototipo domótico. Para lograrlo se realizó un análisis para extraer información relevante de las señales en frecuencia y tiempo, acto seguido se empleó el aprendizaje automático. En concordancia con la bibliografía se consideraron los clasificadores: redes neuronales (angosta, mediana, ancha, bilateral y trilateral) y la máquina de soporte vectorial con variaciones de la función kernel.

Este artículo está organizado de la siguiente manera. La sección dos presenta el marco teórico. La sección tres el trabajo relacionado. La sección cuatro la arquitectura de un prototipo domótico controlado por señales de electrooculografía. Finalmente, las secciones cinco y seis la experimentación y conclusiones.

2 Marco teórico

Una señal es una perturbación física que transmite información, tiene diferentes características de interés para este trabajo las cuales son: valor máximo de amplitud de pico (VMAP), posición de valor máximo de amplitud de pico (PVMAP), valor máximo de amplitud de valle (VMAV), posición de valor máximo de amplitud de valle (PVMAV), cruces por umbral (CU) y área bajo la curva (A). Estás características se observan en Figura 1.

Otro concepto importante en la Figura 1 es el umbral, definido como un valor establecido a partir del cual se produce un efecto, útil para la caracterización de las señales y la posterior identificación en el tipo de movimiento ocular.

Definido el concepto de señal, llegamos a la definición de señal de electrooculografía cuyo origen es el sistema visual, sistema que presenta perturbaciones generadas por movimientos voluntarios e involuntarios, siendo los voluntarios aquellos de interés en este documento: movimientos sacádicos y parpadeo. Los movimientos sacádicos son desplazamientos angulares rápidos y precisos producidos usualmente para observar con detalle objetos y no requieren de un estímulo inicial para generarse, en ellos el cerebro

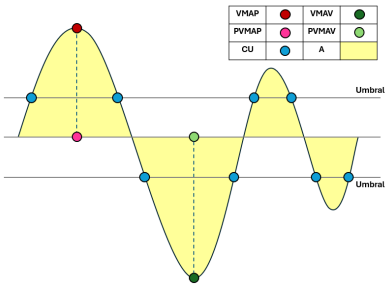


Figura 1. Características de una señal.

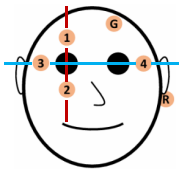


Figura 2. Disposición de electrodos para captar señales EOG (Nathaniel et al., 2018).

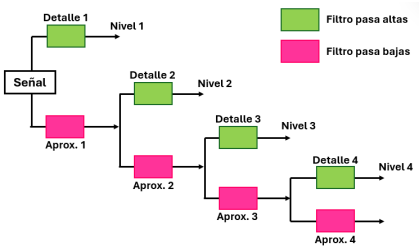


Figura 3. Filtros de la transformada

calcula la posición del ojo en la órbita y la del objetivo, con la diferencia de estas medidas se generan órdenes motoras (Lozano, 2008). El parpadeo es un movimiento realizado por los párpados que se abren y cierran con diferentes fines, como la protección, lubricación y comunicación. Derivado de la actividad cerebral producida por los movimientos oculares se crean señales de electrooculografía (EOG), generadas ya que el ojo humano se comporta como un dipolo eléctrico con la córnea (polo positivo) y la retina (polo negativo), la diferencia de voltaje entre la región frontal y trasera del ojo es llamado potencial corneo-retina y tiene una magnitud variable de 0.4 – 1.0 mV. Los ejes de este dipolo están orientados con el eje ocular, por ello, rotan junto con él (Nathaniel et al., 2018). Las señales EOG se obtienen de electrodos colocados de forma no invasiva en el rostro alrededor de un globo ocular, generalmente el derecho, con una configuración estándar como se muestra en la Figura 2 los movimientos son registrados con 6 electrodos. Los electrodos 1 y 2 colocados en la parte superior e inferior del ojo, los electrodos 3 y 4 adyacentes en las sienes laterales, el electrodo ('G') correspondiente a tierra y el electrodo ('R') como electrodo de referencia.

Las señales EOG usadas para este trabajo son señales discretas cuya información es la diferencia de potencial corneo-retina y su variación en función del tiempo. Los movimientos de interés son el movimiento sacádico hacia arriba, abajo, derecha e izquierda y el parpadeo. Otro concepto importante es la operación de ventaneo, donde la señal $x[n]$ se multiplica por una secuencia finita $w[n]$ llamada ventana (Proakis, J. G y Manolakis, D. G., 2007). produciendo una secuencia $\hat{x}[n] = w[n]x[n]$ que consiste en tomar L muestras de $x[n]$.

Una técnica para el análisis de señales es la transformada wavelet discreta donde se realizan operaciones de estiramiento y desplazamiento. Su expresión matemática se encuentra en la ecuación 1, donde: $x(t)$ es una señal discreta, j y k números enteros

relacionados con los factores de desplazamiento a y escalamiento b en la wavelet continua de: $a = 2^{-j}$ y $b = k2^{-j}$.

$$DWT_{j,k} = 2^{j/2} \int_{-\infty}^{\infty} x(t)\psi(2^j t - k)\, dt \quad (1)$$

La transformada wavelet discreta de una señal se calcula enviando la señal a través de filtros pasa altas y pasa bajas obteniendo coeficientes de salida de aproximación $c_{j,k}$ y detalle $d_{j,k}$ cómo se observa en la Figura 3. Con los atributos de una señal EOG, puede identificarse el tipo de movimiento ocular. Para hacer una distinción de la señal EOG se utilizó un proceso de clasificación, con el objetivo de predecir la etiqueta de clase de una nueva instancia, basado en observaciones pasadas (Raschka, S. y Mirjalili, V., 2019). Los clasificadores empleados son las redes neuronales (NN – Neural Network por sus siglas en inglés) que es un modelo con una estructura de capas interconectada, cada una conteniendo un número de neuronas, su estructura varía según el tipo de red del que se hable y la máquina de soporte vectorial (SVM – Support Vector Machine por sus siglas en inglés) que es un algoritmo que clasifica datos encontrando el mejor hiperplano que separa puntos de información de una clase de los puntos de otra clase. La elección de estos clasificadores está motivada por la revisión del trabajo relacionado.

3 Trabajo relacionado

Esta sección presenta una breve descripción de aquellos trabajos consultados que comparten afinidad en el análisis de señales biomédicas y algunas aplicaciones.

Aungsakul et al. realizaron una evaluación de características en tiempo de señales EOG para conocer cuál parámetro es más significativo para diferenciar movimientos oculares, analizaron parámetros como el valor máximo de amplitud de pico y posición, valor máximo de amplitud de valle y posición, valor del área bajo la curva y número de cruces por umbral. La evaluación se basó en el parámetro F estadístico, siendo el valor más alto el área bajo la curva (*F=594.76*) (Aungsakul et al., 2012).

Bulling, A., et al. analizaron señales EOG para reconocer movimientos oculares. Con la transformada wavelet continua en movimientos sacádicos, tomaron parámetros como la media, varianza, máxima amplitud de la señal, etc. Para la clasificación emplearon SVM con una precisión de 93% y recuerdo de 81.9% (Bulling, A. et al. 2011).

Abbaspour, S., et al. analizan algoritmos para el reconocimiento de movimientos musculares en manos. Los parámetros empleados son en tiempo (valor absoluto medio, desviación estándar, etc.) y frecuencia (longitud de onda, frecuencia media, etc.) con clasificadores como SVM, perceptrón multicapa (Multilayer Perceptron – MLP), análisis discriminatorio lineal (Linear Discriminant Analysis – LDA), K – vecinos más cercanos (K-nearest neighbors – KNN) y árbol de decisión (Decision Tree). La exactitud más alta se obtuvo con el MLP, arriba del 98% (Abbaspour, S., et al. 2020).

Tuay, D., y Quijada, G. realizaron transmisión inalámbrica de datos EOG con un análisis de tiempo – frecuencia y una descomposición wavelet de tres niveles sin familia definida y SVM (Tuay, D., y Quijada, G. (2019).

Mulam, H. y Mudigonda, M. propusieron un método para reconocer señales EOG. Con un método de reducción de señales, descomposición wavelet multinivel y NN. Identificaron diferentes movimientos oculares con una exactitud de entre 76% y 97% y una precisión de entre 80% y 88% (Mulam, H. y Mudigonda, M., 2017).

4 Arquitectura

Esta sección describe la arquitectura del prototipo domótico propuesto ver Figura 4. En las siguientes secciones se explica cada uno de sus componentes.

4.1 Etiquetar señales.

Este módulo recibe las señales EOG discretizadas, para su etiquetamiento; operación realizada bajo dos criterios: ángulos y distancia entre el punto central y un punto al azar. En la Figura 5 se observan dos áreas, los movimientos que caen sobre el área gris son descartados para delimitar los movimientos hacia la derecha, izquierda, arriba y abajo; en el área blanca, los movimientos son seleccionados representando los movimientos del conjunto de datos. El primer criterio consiste en medir un ángulo de 40° como se observa en la Figura 5; el segundo criterio consiste en medir una distancia mínima que deben recorrer los ojos durante su movimiento para considerarla relevante, en el eje horizontal de 0.15 m y el eje vertical de 0.1m. Aplicando los criterios descritos, las

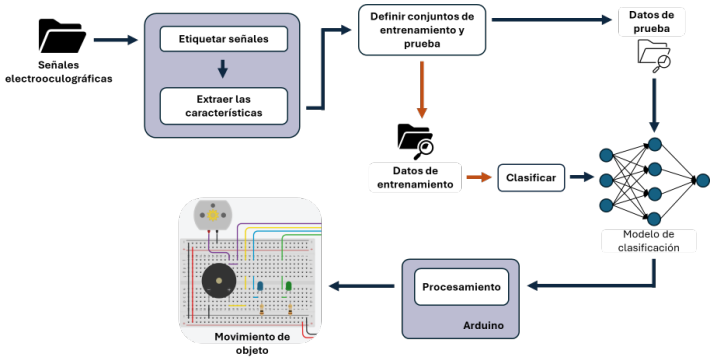


Figura 4. Arquitectura del prototipo domótico.

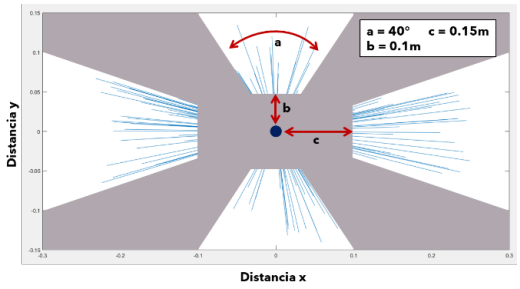


Figura 5. Modelo trigonométrico de los ángulos objetivo sobre la pantalla.

Tabla 1. Características extraídas de las señales EOG.

Característica	Fórmula	Núm. de ecuación
Valor absoluto medio	$MAV = \bar{x} = \frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert$	(2)
Desviación estándar	$STD = \left[\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2\right]$	(3)
Varianza	$Var = \frac{1}{N-1}\sum_{i=1}^{N} x_i^2$	(4)
Longitud de forma de onda	$WL = \sum_{i=1}^{N}\lvert x_i - x_{i-1}\rvert$	(5)
Cruce por cero	$Var = \sum_{i=1}^{N} sgn(-x_i x_{i-1})$ $sgn(x) = \begin{cases}1; & si\ x > 0 \\ 0; & de\ otra\ manera\end{cases}$	(6)
Energía	$E = \left[\sum_{i=1}^{N} x_i\right]^2$	(7)

Tabla 2. Esquema de tiempo de un ensayo. (Bárbara, B., et al., 2020).

Segundo	Acción
1	Movimiento sacádico con origen en el centro de la pantalla y final en un punto aleatorio sobre la pantalla.
2	Movimiento sacádico con origen sobre el punto aleatorio en la pantalla del segundo anterior y final en el punto central de la pantalla.
3-4	Parpadeo al cambio de color del punto.

señales EOG discretas que entran a este módulo se etiquetan según su movimiento sacádico arriba, abajo, derecha e izquierda.

4.2 Extracción de características.

Etiquetas las señales EOG procedemos con la identificación de atributos, las características empleadas son calculadas en función del tiempo y frecuencia, listadas en la Tabla 1. Observando que $x_i = señal\ EOG$ y $N = tamaño\ de\ la\ ventana$ (Abbaspout, S., et al., 2020) (Aungsakul, S. et al., 2012).Con las señales representadas se definieron los conjuntos de entrenamiento y pruebas para el proceso de clasificación.

4.3 Clasificación

Durante este módulo se entrenaron los clasificadores: red neuronal multicapa (NN) y la máquina de soporte vectorial (SVM) empleando una validación cruzada de 5 pliegues. Para las NN se entrenaron redes bicapa, tricapa, angosta, media y ancha. Para la máquina de soporte vectorial se realizaron pruebas modificando la función kernel: lineal, gaussiana, cuadrática y cúbica. En este módulo con las señales de entrada del conjunto de entrenamiento y los diferentes algoritmos de aprendizaje automático generan modelos, éstos son utilizados para predecir el tipo de la señal de entrada tomada del conjunto de pruebas, recuérdese que puede ser un movimiento sacádico hacia arriba,

movimiento sacádico hacia abajo, movimiento sacádico a la derecha, movimiento sacádico a la izquierda o un parpadeo.

4.4 Interfaz de control: procesamiento Arduino.

En este módulo se recibe el tipo de señal al que pertenece el movimiento ocular que será procesada por Matlab para Arduino. La tarjeta envía las señales a el circuito para controlar componentes eléctricos como un ventilador, luces, ventanas automatizadas y un sistema de alarma, cada uno de los cuales se activará según la señal recibida.

5 Experimentación

A continuación, se describen los experimentos realizados para la selección del clasificador. Los datos utilizados en este proyecto son de L- Università tà Malta registrados por los autores Barbara Nathaniel, A. Camilleri Tracey y P. Camilleri Kenneth, las señales fueron recabadas usando el amplificador de bioseñales g.tec g.USBamp con una frecuencia de muestreo de 256 Hz y filtradas con un filtro pasa bandas de entre 0-30 Hz y un filtro notch de 50 Hz (Bárbara, B., et al., 2020). Los datos resultaron de mediciones tomadas a seis personas. Para la toma de señales, los usuarios fueron sentados a 60 cm (D) de un monitor LCD de 24 in con su cabeza inmovilizada con un sistema similar a un oftalmoscopio. El esquema de tiempo de los experimentos se muestra en la Tabla 2. Con una configuración estándar de los electrodos los autores obtuvieron la diferencia de potencial EOG entre la alineación horizontal y vertical para producir los componentes EOG horizontales y verticales, ecuaciones (8) y (9):

$$EOG_h(t) = V_3(t) - V_4(t) \tag{8}$$

$$EOG_v(8) = V_1(t) - V_2(t) \tag{9}$$

Los términos $V_1(t)$, $V_2(t)$, $V_3(t)$ y $V_4(t)$ denotan los potenciales EOG medidos por los electrodos 1, 2, 3 y 4 de la Figura 2. Las señales de esta base de datos fueron seleccionadas, etiquetadas y caracterizadas como fue explicado en la arquitectura del sistema. Se definieron los conjuntos de entrenamiento y pruebas con un total de 365 y 90 señales respectivamente para realizar los siguientes experimentos:

5.1 Experimento 1

En este experimento la señal se caracterizó con cinco métricas en función del tiempo: valor absoluto medio, desviación estándar, varianza, longitud de forma de onda y cruce por cero definidas en la sección 4.2. Para calcularlas se realizó un ventaneo con una variación del tamaño de muestra L (ventana) de 512, 256, 128 y 64. Posteriormente se realizaron experimentos con NN y SVM. Los mejores resultados están presentes en la Tabla 3 y 4 resaltados en negritas y corresponden a las ventanas 128 y 64, los resultados de las ventanas 512 y 256 no fueron incluidos al tener menores porcentajes de exactitud.

Tabla 3. Resultados de clasificación con características de tiempo empleando NN.

Ventana	Exactitud									
	NN Angosta		NN Media		NN Ancha		NN Bilateral		NN Trilateral	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
128	96.4	96.7	98.1	96.7	96.7	96.7	97.0	**97.8**	95.6	96.7
64	96.4	**97.8**	97.5	95.6	97.3	96.7	90.4	96.7	92.9	94.4

Tabla 4. Resultados de clasificación con características de tiempo empleando SVM con diferente kernel.

Ventana	Exactitud					
	SVM Lineal		SVM Cuadrático		SVM Cúbico	
	Train	Test	Train	Test	Train	Test
128	96.4	**97.8**	96.7	96.7	96.4	95.6
64	95.3	93.3	96.2	92.2	94	91.1
Ventana	Exactitud					
	SVM Gaussiano Fino		SVM Gaussiano Medio		SVM Gaussiano Grueso	
	Train	Test	Train	Test	Train	Test
128	77.5	75.6	94.0	**97.8**	91.8	94.4
64	61.1	55.6	97.5	95.6	87.1	91.1

Como se observa en las Tablas 3 y 4, el mejor resultado es de 97.8% de exactitud obtenida con: NN bilateral y NN angosta con ventana de 128 y 64 cada una, y SVM con un kernel lineal y gaussiano con ventana de128 para ambas.

5.2 Experimento 2

En este experimento la señal se caracterizó con siete métricas en función del tiempo: valor máximo de amplitud de pico y valle, posición máxima de amplitud de pico y valle, valor de área bajo la curva, número de cruces por umbral y varianza, definidas en la sección 4.2. Para su cálculo se realizó la operación de ventaneo con variación del tamaño de muestra L (Ventana) de 512, 256, 128 y 64 y de umbral (Umbral) de 285, 255 y 225; valores seleccionados experimentalmente.

Posteriormente se realizaron experimentos con NN y SVM. Los mejores resultados se observan en la Tabla 5 y 6 resaltados en negritas y corresponden a las ventanas 256 y 128, los resultados de las ventanas 512 y 64 no se incluyeron al tener menores porcentajes de exactitud, además se realizaron experimentos con SVM y kernel fino, medio y grueso gaussiano, dichos resultados no fueron incluidos al no superar los porcentajes de la Tabla 6.

Tabla 5. Resultados de clasificación con características de tiempo empleando NN.

Ventana	Umbral	Exactitud									
		NN Angosta		NN Media		NN Ancha		NN Bilateral		NN Trilateral	
		Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
256	285	87.4	88.9	85.2	88.9	88.2	90	83.8	88.9	83.3	90
	255	87.9	86.7	86.6	90	88.8	90	80.5	84.4	83.3	86.7
	225	87.7	86.7	87.7	86.7	89.6	87.8	86.3	85.6	85.5	83.3
128	285	91.5	92.2	94.2	93.3	94	93.3	91	91.1	85.5	88.9
	255	91	91.1	94.5	91.1	92.9	93.3	88.5	88.9	88.8	87.8
	225	90.7	88.9	92.3	92.2	91.5	**94.4**	85.8	85.6	85.5	91.1

Tabla 6. Resultados de clasificación con características de tiempo empleando SVM con diferente kernel.

Ventana	Umbral	Exactitud					
		SVM Lineal		SVM Cuadrático		SVM Cúbico	
		Train	Test	Train	Test	Train	Test
256	285	87.4	91.1	86.3	91.1	83.6	88.9
	255	87.1	88.9	88.2	91.1	84.7	87.8
	225	87.1	88.9	85.8	91.1	84.1	87.8
128	285	91.8	92.2	93.4	92.2	89.6	93.3
	255	91.8	90	92.3	92.2	90.4	92.2
	225	91.2	91.1	90.1	**93.3**	90.1	**93.3**

Como se observa en las Tablas 5 y 6, el mejor resultado se obtuvo con la NN ancha con ventana de 128 y umbral de 225 con una exactitud de 94.4%.

5.3 Experimento 3

En este experimento la señal se caracterizó con seis métricas en función del tiempo – frecuencia: valor absoluto medio, desviación estándar, varianza, longitud de forma de onda y cruce por cero definidas en la sección 4.2. Para calcular las métricas se aplicó la transformada wavelet discreta sobre las señales EOG discretas y se tomaron los coeficientes de aproximación sobre los que se realizó la operación de ventaneo con variación del tamaño de muestra L (Ventana) de 64, 32, 16 y 8. Las operaciones se realizaron con las familias daubechies (db2) y symlet (sym2). Posteriormente se realizaron experimentos con NN y SVM. Los mejores resultados están en las Tablas 7 y 8 resaltados en negritas y corresponden a las ventanas 16 y 8, los resultados de las ventanas 64 y 32 no fueron incluidos al tener menores porcentajes de exactitud.

Tabla 7. Resultados de clasificación con características de tiempo – frecuencia empleando NN.

Ventana	Familia	Exactitud									
		NN Angosta		NN Media		NN Ancha		NN Bilateral		NN Trilateral	
		Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
16	db2	94.5	**96,7**	94.8	95.6	94.8	95,6	92.1	94.4	93.2	94.4
8		94.5	93,3	95.3	91.1	96.2	93,3	94.5	90	93.2	90
16	sym2	94.8	92,2	94	**96.7**	94.8	94,4	89.6	94.4	91.8	91.1
8		94.2	93,3	97	94.4	95.9	93,3	93.7	94.4	91.2	91.1

Tabla 8. Resultados de clasificación con características de tiempo – frecuencia empleando SVM con diferente kernel.

Ventana	Familia	Exactitud					
		SVM Lineal		SVM Cuadrática		SVM Cúbica	
		Train	Test	Train	Test	Train	Test
16	db2	94.2	**96.7**	95.6	**96.7**	93.4	95.6
8		94.8	**96.7**	93.2	94.4	94.0	**96.7**
16	sym2	79.2	85.6	94.2	**96.7**	80.5	87.8
8		95.1	**96.7**	93.4	94.4	94.2	**96.7**
Ventana	Familia	Exactitud					
		SVM Gaussiano Fino		SVM Gaussiano Medio		SVM Gaussiano Grueso	
		Train	Test	Train	Test	Train	Test
16	db2	79.2	85.6	94.2	**96.7**	80.5	87.8
8		78.1	82.2	91.5	93.3	75.3	82.2
16	sym2	81.6	85.6	94.0	**96.7**	82.2	87.8
8		78.1	82.2	93.2	93.3	75.9	82.2

Las Tablas 7 y 8 muestran los mejores resultados en la familia db2 con ventana de 16 con exactitud de 96.7% son con la NN angosta, NN media, y SVM con kernel lineal, cuadrático y gaussiano medio; con ventana de 8 se obtuvo la misma exactitud con SVM lineal y cúbico. En cuanto a la familia sym2 se mantuvo la exactitud con una ventana de 16 para la NN media, SVM cuadrático y SVM gaussiano medio, con una ventana de 8 se alcanzó el mismo resultado con el SVM lineal. Finalmente, los mejores resultados se encuentran en el experimento 1. Con ventana de 128 y los clasificadores NN bilateral, SVM lineal y SVM gaussiano medio con una exactitud de 97.8%. La NN angosta iguala esta exactitud con una ventana de 64. Con un análisis general de los tres experimentos observamos que las características con mejores resultados son calculadas en función del tiempo en el experimento 1, si tomamos en cuenta los experimentos 1 y 2 (ambos realizados en función del tiempo)

Tabla 9. Comparación de clasificadores con mejores resultados de exactitud.

Clasificador	Tiempo de entrenamiento (seg)	Velocidad de predicción (obs/seg)	Tamaño del modelo compacto (kB)
NN Bilateral	1.4202	12000	15
SVM Lineal	4.3384	3500	102
SVM gaussiano medio	3.8164	3300	305

observamos que la ventana de 128 es la que tiene porcentajes de exactitud más altos, por ello dicha ventana es seleccionada para emplear junto al clasificador.

Para seleccionar entre la NN bilateral, SVM lineal y SVM gaussiano medio analizaremos; tiempo de entrenamiento, velocidad de predicción y tamaño del modelo, datos de la Tabla 9. La velocidad de predicción se refiere al número de observaciones procesadas por segundo, el clasificador con mayor velocidad es la NN bilateral. Otro factor importante es el tamaño del modelo ya que este proyecto tiene el objetivo de la construcción de un prototipo domótico, la cantidad de memoria que ocupe el modelo definirá las especificaciones del dispositivo computacional necesario para su funcionamiento, ante estas consideraciones, se seleccionó el clasificador de redes neuronal bilateral (NN bilateral) para la construcción del prototipo domótico.

6 Conclusiones y trabajo futuro

El análisis de señales EOG discretas en función del tiempo fue planteado en los experimentos 1 y 2, las diferencias provienen de los parámetros usados en la extracción de información para la clasificación, siendo en el experimento 1 el uso de estadísticas y en el experimento 2 características de una señal. Finalmente, los resultados del experimento 3 con la wavelet discreta, demuestran recomendaciones de otros autores sobre su uso en señales biológicas (Aungsakul et al., 2012) (Abbaspour, S., et al. 2020). Por otra parte, al trabajar con movimientos oculares deben considerarse factores como la fatiga ocular, ojos llorosos o secos, etc. Sin embargo, según expertos la fatiga no debe llevar a consecuencias graves a largo plazo (Fundación Mayo, 2025).

La arquitectura para la construcción de un prototipo domótico controlado por EOG se mostró viable con los módulos presentados con buenos resultados en la identificación de señales oculares. Usar características en tiempo para las señales EOG discretas genero resultados más exactos. Ambos clasificadores obtuvieron una exactitud de 97.8%, no obstante, se observan diferencias en velocidad de predicción (mayor con NN) y tamaño de los modelos en espacio en disco (menor con NN) observable en la Tabla 9, por ello en esta investigación, el clasificador seleccionado fue la red neuronal bilateral con dos capas conectadas internamente con 10 neuronas. Como trabajo futuro se espera continuar con la construcción del prototipo domótico concluyendo el último módulo de la arquitectura. También se planea experimentar con otro tipo de redes neuronales y otros conjuntos de señales.

Referencias

OMS (2023). *Disability and health*. Recuperado de www.who.int/es/news-room/fact-sheets/detail/disability-and-healt

Instituto Nacional de Estadística y Geografía (México) (2017). "La discapacidad en México, datos al 2014", INEGI, Recuperado de: www.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825094409.pdf

Lozano, F. (2008). "Control de mouse a través de señales EOG y algoritmos de Boosting", Universidad de los Andes, facultad de ingeniería, departamento de ingeniería electrónica, Recuperado de: repositorio.uniandes.edu.co/server/api/core/bitstreams/da02348b-e215-46df-bb27-d1f771db29bb/content.

Nathaniel, B., Camilleri, T. A. y Camilleri, K. P. (2018). "A comparison of EOG baseline drift mitigation techniques", *ELSEVIER*, vol. LVII.

Manolakis, D. y Vinay, I. (2011). "Effects of time-windowing on sinusoidal signals", Applied Digital Signal Processing: theory and practice (pp. CCCXCVII). Cambridge University Press.

Raschka, S. y Mirjalili, V. (2019). "Giving Computers the Ability to Learn from Data", Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2 (pp. I–IV). Packet Publishing.

Aungsakul, S. et al. (2012). "Development of an EOG-based system to control a serious game Evaluating Feature Extraction Methods of Electrooculography (EOG) Signal for Human-Computer Interface", *ELSEVIER*, vol. XXXII, pp. CCXLVI–CCLII.

Bulling, A. et al. (2011). "Eye Movement Analysis for Activity Recognition Using Electrooculography", *IEEE*, vol. XXXIII, pp. DCCXLI–DCCLIII.

Abbaspour, S., et al. (2020). "Evaluation of surface EMG-based recognition algorithms for decoding hand movements. Medical & Biological Engineering & Computing", *Springer*, vol. LVIII, pp. pp. LXXXIII–C.

Tuay, D., y Quijada, G. (2019). "Sistema de transmisión de datos electrooculográficos por medio de tecnología inalámbrica basado en CS y técnicas avanzadas de procesamiento para el desplazamiento de un vehículo", *Revista digital de Semilleros de Investigación REDSI*, vol. II, pp. I–V.

Nathaniel, B., Camilleri, T. A. y Camilleri, K. P., (2020). *Eye movement EOG data*. Recuperado de www.um.edu.mt/cbc/ourprojects/eyecon/eogdataset/.

Mulam, H. y Mudigonda, M. (2017). "A Novel method for Recognizing Eye Movements using NN classifier", *IEEE*, International Conference on Sensing, Signal Processing and Security (ICSSS), pp. CCLXL– CCLXLV.

Fundación Mayo para la Educación y la Investigación Médicas. (2025). *Fatiga ocular*. Recuperado de www.mayoclinic.org/es/diseases-conditions/eyestrain/symptoms-causes/syc-20372397.

Capítulo 7

Optimización con TLBO de los Parámetros de un Algoritmo de Marca de Agua de Imágenes Digitales basado en DWT y SVD

Alejandro Tonatiuh García Espinoza, Maya Carrillo Ruíz, Rogelio González Velázquez, Juan Francisco Leyva Bonilla

Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla, 14 Sur y Av. San Claudio, Col. San Manuel. C. P. 72570. Puebla, Puebla, México
ge224470256@alm.buap.mx, maya.carrillo@correo.buap.mx, rogelio.gonzalez@correo.buap.mx, juan.leyvab@correo.buap.mx

Resumen. La marca de agua de imágenes digitales brinda protección a los derechos de autor de los datos de la imagen, al ocultar la información adecuada de forma que no pueda eliminarse. Con el desarrollo de las redes informáticas y de comunicación, duplicar, alterar o incluso robar datos obtenidos en Internet se realiza con facilidad. En este trabajo, se propone el uso de la optimización basada en enseñanza-aprendizaje para la optimización de un algoritmo de marca de agua de imágenes cimentado en la transformada wavelet discreta y la descomposición en valores singulares. El algoritmo propuesto se compara con métodos existentes en la literatura usando imágenes clásicas del procesamiento digital de imágenes. Los resultados obtenidos permiten observar mejoras, cuando se compara con otras metaheurísticas, en cuanto a tiempo de ejecución, relación señal-ruido máxima y medida del índice de similitud estructural.

Palabras Clave: Optimización con metaheurísticas, TLBO, DWT, SVD.

1 Introducción

Con el rápido desarrollo de las redes informáticas y de comunicación, la distribución de contenido multimedia a través de Internet ha adquirido un gran auge. Dado su carácter natural, recordemos que Internet es un sistema abierto; los datos obtenidos se duplican, alteran o incluso se roban con facilidad (Kumsawat et al., 2005). En consecuencia, la implementación de la protección de los derechos de autor en los medios de comunicación se ha convertido en un asunto sumamente relevante.

La marca de agua en imágenes digitales protege los derechos de autor de los datos de una imagen al esconder información apropiada en la imagen original. Esto debe realizarse de forma que la información agregada no pueda eliminarse y no disminuya la calidad de la imagen original. Los métodos de marcas de agua de imágenes pueden clasificarse en dos categorías dependiendo de si usan la imagen original o no durante

el proceso de detección de marcas de agua. En la mayoría de los casos, la marca de agua se agrega a la imagen en el dominio espacial o en el dominio de transformación.

En esta investigación, se presenta un algoritmo optimizado de marca de agua en el dominio de la transformada discreta wavelet (DWT, por sus siglas en inglés) que hace uso del método de reducción de dimensionalidad de descomposición en valores singulares para la incrustación de la marca de agua. El algoritmo se mejoró utilizando la optimización basada en enseñanza-aprendizaje (TLBO, por sus siglas en inglés).

Este trabajo está organizado como se describe a continuación. En la sección 2 se hace una revisión de la literatura actual en algoritmos de marca de agua de imágenes. La sección 3 introduce los conceptos clave para entender el trabajo. El método propuesto en este trabajo se describe en la sección 4. La sección 5 muestra los resultados obtenidos y una comparación con métodos existentes. Finalmente, en la sección 6 se presentan las conclusiones del trabajo.

2 Estado del arte

Una extensa investigación se ha realizado para encontrar un algoritmo de marca de agua de imágenes que cumpla con los requisitos de invisibilidad y robustez que se busca para la protección de los derechos de autor de los datos de una imagen.

Dugad et al. (1998) proponen una técnica de marca de agua de imágenes en el dominio de la DWT, que no requiere la imagen original en el proceso de detección de marca de agua. Ellos incorporan una marca de agua con un factor de ponderación constante en los coeficientes significativos en el dominio de transformación para garantizar que la marca de agua no se pueda borrar. Lui et al. (2018) introducen un nuevo modelo de marca de agua de imágenes basado en el algoritmo de cifrado asimétrico RSA para garantizar la seguridad de los datos ocultos. Los autores resaltan que la mayoría de los trabajos se enfocan en la robustez y la capacidad de incrustación, pero requieren una cantidad considerable de recursos computacionales en el proceso cifrado o no garantizan la seguridad de los datos ocultos.

Con el fin de mejorar el rendimiento, los investigadores empezaron a optimizar sus algoritmos de marcas de agua de imágenes. Kumsawat et al. (2005) proponen un enfoque de optimización de algoritmos de marca de agua de imágenes digitales, obteniendo una mejora en el rendimiento de un algoritmo en el dominio de la transformada discreta multiwavelet usando optimización por medio de algoritmos genéticos. Sharma y Mir (2022) diseñan un método de incrustación de marcas de agua en imágenes en el dominio de la transformada discreta coseno (DCT, por sus siglas en inglés). Después utilizan la optimización por colonia de hormigas y el algoritmo de refuerzo de gradiente ligero para la predicción de los parámetros de incrustación óptimos para su algoritmo en un conjunto de imágenes. El-Kenawy et al. (2022), introducen un enfoque que hace uso de la DCT, la DWT, la optimización por mirlo acuático y el algoritmo de búsqueda fractal estocástica. La optimización se lleva a cabo para determinar los mejores dos parámetros para su algoritmo, siendo estos un factor de escala y un coeficiente de incrustación.

En la literatura también es posible encontrar trabajos que hacen uso de técnicas de aprendizaje automático en algoritmos de marca de agua de imágenes. Abdelhakim y Abdelhakim (2018) utilizan el método de regresión k-vecinos más cercanos para predecir los parámetros óptimos de incrustación de un algoritmo de marcas de agua de imágenes diseñado en el dominio de la DCT. Por su parte, Agilandeeswari y Ganesan (2016) utilizan un modelo de red neuronal de memoria asociativa bidireccional (BAM, por sus siglas en inglés) para un algoritmo de marca de aguas en múltiples imágenes. En su algoritmo, la matriz de pesos de la BAM se incrusta en los componentes de los coeficientes wavelet escogidos mediante el método de análisis de componentes principales (PCA, por sus siglas en inglés).

Otra técnica de reducción de dimensionalidad que se ha utilizado en algoritmos de marca de agua de imágenes es la descomposición en valores singulares (SVD, por sus siglas en inglés). Zhu et al. (2022) proponen un algoritmo de marca de agua de imágenes optimizado basado en la SVD y la transformada wavelet entera. Su propuesta combina los métodos mencionados con un algoritmo genético para optimizar el algoritmo. Khanna et al. (2016) similarmente crearon un algoritmo de marcas de agua de imágenes que utiliza una DWT de dos niveles y la técnica de la SVD para la incrustación de marcas de agua. También hacen uso de algoritmos genéticos para encontrar la fuerza de incrustación de la marca de agua de su algoritmo.

TLBO ha sido utilizado por Sivananthamaitrey y Kumar (2021) para optimizar un algoritmo de marca de agua dual (es decir, incrustan dos marcas de agua) de imágenes a color. Para esto utilizan la transformada wavelet estacionaria (SWT, por sus siglas en inglés) y la SVD. Por su parte, Krishna y Murty (2017) utilizan TLBO para mejorar un algoritmo de marca de agua de imágenes en el dominio de la DWT. Ellos dividen la imagen anfitriona en bloques para insertar la marca de agua en los bloques seleccionados.

En este trabajo, hacemos uso del algoritmo de optimización TLBO para mejorar un algoritmo de marca de agua que utiliza la DWT y el método de la SVD para la incrustación de la marca de agua.

3 Marco teórico

3.1 Transformada Wavelet Discreta

El objetivo principal de la transformada wavelet es la descomposición en múltiples resoluciones de señales e imágenes. Esto generalmente se realiza al crear un componente de aproximación usando una función escalar (filtro de paso bajo) y componentes de detalle que utilizan funciones wavelet (filtros de paso alto). La DWT es similar a la transformada discreta de Fourier y es ampliamente usada en aplicaciones prácticas (Sundararajan, 2015). Existe un número infinito de señales base para la DWT. Las señales comúnmente utilizadas son Haar, Daubechies, Coiflets y Biortogonal. En este trabajo se hace uso de la señal Haar. La DWT 2-D de una imagen $\boldsymbol{x}$ y su inversa de una imagen, se definen, respectivamente, en la ecuación 1.

$$\boldsymbol{X} = \boldsymbol{W}\boldsymbol{x}\boldsymbol{W}^T \quad \text{y} \quad \boldsymbol{x} = \boldsymbol{W}^T\boldsymbol{X}\,\boldsymbol{W} \tag{1}$$

donde $\boldsymbol{W}$ es la matriz de la transformada de Haar para la DWT de nivel 1. La matriz de la transformada de Haar para la DWT de nivel 1 de una secuencia de longitud N (N es una potencia de 2) se define en la ecuación 2.

$$\boldsymbol{W}_{N,(\log_2(N)-1)} = \begin{bmatrix} \boldsymbol{I}_{\frac{N}{2}} \otimes \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \\ \boldsymbol{I}_{\frac{N}{2}} \otimes \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \end{bmatrix} \tag{2}$$

donde $\boldsymbol{A} \otimes \boldsymbol{B}$ es el producto de Kronecker de las matrices $\boldsymbol{A}$ y $\boldsymbol{B}$, e $\boldsymbol{I}_N$ es la matriz de identidad de $N \times N$.

Con la matriz de la transformada de nivel 1 $\boldsymbol{W}$ particionada como en la ecuación 3.

$$\begin{aligned} \boldsymbol{W} = \begin{bmatrix} \boldsymbol{L} \\ \boldsymbol{H} \end{bmatrix}, \quad \boldsymbol{W}\boldsymbol{x}\boldsymbol{W}^T &= \begin{bmatrix} \boldsymbol{L} \\ \boldsymbol{H} \end{bmatrix} \boldsymbol{x} \begin{bmatrix} \boldsymbol{L} \\ \boldsymbol{H} \end{bmatrix}^T = \begin{bmatrix} \boldsymbol{L}\boldsymbol{x} \\ \boldsymbol{H}\boldsymbol{x} \end{bmatrix} [\boldsymbol{L}^T \quad \boldsymbol{H}^T] \\ &= \begin{bmatrix} \boldsymbol{L}\boldsymbol{x}\boldsymbol{L}^T & \boldsymbol{L}\boldsymbol{x}\boldsymbol{H}^T \\ \boldsymbol{H}\boldsymbol{x}\boldsymbol{L}^T & \boldsymbol{H}\boldsymbol{x}\boldsymbol{H}^T \end{bmatrix} = \begin{bmatrix} X_\phi^j & X_H^j \\ X_V^j & X_D^j \end{bmatrix} \end{aligned} \tag{3}$$

Cada descomposición de una imagen produce cuatro subimágenes de un cuarto de tamaño, X_ϕ^j, X_H^j, X_V^j, y X_D^j. El componente X_ϕ^j es una versión suavizada de la entrada. Los componentes X_H^j, X_V^j, y X_D^j son medidas de diferencias horizontales, verticales y diagonales, respectivamente. La DWT a menudo se usa en algoritmos de marca de agua de imágenes, donde la incrustación de la marca de agua se aplica a uno o más componentes resultantes de la descomposición de la imagen de entrada (Sundararajan, 2015).

3.2 Descomposición en valores singulares

La descomposición en valores singulares es un método de reducción de dimensionalidad comúnmente usado en álgebra lineal. Esta herramienta matemática puede ser usada en muchas aplicaciones, incluyendo el procesamiento de imágenes digitales. En el procesamiento de imágenes, la SVD tiene buena estabilidad porque los valores singulares prácticamente no cambian (Zhu et al., 2022). Por lo tanto, si se alteran ligeramente estos valores, para incrustar información, la calidad visual de la imagen no se verá degrada de manera perceptible.

La fórmula para descomponer una imagen I de $M \times N$ se muestra a continuación

$$I = U \cdot S \cdot V^T \tag{4}$$

donde U y V representan matrices de $M \times M$ y $N \times N$, respectivamente, y S es una matriz diagonal de $M \times N$ que contiene los valores singulares de la imagen.

3.3 Optimización basada en enseñanza-aprendizaje

Rao et al. (2011) propusieron la optimización basada en enseñanza-aprendizaje inicialmente para la optimización de problemas de diseño mecánico. El método TLBO se basa en la influencia del profesor en el rendimiento académico de los alumnos. El profesor es generalmente considerado una persona con un alto nivel de formación que comparte sus conocimientos con los alumnos. La calidad del profesor influye en el rendimiento académico de los alumnos. Es evidente que un buen profesor capacita a sus alumnos para que obtengan mejores resultados en sus calificaciones.

El proceso de TLBO se divide en dos fases. La primera parte consiste en la "fase de enseñanza" y la segunda parte consta de la "fase de aprendizaje". La "fase de enseñanza" significa aprender del profesor y la "fase de aprendizaje" significa aprender a través de la interacción entre alumnos.

La figura 1 muestra el diagrama de flujo para la optimización basada en enseñanza-aprendizaje. Un buen profesor es aquel que eleva el nivel del conocimiento de sus alumnos a su nivel. Sin embargo, en la práctica, esto no es posible, y un profesor solo puede mejorar el promedio de una clase hasta cierto punto. En la fase de enseñanza de la TLBO cada solución se actualiza según la ecuación 5.

$$X_{new} = X_{old} + r \cdot (X_{teacher} - T_F \cdot \mu) \tag{5}$$

donde X_{new} es el nuevo valor de X_{old}, $T_F = \text{round}[1 + \text{rand}(0,1)]$ es un factor de enseñanza que decide el valor de la media que se va a cambiar, μ es la media de la población y r es un número aleatorio en el rango $[0, 1]$.

Un alumno interactúa aleatoriamente con otros alumnos mediante debates grupales, presentaciones, comunicaciones formales, etc. Este aprende algo nuevo si el otro alumno tiene más conocimientos que él. La actualización de cada solución en TLBO en la fase de aprendizaje se expresa en la ecuación 6.

$$X_{new} = \begin{cases} X_{old} + r \cdot (X_i - X_j), & \text{si } f(X_i) < f(X_j) \\ X_{old} + r \cdot (X_j - X_i), & \text{si } f(X_j) < f(X_i) \end{cases} \tag{6}$$

donde X_{new} es el nuevo valor de X_{old}, X_i y X_j son dos alumnos seleccionados aleatoriamente, r es un número aleatorio en el rango $[0, 1]$ y $f(x)$ es la función objetivo a minimizar. El algoritmo termina al cumplir los criterios de parada deseados, como alcanzar un número máximo de iteraciones.

4 Método propuesto

En esta sección se presenta un esquema que contiene dos procesos principales: el proceso de incrustación de marca de agua y el proceso de extracción de marca de agua. El preprocesamiento y procedimiento de incrustación se llevan a cabo en el lado del emisor, y la imagen con marca de agua se transmite al receptor a través de Internet. En el punto receptor, se lleva a cabo el proceso de extracción de marca de agua y

recuperación. En el proceso de incrustación de marca de agua, una transformada wavelet discreta con señal base Haar se aplica a la imagen anfitriona y se obtiene el componente de aproximación X_{ϕ}^{j}. Este componente se procesa más adelante mediante una descomposición en valores singulares al igual que la marca de agua a incrustar. Sumando el valor singular del componente y el de la marca de agua, se obtiene un nuevo valor singular. En la operación de adición, un factor escalar se escoge para controlar la fuerza de incrustación de la marca de agua, este parámetro es el que se busca optimizar usando la TLBO. Este nuevo valor singular se utiliza para obtener un nuevo componente de aproximación aplicando la inversa de la SVD. Finalmente, la imagen con marca de agua se obtiene al aplicar la transformada wavelet discreta inversa con el nuevo componente de aproximación.

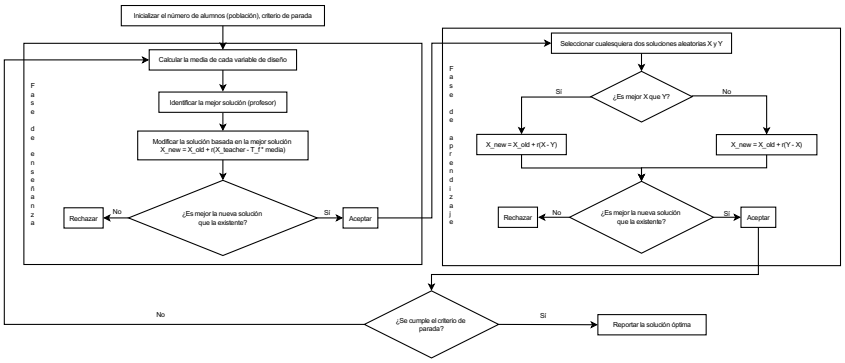


Figura 1. Diagrama de flujo para la optimización basada en enseñanza-aprendizaje (creación propia).

1.1 Proceso de incrustación de la marca de agua

Los pasos específicos para la inserción de la marca de agua se describen a continuación.

1. La imagen anfitriona I, en escalas de grises, se descompone en cuatro componentes $X_{\phi}^{j}, X_{H}^{j}, X_{V}^{j}, X_{D}^{j} = \mathrm{DWT}(I)$.
2. La SVD se aplica en el componente de aproximación X_{ϕ}^{j} y en la marca de agua W, de tal forma que $U_I, S_I, V_I = \mathrm{SVD}(X_{\phi}^{j})$ y $U_W, S_W, V_W = \mathrm{SVD}(W)$.
3. Se calcula un nuevo valor singular S_{new} al sumar S_I y el valor singular de la marca de agua S_W con un factor α. Así, $S_{new} = S_I + \alpha S_W$.
4. Se aplica la inversa de la SVD al nuevo valor singular, para obtener un coeficiente de aproximación modificado $X'^{j}_{\phi} = U_I \cdot S_{new} \cdot V_I^T$.
5. Se obtiene la imagen con marca de agua I_W al realizar la DWT inversa con el coeficiente de aproximación modificado, $I_W = \mathrm{IDWT}(X'^{j}_{\phi})$.

1.2 Proceso de extracción de la marca de agua

Para la extracción de la marca de agua, la DWT se aplica sobre la imagen con marca de agua, y se descompone en el componente de aproximación y los componentes de detalle. La marca de agua se obtiene mediante el uso de la imagen anfitriona original, la marca de agua original y el reciente valor singular S_{new}. Los pasos detallados son como a continuación.

1. La imagen con marca de agua I_W y la imagen anfitriona original I se descomponen en sus cuatro componentes $X'^{j}_{\phi}, X'^{j}_{H}, X'^{j}_{V}, X'^{j}_{D} = \text{DWT}(I_W)$ y X^{j}_{ϕ}, $X^{j}_{H}, X^{j}_{V}, X^{j}_{D} = \text{DWT}(I)$.
2. La SVD se aplica en los componentes de aproximación X^{j}_{ϕ}y X'^{j}_{ϕ}, de tal forma que $U_I, S_I, V_I = \text{SVD}(X^{j}_{\phi})$ y $U_{I_W}, S_{I_W}, V_{I_W} = \text{SVD}(I_W)$.
3. Se calcula un nuevo valor singular S_{new} con la siguiente fórmula $S_{new} = (S_{I_W} - S_I)/\alpha$.
4. Se obtiene un coeficiente de aproximación modificado $\hat{X}^{j}_{\phi}$ aplicando la inversa de la SVD al nuevo valor singular $\hat{X}^{j}_{\phi} = U_W \cdot S_{new} \cdot V_W^T$.
5. Se obtiene la marca de agua incrustada W' en la imagen anfitriona al realizar la DWT inversa con el coeficiente de aproximación modificado $W' = \text{IDWT}(\hat{X}^{j}_{\phi})$.

1.3 Mejorando el rendimiento de la marca de agua usando metaheurísticas

Este trabajo emplea la TLBO para buscar parámetros apropiados con el fin de lograr el rendimiento óptimo del algoritmo de marca de agua de imágenes digitales descrito anteriormente. El parámetro que se busca optimizar es el factor escalar para controlar la fuerza de incrustación de la marca de agua. Los detalles de la metaheurística se describen más adelante.

Función objetivo. La función objetivo consiste en una combinación de la relación señal-ruido máxima (PSNR por sus siglas en inglés) y el coeficiente de correlación normalizado (NCC, por sus siglas en inglés). La PSNR compara la marca de agua original W y la marca de agua extraída W', ambas de tamaño $M \times N$, de la forma que se muestra en la ecuación 7.

$$\text{PSNR} = 10 \log \frac{255^2}{\frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(W(i,j) - W'(i,j)\right)^2} \tag{7}$$

Note que el denominador es igual al error cuadrático medio (MSE, por sus siglas en inglés). Por su parte, el coeficiente de correlación se define en la ecuación 8.

$$\mathrm{NCC} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(W(i,j) - \bar{W})\left(W'(i,j) - \overline{W'}\right)}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(W(i,j) - \bar{W})^2}\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(W'(i,j) - \overline{W'})^2}} \quad (8)$$

Así, la función objetivo $f(\alpha)$ a optimizar es la que se observa en la ecuación 9.

$$f(\alpha) = \mathrm{PSNR}(W, W'_{\alpha}) + \frac{1}{K}\sum_{i=1}^{K}\mathrm{NCC}(W, W_{\alpha}^{i}), \quad (9)$$

donde W'_{α} es la marca de agua extraída al usar α como fuerza de incrustación y W_{α}^{i} es la marca de agua extraída de una imagen atacada por el i-enésimo tipo de ataque, al usar α como fuerza de incrustación. Los tipos de ataques utilizados en este trabajo se mencionarán más adelante.

Codificación de los individuos (alumnos). Cada individuo consiste en un número real en el rango $[-5, 5]$. Cada individuo representa el factor escalar para controlar la fuerza de incrustación de la marca de agua.

Ajuste de parámetros. La TLBO carece de parámetros específicos, y los únicos parámetros que se deben ajustar son los criterios de parada y el tamaño de población. Como criterio de parada utilizamos un máximo de iteraciones. Se decidió por usar un máximo de 30 iteraciones y una población de 5 individuos. Estos valores se determinaron de manera empírica.

5 Resultados y discusión

El rendimiento del método propuesto se evalúa a través de simulaciones experimentales usando el lenguaje de programación Python en una computadora Windows con un procesador 11th Gen Intel(R) Core(TM) i5-1135G7 @ 2.40GHz y 8.00 GB RAM. Ocho imágenes clásicas en el procesamiento de imágenes digitales, en escala de grises y de tamaño 512×512, fueron seleccionadas como imágenes anfitrionas. Como marca de agua se seleccionó la imagen "Yacht", también clásica en el procesamiento de imágenes digitales. Las imágenes de prueba se muestran en la figura 3.

Las métricas utilizadas para medir el rendimiento de ambos algoritmos son: PSNR, NCC, SSIM y tiempo de ejecución. SSIM son las siglas en inglés de la medida del

índice de similitud estructural y se define en la ecuación 10.

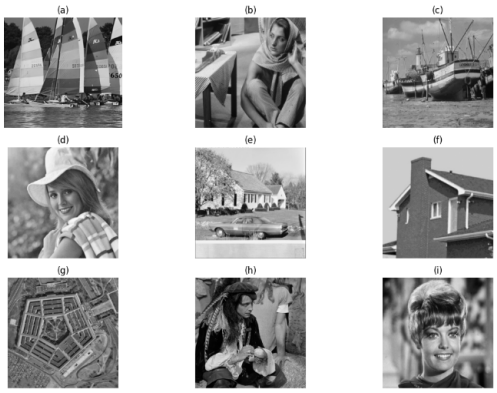


Figura 3. Las imágenes del conjunto de prueba: (a) Yacht; (b) Barbara; (c) Boats; (d) Elaine; (e) House 1; (f) House 2; (g) Pentagon; (h) Pirate; (i) Zelda.

$$SSIM = \frac{(2\mu_W\mu_{W'} + C_1)(2\sigma_{W,W'} + C_2)}{(\mu_W^2 + {\mu'_W}^2 + C_1)(\sigma_W^2 + {\sigma'_W}^2 + C_2)} \quad (10)$$

donde μ_W y $\mu_{W'}$ representan la media de la luminancia de la marca de agua original W y la marca de agua extraída W', respectivamente. σ_W y $\sigma_{W'}$ denotan la varianza de los valores de los píxeles, $\sigma_{W,W'}$ es la covarianza entre las marcas de agua, y C_1 y C_2 son constantes de estabilización. La tabla 1 muestra los tiempos de ejecución del método propuesto y otros algoritmos de marca de agua de imágenes en la literatura. Se observa que el método propuesto en este trabajo consigue una mejora significativa en tiempo de ejecución en comparación con otras metaheurísticas. Sin embargo, los métodos de aprendizaje automático son considerablemente más rápidos que los optimizados por metaheurísticas, como K-vecinos más cercanos y el algoritmo de refuerzo de gradiente de luz (K-NN y LGBA, respectivamente).

Tabla 1. Comparación del método propuesto con otros métodos, en términos de tiempo de ejecución en segundos.

Abdelhakim y Abdelhakim (2018)		Sharma y Mir (2022)		Método propuesto optimizado
ABC	K-NN	ACO	LGBA	
1393.1	23.4	1169.9	**16.512**	184.913

La tabla 2 muestra los resultados de la evaluación de similitud entre la marca de agua original y la extraída. Un valor 1, en similitud, significa que las marcas de agua son idénticas. Una comparación con respecto a la calidad de la marca de agua en términos de la PSNR se observa en la tabla 3. El PSNR se expresa en decibelios (dB) y a mayor valor indica menor error entre las imágenes. Valores mayores a 50 indican que la calidad de la imagen es casi idéntica a la original. Un mayor tiempo de ejecución de

nuestro método se compensa con un aumento en la calidad de la marca de agua, con respecto a otras metaheurísticas y algoritmos de aprendizaje. También se llevó a cabo la evaluación de robustez utilizando una imagen. La evaluación de robustez se realiza con ataques a la imagen Barbara, que se encuentra entre las imágenes con mejor rendimiento, como se observa en la tabla 3. Un ataque es cualquier modificación accidental o intencional de la imagen. La tabla 4 muestra los resultados de una evaluación de robustez en términos del NCC en la imagen Barbara. Los tipos de ataque fueron seleccionados de Agilandeeswari y Ganesan (2016). En la figura 4 se observan las marcas de agua extraídas de esta evaluación de robustez.

Tabla 2. Evaluación de la similitud entre la marca de agua original y la extraída en términos del SSIM.

	Barbara	Boat	Elaine	House 1	House 2	Pentagon	Pirate	Zelda
SSIM	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999

Los valores arrojados en las diferentes evaluaciones muestran un desempeño prometedor del algoritmo propuesto en este trabajo, pero con oportunidad de mejora en la robustez del algoritmo, en comparación con métodos existentes.

Tabla 3. Comparación de la calidad de la marca de agua del método propuesto y otros métodos en la literatura, en términos de la PSNR.

Imagen de prueba	Abdelhakim y Abdelhakim (2018)		Sharma y Mir (2022)		Método propuesto optimizado
	ABC	K-NN	ACO	LGBA	
Barbara	40.552	40.583	44.882	44.983	**78.130**
Boat	40.683	40.629	44.418	44.798	**71.467**
Elaine	41.165	41.239	46.381	46.498	**85.882**
House 1	40.513	40.493	44.187	44.582	**74.420**
House 2	45.269	45.325	49.512	49.822	**54.420**
Pentagon	38.600	38.727	42.612	42.819	**66.080**
Pirate	41.043	41.203	45.685	45.791	**72.110**
Zelda	43.282	43.347	47.651	47.898	**73.881**

6 Conclusiones

En este trabajo, se propone el uso de la optimización basada en enseñanza-aprendizaje para la optimización de un algoritmo de marca de agua en imágenes digitales. Obteniendo un desempeño comparable con los métodos existentes en la literatura y logrando una mejora en la calidad de la imagen extraída de una imagen anfitriona. Se logra una reducción en tiempo de ejecución en comparación con otros algoritmos optimizados mediante metaheurísticas. Como trabajo futuro, se establece como objetivo mejorar la robustez del algoritmo y realizar experimentos con más tipos de ataques mencionados en la literatura.

Tabla 4. Evaluación de robustez en términos del NCC en la imagen Barbara.

Índice	Tipo de ataque	Parámetros	NCC
(a)	Ruido gaussiano	$\sigma^2 = 0.1$	0.191
(b)	Ruido gaussiano	$\sigma^2 = 0.5$	0.259
(c)	Ruido sal y pimienta	Densidad $= 0.02$	0.810
(d)	Ruido sal y pimienta	Densidad $= 0.06$	0.614
(e)	Ruido de Poisson	-	0.840
(f)	Filtro de mediana	Vecindad de 3×3	0.774
(g)	Filtro de mediana	Vecindad de 5×5	0.718
(h)	Ecualización de histograma	-	0.777
(i)	Ajuste de contraste	-	0.785

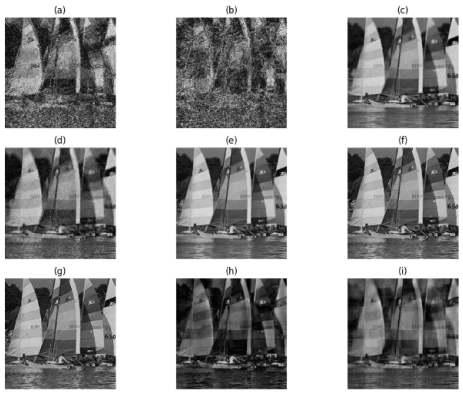


Figura 4. Resultados de la extracción de las marcas de agua de la imagen Barbara atacada, indexados por los tipos de ataque de la tabla 4.

7 Agradecimientos

El primer autor agradece al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo otorgado con la beca 2008590 para la realización de este trabajo.

Referencias

Abdelhakim, A.M. y Abdelhakim, M. (2018) "A time-efficient optimization for robust image watermarking using machine learning," Expert Systems with Applications, 100, pp. 197–210. Disponible en: https://doi.org/https://doi.org/10.1016/j.eswa.2018.02.002.

Agilandeeswari, L. y Ganesan, K. (2016) "A bi-directional associative memory based multiple image watermarking on cover video," *Multimedia Tools and Applications*, 75(12), pp. 7211–7256. Disponible en: https://doi.org/10.1007/s11042-015-2642-1.

Dugad, R., Ratakonda, K. y Ahuja, N. (1998) “A new wavelet-based scheme for watermarking images,” en *Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No.98CB36269)*, vol. 2, pp. 419–423. Disponible en: https://doi.org/10.1109/ICIP.1998.723406.

El-Kenawy, E.-S.M. *et al.* (2022) “Advanced Dipper-Throated Meta-Heuristic Optimization Algorithm for Digital Image Watermarking,” *Applied Sciences*, 12(20). Disponible en: https://doi.org/10.3390/app122010642.

Khanna, A.K., Roy, N.R. y Verma, B. (2016) “Digital image watermarking and its optimization using genetic algorithm,” en *2016 International Conference on Computing, Communication and Automation (ICCCA)*, pp. 1140–1144. Disponible en: https://doi.org/10.1109/CCAA.2016.7813888.

Krishna, A.V.N.D. y Murty, P.S. (2017) “Authentication of digital visual data using discrete wavelet transform and TLBO,” in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1706–1712. Available at: https://doi.org/10.1109/ICACCI.2017.8126089.

Kumsawat, P., Attakitmongcol, K. y Srikaew, A. (2005) “A new approach for optimization in image watermarking by using genetic algorithms,” *IEEE Transactions on Signal Processing*, vol. 53(12), pp. 4707–4719. Disponible en: https://doi.org/10.1109/TSP.2005.859323.

Liu, Y. *et al.* (2018) “Secure and robust digital image watermarking scheme using logistic and RSA encryption,” *Expert Systems with Applications*, vol. 97, pp. 95–105. Disponible en: https://doi.org/https://doi.org/10.1016/j.eswa.2017.12.003.

Rao, R. v, Savsani, V.J. y Vakharia, D.P. (2011) “Teaching–learning-based optimization: A novel method for constrained mechanical design optimization problems,” *Computer-Aided Design*, vol. 43(3), pp. 303–315. Disponible en: https://doi.org/https://doi.org/10.1016/j.cad.2010.12.015.

Sharma, V.K. y Mir, R.N. (2022) “An enhanced time efficient technique for image watermarking using ant colony optimization and light gradient boosting algorithm,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34(3), pp. 615–626. Disponible en: https://doi.org/https://doi.org/10.1016/j.jksuci.2019.03.009.

Sivananthamaitrey, P. y Kumar, P.R. (2021) “Teaching-Learning Based Optimization of Dual Watermarking of Color Images,” in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 1204–1209. Available at: https://doi.org/10.1109/ICCMC51019.2021.9418382.

Sundararajan, D. (2015) “The Haar Discrete Wavelet Transform,” en *Discrete Wavelet Transform: A Signal Processing Approach*. Wiley, pp. 97–130. Disponible en: https://doi.org/10.1002/9781119113119.ch8.

Zhu, T., Qu, W. y Cao, W. (2022) “An optimized image watermarking algorithm based on SVD and IWT,” The Journal of Supercomputing, vol. 78(1), pp. 222–237. Disponible en: https://doi.org/10.1007/s11227-021-03886-2.

Capítulo 8

Q-learning aplicado a un modelo de crecimiento económico

Gustavo Portillo Ramírez[1], Hugo A. Cruz Suárez[2]

[1] Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla,
[2] Facultad de Ciencias Físico Matemáticas, Benemérita Universidad Autónoma de Puebla

[1]gustavo.portillora@correo.buap.mx, [2]hcs@fcfm.buap.mx

Resumen. En este trabajo se presenta un modelo de crecimiento económico que se aborda como un problema de control óptimo vía la teoría de procesos de decisión de Markov. Este problema se resuelve mediante aprendizaje por refuerzo, específicamente se utiliza la metodología de Q-learning para encontrar políticas casi óptimas y aproximaciones al costo óptimo. Finalmente se proporciona la implementación numérica, la cual se desarrolló en el software R.

Palabras Clave: Q-learning, control óptimo, proceso de decisión de Markov.

1 Introducción

En el área de optimización estocásticas se pueden encontrar modelos de tipo secuencial, en los que en cada etapa de observación del sistema se debe proporcionar una decisión que generará un costo o una recompensa. Además, cada decisión puede repercutir en el contexto de decisiones futuras, por ejemplo, si se requiere minimizar un costo total acumulado, debería ponderarse la necesidad de minimizar el costo de la decisión actual con necesidad de eludir situaciones futuras en las que el costo alto no se puede evitar. Algunos ejemplos de problemas de decisión secuencial se mencionan a continuación: problemas de control óptimo determinista y problemas de control óptimo estocástico, problemas de decisión de Markov, problemas de decisión semi-Markovianos, problemas de control minimax y juegos secuenciales, Bertsekas y Shreve (1996). En este escrito utilizamos procesos de decisión de Markov (PDMs) para abordar un modelo de crecimiento económico que es resuelto vía Q-learning. Los PDMs son una herramienta poderosa que ha sido aplicada en una gran diversidad de aplicaciones, entre ellas, la industria Sodachi et al. (2024), inteligencia artificial Suilen et al. (2024), robótica Cao (2021), Kurniawati (2022) y economía Boucherie y Van Dijk (2017), Hernández-Hernández (2012).

Uno de los métodos principales para el análisis de problemas de decisión secuencial es la programación dinámica Bellman (2003). Esta metodología se popularizó como consecuencia de los estudios sistemáticos que desarrolló Bellman. Durante los primeros años de la teoría de procesos de decisión de Markov gran parte de la investigación se

centró solo en el estudio de ecuaciones de optimización y de los métodos para su solución, la cual incluye la iteración de políticas e iteración de valores. Por lo general, programación dinámica se aplica a problemas de horizonte infinito de criterio único con un costo/recompensa total esperado o un costo/recompensa promedio por unidad de tiempo, para consultar detalles ver Puterman (2005), Feinberg y Shwartz (2012).

En la literatura, se han documentado situaciones en las que los PDMs presentan algunas problemáticas en la búsqueda de soluciones óptimas. Esto suele ocurrir, cuando se presentan dimensiones altas en el espacio de estados o si existen dificultades con el modelaje, por ejemplo, si no la matriz de transición se encuentra ausente. Las dos limitaciones anteriores se denominan en la literatura la *doble maldición de la programación dinámica* Gosavi (2015). Una opción útil para combatir esta doble maldición es la metodología de Aprendizaje por Refuerzo (AR) a través de Q-learning. Lo anterior es posible debido a que el cómputo de modelado en AR es menor que el proporcionado por programación dinámica. Se debe destacar que en PDMs desarrollados a gran escala, la metodología de programación dinámica ya no es viable, en cambio AR lo sigue siendo Gosavi (2015). Adicionalmente, para enfrentar problemas donde las probabilidades de transición sean complicadas de estimar, el enfoque que provee AR es sugestivo y tiene la capacidad de aportar soluciones casi óptimas. AR utiliza como herramienta principal a la simulación, que es utilizada para eludir los cálculos desarrollados en las probabilidades de transición. Actualmente Q-learning sigue en desarrollo, por ejemplo, Neufeld (2024) presenta la viabilidad de un algoritmo de Q-learning, así como las ventajas de considerar la robustez distribucional al resolver problemas de control óptimo estocástico, en particular cuando las distribuciones estimadas resultan estar mal especificadas en la práctica. Además, recientemente desde el punto de vista teórico, Wan et al. (2024) han extendido el análisis de convergencia casi-segura de los algoritmos de Q-learning basados en iteración relativa.

En Gosavi (2015), el AR se establece como una rama de programación dinámica, por lo que el algoritmo correspondiente se origina de su contraparte de programación dinámica. En el enfoque de programación dinámica se requiere, desde el inicio, construir dos matrices; matriz probabilidad de transición y matriz de transición de costos/recompensas, posteriormente, se introducen ambas matrices en un algoritmo que generará la solución óptima. En cambio, el AR no se estima si quiera una matriz, sino que simula el sistema vía las distribuciones de probabilidad de las variables aleatorias que gobiernan el comportamiento de este. Finalmente, en el interior de un simulador, se aprovechas un algoritmo apropiado que provea la solución óptima.

En este trabajo, se utiliza la metodología proporcionada por AR para dar solución numérica al modelo de crecimiento económico utilizando el software R, R Core Team (2024). Para lograr lo anterior, es necesario identificar el modelo de crecimiento económico proporcionado por Brock y Mirman (1972) con un proceso de decisión de Markov, ya que esta herramienta nos provee de la política que maximiza la utilidad esperada descontada. Sin embargo, al final de este procedimiento, no se tiene acceso a la ley de transición de los estados de la cadena de Markov asociada. Aun así, con el enfoque de PDMs se tiene acceso a la ecuación en diferencias que describe la evolución del sistema estocástico que modela el capital en cada etapa del tiempo. Esta información

es suficiente para el aprendizaje por refuerzo ya que producirá una política casi óptima. En este contexto, se analiza el comportamiento del modelo de crecimiento económico cuando la desviación estándar σ de los choques tecnológicos estocásticos (un choque tecnológico ocurre cuando el desarrollo tecnológico afecta a la productividad) convergen a cero. En algunas situaciones, en la línea de investigación de sistemas estocásticos perturbados con ruidos pequeños se presenta una comparación con el sistema determinista, esto ocurre cuando las variables de perturbación se desvanecen. Esto es de gran relevancia ya que en la práctica se puede utilizar la solución del sistema aleatorio perturbado para el sistema determinista y viceversa. En este contexto, PDMs a tiempo discreto han sido estudiados por Cruz-Suárez et al. (2009), Cruz-Suárez y Ilhuicatzi-Roldán (2010), en los cuales la dinámica del sistema está inducida por una ecuación en diferencias. Recientemente, en el estudio Portillo-Ramírez et al. (2023), se incorporaron PDMs que se desarrollan mediante un par de ecuaciones en diferencias con dinámicas acopladas. En el enfoque propuesto se utilizan restricciones de continuidad de Lipschitz para los elementos que constituyen el modelo de control. Posteriormente se utilizan procedimientos de PD para garantizar: convergencia del tipo uniforme de las políticas óptimas estocásticas a la política determinista sobre subconjuntos compactos del espacio de estados, esto cuando dos parámetros numéricos dependientes de la estocasticidad convergen a cero; y se proporciona una tasa de convergencia para el comportamiento de la función de costo óptimo del sistema aleatorio con respecto a la función de costo del sistema determinista.

En este trabajo se aplica Q-learning al modelo de crecimiento económico considerando la relación entre sistemas estocásticos perturbados por ruidos pequeños y por sistemas deterministas. En este contexto se obtiene el valor del capital promedio y se presentan realizaciones de la trayectoria óptima del proceso estocástico, cuando $\sigma \to 0$, es decir, se presenta la trayectoria del proceso cuando se aplica la política casi óptima y el estado inicial corresponde al estado estable del sistema determinista asociado.

Este escrito se encuentra organizado de la manera siguiente. En la Sección 2 se presenta la teoría de procesos de decisión de Markov asociados a un criterio de rendimiento descontado. En la Sección 3, se aborda la metodología del aprendizaje por refuerzo, que es deducida de su contraparte; programación dinámica. En la Sección 4, se describe el modelo de crecimiento económico que se resuelve de manera numérica, el modelo corresponde a una economía especializada de Brock y Mirman (1972). En la Sección 5 se muestran los resultados de la implementación numérica; se proveen algunas realizaciones de la trayectoria óptima del sistema estocástico, bajo el uso de la solución casi óptima que proporciona el algoritmo Q-learning, cuando el parámetro de ruido pequeño converge a cero y presentamos el valor del costo del capital promedio. Finalmente, en la Sección 6 se proporcionan los comentarios finales del uso de AR en la resolución del problema de crecimiento económico.

2 Procesos de decisión de Markov

La teoría de procesos de decisión de Markov presentada en este trabajo es referente a optimización secuencial de sistemas estocásticos a tiempo discreto. En general, los PDMs se definen a través de los siguientes objetos: un espacio de estados X; conjunto

de acciones A; una familia $\{A(x)|\ x \in X\}$ de subconjuntos medibles de A, en este punto $A(x)$ denota el conjunto de acciones admisibles cuando el estado actual es $x \in X$; $p_{x,y}(a)$ denota a las probabilidades de transición donde $x, y \in X$ y $a \in A(x)$, a este tipo de probabilidades de transición se denominan controladas; una función de costo $C: X \times A \to R$, donde $C(x,a)$ denota el costo en un paso cuando se aplica la acción a en el estado x, $x \in X$ y $a \in A(x)$. Adicionalmente, se considera el conjunto de pareja estado-acción admisible, denotado por K, y definido por $K = \{(x,a)|\ x \in X, a \in A(x)\}$. K es un subconjunto medible de $X \times A$. En resumen, denotamos un PDM por el modelo $\mathcal{M} = (X, A, \{A(x)|x \in A\}, C, [p_{xy}(a)])$, el cual queda totalmente determinado por las componentes antes descritas.

La interpretación del modelo es la siguiente: cuando el sistema se encuentra en el estado x un controlador selecciona una acción $a \in A(x)$. Después, ocurren dos cosas: (a) el sistema transita al siguiente estado y, de acuerdo con la ley de transición controlada $p_{x,y}(a)$ y (b) se incurre en un costo $C(x,a)$. Finalmente, y es el estado actual y luego el proceso se repite nuevamente.

En PDMs el objetivo es determinar una estrategia de operación óptima del sistema en estudio. Las estrategias son seleccionadas en función de la información que se cuenta en cada instante de tiempo, por ello es necesario definir el espacio de historias. Para cada $t \geq 0$, *el espacio de historias,* hasta el tiempo t, denotado por $\mathcal{H}_t$, se define como $\mathcal{H}_0 = X$ y $\mathcal{H}_t = K \times X$ si $t \geq 1$, donde $K = \{(x,a): x \in X, a \in A\}$. Una política es una sucesión $\pi = \{\pi_t\}$ de *kérneles* estocásticos, en la que π_t está definido sobre el espacio de acciones A dada la historia $\mathcal{H}_t$ que cumple la condición $\pi_t(A(x_t)|h_t) = 1$ para toda $h_t \in \mathcal{H}_t$, $t \geq 0$. Además, se denota por $P(A|X)$ a la colección de probabilidades condicionales sobre A dado X. Considere Φ el conjunto que contiene a cualquier probabilidad condicional ϕ en $P(A|X)$ tal que para todo $x \in X : \phi(A(x)|x) = 1$.

En el contexto anterior, se distinguen tres tipos de política $\pi = \{\pi_t\}$: *Markoviana aleatorizada*, si existe una sucesión $\{\phi_t\} \subseteq \Phi$ tal que $\pi_t(\cdot\,|h_t) = \phi(\cdot\,|x_t)$, para todo $h_t \in \mathcal{H}_t$ y $t \geq 0$. *Determinista*, si existe una sucesión de funciones medibles $g_t: \mathcal{H}_t \to A$, tales que cualquiera sea $h_t \in \mathcal{H}_t$ y $t \geq 0$ se obtiene que $g_t(h_t) \in A(x_t)$ y $\pi_t(\cdot\,|\mathrm{h_t})$ está concentrada en $g_t(h_t)$. *Determinista Markov-estacionaria* si existe una función medible $f: X \to A$, que satisface $f(x_t) \in A(x_t)$ y $\pi_t(\cdot\,|h_t)$ está concentrada en $f(x_t)$ para cada $h_t \in \mathcal{H}_t$ y $t \geq 0$. El conjunto de políticas estacionarias se denota por F.
Un PDM se dice que es *finito* si el espacio de estados X y el conjunto de acciones A son finitos. Adicionalmente, un PDM es llamado *discreto* si el espacio de estados y el conjunto de acciones son conjuntos discretos.
Un elemento vital en PDMs es el criterio de rendimiento, que toma el papel de la herramienta básica que permite comparar políticas. A continuación, se describe el criterio de rendimiento que utilizaremos en este trabajo. El costo total esperado descontado sobre un horizonte infinito se define como

$$V(x,\pi) = E_x^{\pi}\left[\sum_{n=1}^{\infty} \beta^n C(x_n, a_n)\right], \qquad (1)$$

donde E_x^{π} denota la esperanza condicional dado el estado $x \in X$ y la política $\pi \in \Pi$, Π denota el conjunto de todas las políticas y $\beta \in [0,1]$ es el factor de descuento, Feinberg y Shwartz (2012), Puterman (2005). La siguiente función V, se conoce como la función de valor óptimo del criterio de rendimiento descontado que aparece en (1):

$$V(x) = \inf_{\pi \in \Pi} V(x, \pi), \tag{2}$$

donde $x \in X$ y la notación $\inf_{\pi \in \Pi} V(x, \pi)$, es referente al ínfimo sobre el conjunto Π. Adicionalmente, si existe $\pi^* \in \Pi$ tal que $V(x, \pi^*) = V(x)$, a π^* se le llama *política óptima*. Observe que lo anterior en conjunto con la expresión (2) se concluye que la desigualdad $V(x, \pi^*) \leq V(x, \pi)$ ocurre para cualquier $\pi \in \Pi$.

Por lo tanto, en un PDM, se necesitan algoritmos que puedan proporcionar tanto la función de valor óptimo como la política óptima. Esto es lo que nos interesa en el modelo de crecimiento económico y daremos solución vía la metodología de aprendizaje por refuerzo, que se discute en la siguiente sección.

3 Aprendizaje por refuerzo

Para evitar la denominada *doble maldición* de la programación dinámica se tiene la metodología de aprendizaje por refuerzo que provee de soluciones casi óptimas para algunos problemas de control. La herramienta principal que emplea AR es la simulación, que se aprovecha para evadir cálculos intensivos de las probabilidades de transición. Gosavi (2015) muestra al AR como un área de programación dinámica, por lo cual el algoritmo de AR es deducido de su correspondiente parte programación dinámica. A continuación, se discutirán como se usan los Q-factores y el tamaño de paso para resolver un PDM dentro de un simulador.

La función de valor de costo descontado se puede definir por la ecuación que sigue:

$$V^*(i) = \max_{a \in A(i)} \left\{ \sum_{j=1}^{n} p_{ij}(a) \; [r(i,a,j) + \beta V^*(j)] \right\}, \tag{3}$$

donde $i \in X$, $n = |X|$ denota el número de estados en la cadena de Markov; $V^*(i)$ denota el elemento i-ésimo del vector de la función de valor; el conjunto de acciones admisibles en el estado i es denotado por $A(i)$; la probabilidad de transición en un paso del estado i al estado j cuando se aplica la acción a es $p_{ij}(a)$; la recompensa inmediata cuando se selecciona la acción a y el sistema se mueve al estado j, se denota por $r(i,a,j)$. En programación dinámica, dado $i \in X$, se asocia un único componente de la función de valor $V^*(i)$.

Por otra parte, en AR la función de valor del PDM se almacena en los llamados Q-factores. En la metodología de AR es usada una pareja de estado-acción para identificarlo con un vector que se denomina Q-factor. Se define el Q-factor para una pareja de estado-acción $(i,a) \in K$ como sigue

$$Q(i,a) = \sum_{j=1}^{n} p_{ij}(a) \; [r(i,a,j) + \beta V^*(j)]. \tag{4}$$

Combinando (3) y (4), se consigue que

$$V^*(i) = \max_{a \in A(i)} Q(i,a). \quad (5)$$

La expresión en (5) muestra explícitamente que si se conocen los Q-factores, se puede conseguir la función de valor V^*. Empleando la igualdad que proporciona (5), la igualdad (4), se reescribe de la manera siguiente

$$Q(i,a) = \sum_{j=1}^{n} p_{ij}(a)\left[r(i,a,j) + \beta \max_{b \in A(i)} Q\,(i,b)\right], \quad (6)$$

para todo $(i,a) \in K$. La ecuación (6) se entiende como la versión de Q-factor para la ecuación de optimalidad de Bellman asociada al costo descontado del PDM. Los Q-factores se actualizan en cada etapa a lo largo del tiempo, por ello, super-indexamos a Q con el número de la etapa actual. En programación dinámica, Q se actualiza mediante la asignación siguiente:

$$Q^{k+1}(i,a) \leftarrow \sum_{j=1}^{n} p_{ij}(a)\left[r(i,a,j) + \beta \max_{b \in A(j)} Q^k(j,b)\right],$$

para $k = 1, \ldots, k_{max}$, donde k_{max} es el número máximo de iteraciones. Cuando se usa AR, se realiza un cambio en la regla de actualización de Q al aplicar el algoritmo de Robbins-Monro, Gosavi (2015) y se obtiene la siguiente expresión para la actualización de Q que sigue*:*

$$Q^{k+1}(i,a) \leftarrow (1-\alpha_{k+1})Q^k(i,a) + \alpha_{k+1}\left[r(i,a,j) + \beta \max_{b \in A(j)} Q^k(j,b)\right], \quad (7)$$

para cada $(i,a) \in K$.

El procedimiento esencial que provee la solución del problema de optimización a través de Q-learning se refleja en los siguientes pasos.

Los Q-factores se inicializan en 0: $Q(i,a) = 0$, para cualesquiera $i \in X$, $a \in A(i)$. El tamaño de paso es α, fije $k_{máx}$: como el número máximo de iteraciones e i el estado inicial. Tome $k = 1$.

Simular una acción $a \in A(i)$, utilizando la probabilidad $\frac{1}{|A(i)|}$.

El estado siguiente es j, a través de la dinámica de la ecuación en diferencias que modela el capital, ver por ejemplo (9), considerando los ajustes necesarios para obtener un elemento de X. Luego, se obtiene la utilidad $u(i,a,j)$.

Para $(i,a) \in K$, actualice (7), esto es calcule

$$Q^{k+1}(i,a) \leftarrow (1-\alpha)Q^k(i,a) + \alpha[\,u(i,a,j) + \beta\, máx_{b \in A(j)}\, Q^k(j,b)].$$

Actualice $k \leftarrow k+1$. Si $k < k_{máx}$ entonces $i \leftarrow j$ y regrese al segundo paso, en otro caso vaya al último paso.

Para cada $i \in X$, $d(i) = argmax_{b \in A(i)}\, Q(i,b)$, donde d denota la ϵ-óptima política y pare. En este paso $argmax$ denota el argumento máximo.

En Q-learning, la tasa de aprendizaje o tamaño de paso tiene un papel relevante al implementar el algoritmo y que repercute directamente en la solución del problema. Desde el punto de vista teórico es necesario que los tamaños de paso cumplan las siguientes condiciones

$$\sum_{k=1}^{\infty} \alpha^k = \infty \text{ y } \sum_{k=1}^{\infty} (\alpha^k)^2 < \infty,$$

Las cuales aseguran la convergencia al óptimo de las soluciones. Algunas tasas de aprendizaje exploradas en la literatura son: $\alpha = \frac{1}{k}$; $\alpha = \frac{A}{k+B}$ con A y B constantes mayores que cero; y $\alpha = \frac{A}{V(x,a)}$, A es una constante positiva y $V(x,a)$ denota el número de visitas a la pareja estado-acción $(x,a) \in K$, $k = 1,2,3,\ldots$. En Gosavi (2008), se realizaron pruebas con diferentes tasas de aprendizaje, particularmente las mencionadas anteriormente y se reportó que la tasa que produce las mejores soluciones fue $\alpha = \frac{A}{k+B}$, con $A = 100$ y $B = 150$.

4 Un modelo de crecimiento económico

El modelo que será analizado es una economía especializada del trabajo proporcionada por Brock y Mirman (1972) descrita con producción, acumulación de capital y crecimiento estocástico de la productividad. Se asume que producción se provee mediante una función de producción de tipo Coob-Douglas que cuenta con rendimientos constantes a escala con parámetro α;

$$F(K,L) = K^{\alpha}(AL)^{1-\alpha},$$

donde K denota el stock de capital, la oferta de mano de obra es L y A es el parámetro tecnológico de aumento de la mano de obra. Para facilitar el estudio se fija la oferta de mano de obra como 1. Adicionalmente, se asume que el parámetro tecnológico A evoluciona como sigue:

$$log(A_{t+1}) = \kappa + log(A_t) + \sigma Z_{t+1},$$

donde Z es la variable exógena y denota una variable aleatoria normal estándar, esto es denotado por $Z \sim N(0,1)$, $\kappa \geq 0$ denota la tasa media de crecimiento tecnológico y $\sigma \geq 0$ es un parámetro numérico que involucra la estocasticidad en el sistema. Considere a δ como la tasa de depreciación del capital y denote al consumo por C_t. Bajo este contexto, es posible describir la evolución del capital mediante la ecuación:

$$K_{t+1} = A_t^{1-\alpha} K_t^{\alpha} - C_t + (1-\delta)K_t. \quad (8)$$

La estacionariedad de los cocientes $k_t = K_t/A_t$ y $c_t = C_t/A_t$, asociados al capital y tecnología, y al del consumo a la tecnología, respectivamente, permite la escritura del problema en términos de variables estacionarias. Si se normaliza el nivel de tecnología (8) se genera la evolución del capital que sigue:

$$k_{t+1} = \theta Z_{t+1}^{\sigma}(k_t^{\alpha} - c_t + (1-\delta)k_t), \quad (9)$$

donde $\theta = e^{-\kappa}$ y Z^{σ} tiene distribución log-normal.

Un agente representativo de la economía tomará decisiones de consumo separables en el tiempo haciendo uso de la utilidad periódica: $U(C) = \frac{C^{1-\gamma}}{1-\gamma} = \frac{A^{1-\gamma}c^{1-\gamma}}{1-\gamma}$, con $\gamma \in (0,1)$. El problema del planificador social se basa en la elección de una sucesión de consumo a lo largo del tiempo que logre maximizar la utilidad descontada esperada del agente. De esta manera, se debe resolver:

$$sup_{C_t} E \sum_{t=0}^{\infty} \beta^t U(c_t),$$

sujeto a (9).

En Williams (2004) se presentan las condiciones bajo las cuales este problema de optimización tiene una solución y además muestra que la solución es un control de retroalimentación de la forma $c_t = c^\sigma(k_t)$. Por lo tanto, es posible reescribir la evolución del capital mediante la expresión:

$$k_{t+1} = \theta\, Z_{t+1}^\sigma(\, k_t^\alpha \, - c_t \, + \, (1-\delta)k_t) \equiv \overline{f}^\sigma(\, Z_{t+1}^\sigma, k_t, c^\sigma), \qquad (10)$$

donde la última notación exhibe la dependencia de la política de consumo c^σ (desconocida). La política que proporciona la solución al problema estocástico verifica la ecuación estocástica de Euler:

$$c^\sigma(k)^{-\gamma} = \beta \int (\theta Z^\sigma)^\gamma\, c^\sigma\left(\overline{f}^\sigma(Z^\sigma, k, c^\sigma)\right)^{-\gamma}\left[\alpha \overline{f}^\sigma(Z^\sigma, k, c^\sigma)^{\alpha-1} + 1 - \delta\,\right] dG^\sigma(Z^\sigma), \qquad (11)$$

donde G^σ es la función de distribución log-normal.

A la vez será utilizado el modelo determinista asociado al sistema aleatorio que se encuentra caracterizado por la ecuación en diferencias que modela el capital y que está dada por:

$$k_{t+1} = \theta(\, k_t^\alpha \, - c_t^0 + (1-\delta)k_t) = \overline{f}^0(k_t)^2,$$

que se obtiene al sustituir σ por 0 en la dinámica expresada en (10). Cálculos no tan complicados muestran que la solución para el sistema determinista es un control de retroalimentación $c_t = c^0(k_t)$. En este contexto, se verifica que la política de consumo óptimo satisface la ecuación de Euler:

$$c^0(k)^{-\gamma} = \beta\theta^\gamma c^0\left(\overline{f}^0(\, k, c^0)\right)^{-\gamma}\left[\,\alpha\, \overline{f}^0(\, k, c^0)^{\alpha-1} \, + 1 - \delta\right], \qquad (12)$$

la cual resulta ser similar a (11). Finalmente, se deduce una expresión analítica que caracteriza el estado estable del sistema determinista. Inicialmente, sustituya k^* en lugar de k, en la ecuación de Euler localizada en la igualdad (12), de lo cual se obtiene que

$$c^0(k^*)^{-\gamma} = \beta\theta^\gamma\, c^0(k^*)^{-\gamma}[\,\alpha\,(k^*)^{\alpha-1} \, + 1 - \delta]. \qquad (13)$$

Después, despeje a k^*de la igualdad anterior. Por lo tanto, el único estado estable del sistema determinista que resulta ser interior es:

$$k^* = \left(\frac{1 + (\delta \, - 1)\beta\theta^\gamma}{\alpha\, \beta\theta^\gamma}\right)^{\frac{1}{\alpha-1}}. \qquad (14)$$

Finalmente, al trabajar con el modelo determinista, si establece $k = k^*$ en (13) y se despeja al consumo $c^0(k^*)$, se consigue que

$$c^0(k^*) = (k^*)^\alpha + \frac{(\theta(1-\delta)-1)}{\theta} k^*.$$

La igualdad anterior en conjunto con la representación del estado estable determinista descrito en (14) serán suficientes para comparar el modelo aleatorio con el modelo determinista cuando el ruido es pequeño, i.e. $\sigma \to 0$. La comparación entre ambos modelos: perturbado y determinista se realizará mediante simulación. A continuación, se identifica el problema de crecimiento económico estocástico con las componentes de un PDM estocástico discreto.

Se propone que el espacio de estados, denotado por X_C sea el intervalo cerrado $[0, K]$, el espacio de acciones será $A_C = [0, K]$, donde $K > 0$ es una constante fija. El conjunto de acciones admisibles será $A_C(k) = [0, k]$, para cualquier estado $k \in X$. La función

de utilidad $u(c)$, tiene regla de correspondencia $\frac{c^{1-\gamma}}{1-\gamma}$ para cada $c \in A_C(k)$ y $k \in X_C$, por lo tanto c depende del estado k. La ley de transición Q_l es inducida por la ecuación en diferencias descrita por (9). En resumen, el PDM discreto con las componentes $(X_C, A_C, \{A_C(k) \mid k \in X\}, Q_l, u)$ será denotado por $\mathcal{M}$ y será nuestra base para la implementación numérica tanto de PD como AR.

El modelo estocástico original es continuo, ya que tanto el espacio de estados como acciones son intervalos, de modo que se discretizan las componentes de la manera que sigue: el conjunto de estados será $X = \{k_0, k_1, \ldots, k_n\}$, donde $k_0 < k_1 < \cdots < k_n$ con $k_0 = 0$ y $k_n = K$, $n \in \mathbb{N}$ fijo; $A = \{c_0, c_1, \ldots, c_m\}$ es el conjunto de acciones, con $c_0 < c_1 < \cdots < c_m$ donde $c_0 = 0$ y $c_m = K$, $m \in \mathbb{N}$ fijo: $A(k) = \{0, c_1, \ldots, c_j\}$, con $c_j = k$, será el conjunto de acciones admisibles cuando el estado actual es $k \in X$. La función de utilidad será $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$, $\gamma \in (0,1)$ con $c \in A(k)$ y $k \in X$.

Las componentes faltantes para el PDM son probabilidades de transición controladas: $p_{kl}(c)$, que representan la probabilidad de que el estado actual sea k se usa la acción c y como consecuencia se transitar al estado l, para $k, l \in X$ cualesquiera y $c \in A(k)$. En el contexto del modelo de crecimiento económico estocástico no contamos de manera explícita con la matriz de transición, aunque se sabe que de manera teórica está inducida por la ecuación en diferencias (9), este el nuestro motivo principal para aplicar Q-learning. Con el enfoque de AR, únicamente será necesario implementar realización del proceso estocástico determinado en (9) una cantidad alta de veces, promediar y poner a prueba el simulador.

5 Experimentos numéricos

Los experimentos desarrollados requieren la discretización de las componentes del modelo presentado en la sección anterior. Los experimentos numéricos consisten en determinar la solución del modelo de crecimiento económico, esta solución consta de la política óptima y de la función de costo. Específicamente, se exhibe el comportamiento de la trayectoria óptima del proceso estocástico cuando el estado inicial de la dinámica (9) corresponde al estado estable del sistema determinista en conjunto de la condición de que el ruido converge a cero, i.e. $\sigma \to 0$.

Los procedimientos desarrollados que permiten encontrar una política casi-óptima y generar la trayectoria optima del modelo de crecimiento económico estocástico se muestran en los Algoritmos 1 y 2. Estos se implementaron en el lenguaje de programación de R.

Algoritmo 1: *Función principal*

Datos: $\alpha, \beta, \delta, \gamma, \kappa, max.ite, \Delta_e, \Delta_a, k_0, \sigma$
Resultado: Capital promedio, trayectoria óptima del Capital
Política$= GenerarPolitica(\alpha, \beta, \delta, \gamma, \kappa, max.ite, \Delta_e, \Delta_a, k_0, \sigma)$;
N=length(Política);

for $1 \leq j < M$ ***do***
$T = matrix(0, N, M)$;
$T[1, j] = k_0$;
$T[2, j] = e^{-\sigma rnorm(1)-\kappa}(k_0^{\alpha} - Política[1] + (1-\delta)k_0)$;
for $2 \leq i < N$ ***do***
$T[i+1, j] = e^{-\sigma rnorm(1)-\kappa}(k_0^{\alpha} - Política[i] + (1-\delta)k_0)$;
end for
end for
for $1 \leq i < N$ ***do***
Ta[i]=mean(T[i]);
end for
plot(Ta): genera la trayectoria óptima
print(mean(Ta)): muestra el capital promedio

Algoritmo 2: *Generar Política*
Datos: $\alpha, \beta, \delta, \gamma, \kappa, max.ite, \Delta_e, \Delta_a, k_0, \sigma$
Resultado: Política óptima
Política$= GenerarPolitica(\alpha, \beta, \delta, \gamma, \kappa, max.ite, \Delta_e, \Delta_a, k_0, \sigma)$;
$r(k, c) = \frac{c^{1-\gamma}}{1-\gamma}$*: función de utilidad:*
for $1 \leq j < M$ ***do***
$T = matrix(0, 2, max.ite)$;
$T = Trayectoria(100{,}000, \alpha, \sigma, \kappa, \delta, \gamma, k_0)$; *simula una trayectoria del capital.*
$Q = matrix(0, N_e + 1, N_a + 1), N_e$*: número de estados,* N_a *número de acciones* ;
$\alpha_t = \frac{150}{300}$;
$i = \frac{trunc(k_0) + 1}{\Delta_e}$
$p = array\left(\frac{1}{i+1}, i+1\right)$
$a = sample(c(0{:}\, i), 1, replace = TRUE, prob = p$): *genera un número entre* 0 *e* i*;*
$j = \frac{trunc\,(e^{-\sigma rnorm(1)-\kappa}(k_0^{\alpha} - a\Delta_a + (1-\delta)k_0 + 1))}{\Delta_e}$;
$Q[i, a] \leftarrow (1-\alpha_t)Q[i, a] + \alpha_t(r(i\Delta_e, a\Delta_a) + \beta\ máx(Q[j,])$ *);*
for $2 \leq s < max.ite$ ***do***
$i = j$;
$p = array\left(\frac{1}{i+1}, i+1\right)$
$a = sample(c(0{:}\, i), 1, replace = TRUE, prob = p$)
$j = \frac{trunc\,(e^{-\sigma rnorm(1)-\kappa}(k_0^{\alpha} - a\Delta_a + (1-\delta)k_0 + 1))}{\Delta_e}$;
$\alpha_t = \frac{150}{s+300}$;
$Q[i, a] \leftarrow (1-\alpha_t)Q[i, a] + \alpha_t(r(i\Delta_e, a\Delta_a) + \beta\ máx(Q[j,])$ *);*
end for
for $1 \leq s < max.ite$ ***do***
$pol[s] = which.\max(Q[s,])$;
end for

Return pol: genera la política óptima

El Algoritmo 1 está estructurado como una función principal que tiene como salidas a la trayectoria óptima del proceso y el capital promedio. En el algoritmo 2, se muestra el procedimiento numérico que genera el consumo óptimo (política casi óptima). Dentro de este algoritmo tiene lugar la implementación del procedimiento Q-learning. También, es aquí donde se simula la trayectoria óptima del proceso estocástico que modela el capital. La trayectoria óptima se consigue al aplicar la sucesión de consumo en la ecuación que modela el capital, ver ecuación (9). Además, el capital promedio se obtuvo repitiendo este procedimiento una cantidad finita de veces y se promedió el valor del capital sobre las repeticiones elaboradas. Los resultados numéricos fueron obtenidos con los estimadores descritos en la Tabla 1. Estos estimadores son los que produjo el estudio de Williams (2004) el cual se basó en datos reales. La variante es una actualización del valor del parámetro γ, ya que considerar el valor original en la investigación se tiene como resultado que las funciones de costo serán negativas y en esta economía no existe interpretación para costos negativos.

Tabla 1. Valores numéricos de los parámetros del modelo

Parámetro	Valor numérico
α	0.6500
β	0.7638
γ	0.4500
δ	0.0517
κ	0.0176

Observe que a través de los estimadores para los parámetros del modelo que se describen en la Tabla 1, la expresión (14) produce el estado estable del sistema determinista $k^* = 4.950451$.

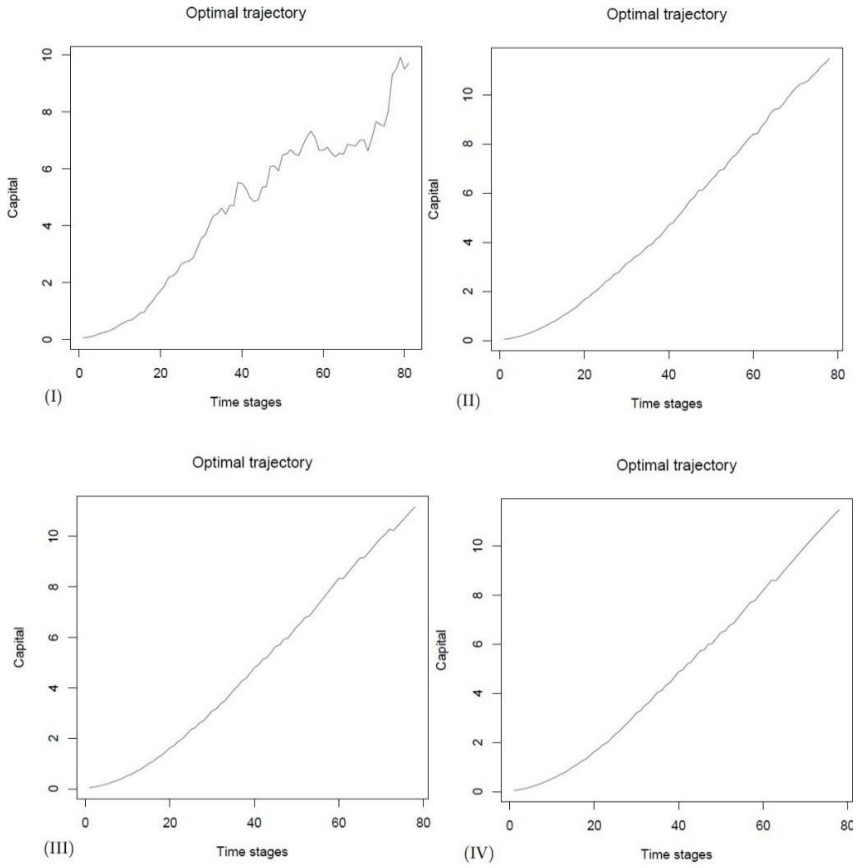


Figura 1. Trayectoria óptima del sistema estocástico cuando: (I) $\sigma = 0.00492$, (II) $\sigma = 0.000492$, (III) $\sigma = 0.0000492$, (IV) $\sigma = 0.00000492$.

La Figura 1 presenta algunas realizaciones de la trayectoria óptima del proceso asociado al modelo de crecimiento económico estocástico, el cual se desarrolla por la ecuación en diferencias (9), el periodo de observación fue de 300 unidades de tiempo. La implementación de los algoritmos 1 y 2 reflejan aproximaciones del capital de 4.540348, 4.915716, 4.853004 y 4.921873, para las representaciones (I)-(IV), respectivamente.

Para generar las gráficas se estableció lo siguiente: iteración máxima igual a 1000, incrementos de 0.4 unidades para la discretización de los espacios de estados acciones, se implementaron 1000 trayectorias para promediar, con ello se construyó la trayectoria óptima promedio.

Adicionalmente, en la Figura 1 se observa que conforme el ruido σ converge a cero, la trayectoria óptima del sistema estocástico tiende a comportarse de manera determinista. Esto también es visible en los parámetros que intervienen en el grado de estocasticidad del modelo, ya que las componentes aleatorias del sistema convergen a cero. Con los

experimentos desarrollados se puede observar que cuando σ converge a cero, la sucesión de consumos (política casi óptima) que proporcionó Q-learning nos provee de un capital promedio aproximado al valor del estado estable estacionario del sistema determinista, que es congruente con lo establecido en la teoría.

6 Conclusiones

En este trabajo se aprovechó la perspectiva de PDMs para replantear y proporcionar soluciones casi óptimas a un modelo de crecimiento económico, lo cual permite determinar soluciones aproximadas a tal modelo. Con el enfoque abordado, además de la garantía de que existe un único estado estacionario en el modelo determinista asociado, se muestra que conforme la desviación estándar σ del proceso de choque tecnológico converge a cero, el proceso de stock del capital modelado convergerá al único estado estacionario del sistema determinista asociado. La validación de la convergencia se realiza mediante experimentos numéricos, en los cuales fueron retomados valores estimados para los parámetros del modelo proporcionados por Williams (2004), y como resultado se obtiene que, en efecto, capital (en promedio) del sistema estocástico se aproxima al valor del estado estable del sistema determinista. La solución casi óptima que Q-learning provee es utilizada para generar simulaciones de la trayectoria óptima del capital promedio, la cual nos permite visualizar posibles situaciones del desarrollo del capital, bajo los parámetros antes fijados. Algunos trabajos futuros, consideran implementar Q-learning en modelos de control más sofisticados y ejecutarlo con datos totalmente reales.

Referencias

Bellman, R. (2003). Dynamic Programming. Dover.

Bertsekas, D. P., y Shreve, S. E. (1996). Stochastic Optimal Control: The Discrete-Time Case. Athena Scientic.

Boucherie, R. J., y Van Dijk, N. M. (2017). Markov Decision Processes in Practice. Springer.

Brock, W. A., y Mirman, L. J. (1972). "Optimal economic growth and uncertainty: the discounted case", *Journal of Economic Theory*, vol. 4(3), pp. 479-513.

Cao, X. R. (2021). Foundations of average-cost nonhomogeneous controlled Markov chains. Springer International Publishing.

Cruz-Suárez, H., Gordienko, E., y Montes-de-Oca, R. (2009). "A note on deterministic approximation of discounted Markov decision processes", *Applied mathematics letters*, vol. 22(8), pp. 1252-1256.

Capítulo 9

Expectativas de los estudiantes de Ingeniería en Ciencia de Datos

Jaime Alejandro Romero Sierra, Daniel Marcelo González Arriaga, Carlos Leopoldo Carreón Díaz de León, Juan Carlos Conde Ramírez

Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla, México

jaime.romeros@correo.buap.mx, daniel.gonzalezarr@correo.buap.mx, carlos.carreon@correo.buap.mx, carlos.conder@correo.buap.mx

Resumen. En el presente trabajo se aborda como la ciencia de datos se ha posicionado como una de las disciplinas más relevantes dentro de esta era digital, convirtiendo los paradigmas que las empresas, organizaciones y sociedades, al hacer uso de grandes volúmenes de datos, y transformándolos en conocimiento utilizable. Abordando las expectativas de los estudiantes que ingresan a la ingeniería en ciencia de datos, desde la elección de esta carrera no solo por las amplias oportunidades laborables y alta demanda de la misma, sino que, también es una de las pocas carreras que puede generar un impacto social en las diferentes ramas de estudio, pudiendo impactar desde un sistema de salud, hasta la creación de soluciones sostenibles. A través de diferentes análisis, encuestas y metodologías, se buscan comprender las motivaciones y preocupaciones de las nuevas generaciones, que buscan ingresar a esta carrera y ser parte de la transformación digital del futuro

Palabras Clave: Ciencia, Datos, Motivación, Expectativas, Digital.

1 Introducción

Actualmente vivimos en una era digital, como parte de la actualización de los conocimientos ha emergido la ciencia de datos, esta disciplina la podemos contemplar como una de las ciencias transformadoras y relevantes de estos tiempos, por que, con todos los conocimientos, redefine la forma de tomar decisiones tanto de las empresas, organizaciones y sociedades, recopilando datos y haciendo un análisis de los mismos. Haciendo sinergia de procesos como: el procesamiento, análisis y extracción de conocimientos, a partir de grandes volúmenes de datos, ha convertido en un pilar fundamental esta disciplina para el avance de la tecnología, economía y la sociedad, logrando posicionar a la ciencia de datos, como una de las carreras emergentes con un gran potencial de crecimiento. No sólo enfocándose en el área de la computación, si no extendiéndose su aplicación a diferentes ramas como lo son: salud, biología, economía, educación y finanzas, logrando un impacto en la forma de abordar las problemáticas y generando soluciones innovadoras.

Este trabajo se enfoca a explorar como la ciencia de datos se ha ido posicionando como una de las carreras para las nuevas generaciones, con la característica de ser atractiva y presentar un amplio campo laboral lleno de oportunidades. Algunas de las razones por las que los estudiantes eligen esta carrera no solo tienen que ver con las mencionadas anteriormente, si no que, el potencial que ven en la carrera para generar cambios significativos en la sociedad, generando impactos en todos los ámbitos profesionales. Abarcando desde la optimización del sector salud hasta temas relacionados con la sostenibilidad que busquen el cuidado del medio ambiente, la ciencia de datos busca diversificar su campo de acción, buscando una sinergia entre el conocimiento técnico y la posibilidad de contribuir a bienes en común.

En esta era digital, la ciencia de datos ha ido emergiendo como una de las disciplinas base, permitiendo que las organizaciones, tomen en cuenta toda su información agrupándola en grandes volúmenes de datos y transformándola en conocimiento que conlleve a la acción. Se han logrado avances significativos en el sector salud, finanzas, biología y comercio electrónico, logrando implementar el análisis de datos que generan una optimización de procesos y produce una mejora en la toma de decisiones (Universidad San Sebastián, 2023).

La relevancia de la ciencia de datos ha ido creciendo en los últimos años, tanto así que estudiantes de carreras de diferentes áreas, se han tomado el tiempo de introducirla en sus estudios. La ingeniería en ciencia de datos es una opción atractiva por su alta demanda laboral y su capacidad de aplicarse en una variedad de ramas dentro del ámbito profesional (Politécnico Grancolombiano, 2023). Además, de que esta diciplina debido a su enfoque, permite que los profesionistas aborden problemáticas complejas, auxiliándose de herramientas computacionales, matemáticas y estadística, se llegan a soluciones innovadoras que generan un impacto positivo en las decisiones (Universidad de Ingeniería y Tecnología, 2023).

Dentro de lo que las motivaciones que tienen los estudiantes para elegir esta carrera no sólo se limitan al campo laboral. Muchos de estos estudiantes buscan contribuir a la sociedad para generar un bienestar aplicando las técnicas de ciencia de datos en las áreas de educación, medio ambiente y salud (Centro Competencia, 2023). Esto genera una conciencia creciente sobre el potencial que nos puede generar la ciencia de datos para generar un cambio significativo dentro de la sociedad.

Sin embargo, no solamente debemos ver como una oportunidad, sino que también surgen los desafíos éticos, estos se relacionan con la privacidad y la responsabilidad del uso de la información que estamos ocupando. Es primordial que las personas que se interesen en el estudio de esta profesión, estén preparadas para afrontar estas interrogantes y promuevan una práctica ética en el uso de estas tecnologías (Data Universe, 2023).

En este estudio nos auxiliamos de análisis, encuestas y metodologías que buscan comprender las motivaciones, expectativas y las preocupaciones de los estudiantes que se

arriesgaron a estudiar esta nueva carrera y decidieron adentrarse en el mundo de la ciencia de datos. Unas de las preguntas clave en la cuales nos vamos a enfocar, son: ¿Qué los impulsa a elegir esta carrera? ¿Qué esperan lograr al formarse como científicos de datos?, al responder estas interrogantes, nosotros podemos comprender cuál es el perfil de las nuevas generaciones que buscan adentrarse a esta carrera, y así, identificar qué rasgos necesidades y aspiraciones nosotros podíamos enfocarnos en los programas educativos para satisfacer las necesidades en la industria.

Por lo consiguiente en este artículo, se enfocará en analizar el papel de la ciencia de datos, como un agente de transformación en este futuro digital, vamos a destacar la relevancia de cómo se aborda como un área multidisciplinar, En la cual como principal objetivo busca un impacto positivo en las diversas áreas del conocimiento en las cuales se podría aplicar. Como rasgo esencial se busca una reflexión de cómo las expectativas de los estudiantes que ingresan a la carrera de ciencia de datos pueden repercutir en su desarrollo profesional así también en cómo se van desenvolviendo en esta disciplina. Finalmente, con un análisis se espera generar una contribución a la comprensión de los desafíos y oportunidades que enfrenta la ciencia de datos en este mundo impulsado cada vez más por la tecnología y la información

2 Preliminares

En el siglo XXI una de las disciplinas que han sido las más transformadoras y relevantes dentro del ámbito de la computación es la ciencia de datos, A raíz de que la era digital se empezaron a ocupar grandes volúmenes de datos. Según Smith y Johnson (2020), No menciona que el volumen de datos que se produce a nivel global se duplica cada 2 años, este es un fenómeno que como humanidad nos ha creado la necesidad de desarrollar técnicas y herramientas para aprovechar estos datos y analizarlos de manera de que nos den conclusiones significativas de los mismos. A este fenómeno de crecimiento exponencial le hemos llamado el Big Data, Con el cual organizaciones gobiernos y sociedades ante redefinido la forma en que toman sus decisiones, generando un impacto significativo en varios sectores como es el de la salud, la educación, las finanzas y el del desarrollo sustentable (García et al., 2019).

Los orígenes de la ciencia de datos, parten de disciplinas como: la estadística, las matemáticas aplicadas y la informática, sin embargo, esta evolución ha sido marcada por la integración de áreas como la inteligencia artificial, el machine learning y la visualización de datos (Brown, 2021). En 1960 John Tuckey presentó las bases de lo que es el análisis exploratorio de datos remarcando que el uso de técnicas estadísticas es esencial para descubrir patrones y tendencias que nos ayudan a ver el comportamiento de datos complejos (Tukey, 1977). Desde ese entonces, ha sido un parteaguas para la disciplina, ya que, a partir

de este momento, se ha registrado un crecimiento acelerado especialmente con la integración del internet y la de la digitalización en la década de los noventas.

Actualmente la ciencia de datos se ha estado consolidando, ya que se ve como una herramienta indispensable la cual ocupamos para tomar decisiones basándonos en evidencias encontradas en patrones de la visualización de datos. Un estudio en la universidad de Stanford nos indica que el 85% de las empresas de la revista Fortune 500, se han visto en la necesidad de adoptar herramientas que involucren la ciencia de datos para optimizar sus procesos mejorar la competitividad de las mismas y el desarrollo de productos innovadores (Lee & Martínez, 2018). Ahora no solamente pasa esto en el sector privado, sino que también en el sector gubernamental y en organizaciones sin fines de lucro las cuales todas han empezado a utilizar la ciencia de datos para abordar estos desafíos sociales como lo es reducir la pobreza buscar la optimización de los sistemas de salud y abordar la mejora del cambio climático (UNESCO, 2022).

En el panorama educativo con lo que respecta a la ciencia de datos esta ha ido ganando popularidad entre los estudiantes debido a que gracias a su potencial y su versatilidad ha tenido un impacto social. Un informe de la UNESCO (2022), nos dice que la disciplina de ciencia de datos, no solo nos ofrece oportunidades laborales, sino que también nos permite visualizar qué desafíos nos encontramos en un panorama global y resolverlo de manera innovadora. Por ejemplo, en el sector salud al analizar estos grandes volúmenes de datos ha permitido la capacidad de predecir brotes de enfermedades, optimizar la distribución de recursos y generar una personalización de tratamientos médicos (Pérez & González, 2021). Otra de las ramas estudiadas es en la parte ambiental, la ciencia de datos nos permite el monitoreo del cambio climático, haciendo modelos para desarrollar una predicción del clima y también diseñar soluciones para el desarrollo sostenible de la sociedad (García et al., 2019).

Por otra parte, este rápido crecimiento del área de la ciencia de datos también produce un desafío significativo, siendo este que la brecha de habilidades tecnológicas sea más evidente dentro de la demanda profesional, ya que personas capacitadas en esta disciplina van a superar con creces a la oferta actual que hay en otras áreas (Brown, 2021). Además, una de las preocupaciones relacionadas con esta parte de la ciencia de datos, es las implicaciones éticas del uso de los datos, ya que la privacidad, la transparencia y el sesgo de los algoritmos representan una importante temática de estudio. Según Pérez y González (2021), es importante que, dentro de los programas académicos, no solo se busque formar a los estudiantes en aspectos técnicos o matemáticos, sino que también, se les brinden herramientas para llegar a una comprensión profunda de los principios éticos y de las implicaciones sociales que traen sus trabajos al utilizar la ciencia de datos.

Además, la utilización de la ciencia de datos, permite el desarrollo de plataformas educativas las cuales utilizan análisis predictivos para identificar, qué estudiantes tienen un

riesgo de deserción, una baja de nivel de desarrollo cognitivo dentro de las materias y la eficiencia de los profesores que imparten las materias. Según Chassignol et al. (2018), el uso de estas herramientas, nos han ayudado a mejorar la efectividad para intervenir de manera oportuna y así poder ofrecer un apoyo adicional a quienes lo necesitan, así como también, al utilizar estos datos nos facilita la evaluación continua de los programas educativos, las instituciones pueden identificar áreas de mejora y así adoptar estrategias pedagógicas para los estudiantes en la demanda laboral en un mundo global (Dickson, 2017).

La ciencia de datos gracias a su habilidad de analizar grandes volúmenes de datos, permite analizar la interacción entre plataformas de aprendizaje en línea, accediendo a comprender qué patrones de comportamiento y preferencias de los usuarios, nos permiten aprender de una manera más eficiente (Akharraz et al., 2018). Ahora con esto no solo se puede mejorar el diseño de los cursos, sino que también, nos proporcionan insights valiosos para tomar decisiones en cómo implementar de manera eficiente políticas educativas.

Además de que hay una preocupación en especial a la privacidad y la seguridad de los datos de los estudiantes, ya que, al usar nuestras herramientas de ciencia de datos implica que recopilamos y analizamos información que puede ser sensible (Chassignol et al., 2018). Una de las maneras más eficientes para abordar estos desafíos es establecer Marcos regulatorios que sean robustos y promover prácticas éticas para el manejo de estos datos.

Una de las tendencias más relevantes dentro de este sector, es que los profesionales en ciencia de datos sean cada vez más, lo que ha llevado que diferentes escuelas partan a una creación de programas académicos que se especialicen en esta área, en las universidades a lo largo del mundo. Según un informe de la UNESCO (2022), al formar un científico de datos no solo nos debemos de enfocar en el aspecto cognitivo o técnico, como temas en programación o análisis estadísticos sino que también, nos debemos enfocar en habilidades blandas como lo es el pensamiento crítico, la comunicación efectiva y la ética en el uso de datos, esto es muy importante ya que en el contexto actual donde los algoritmos y los modelos predictivos pueden generar un impacto directo en la vida de las personas, se genera una responsabilidad en el estudiante para aplicar sus conocimientos.

3 Metodología

Para realizar la presente investigación se tomará un análisis cuantitativo y cualitativo, ya que se pretende que a través de encuestas recopilar datos, teniendo estas como objetivo comprender que motivó al alumno a estudiar la carrera de ciencia de datos, que expectativas y preocupaciones tuvieron al momento de elegir estudiar la carrera. Para esto actualmente en la primera generación se contaba con 60 alumnos, de los cuales para poder hacer un

estudio estadísticamente relevante se aplico la encuesta a 52 alumnos como se calculó por medio de la formula para poblaciones finitas, y obtener un resultado representativo.

El instrumento de recolección de datos, se diseño una encuesta la cual consta de preguntas cerradas para permitir un análisis que no fuera tan disperso, algunas preguntas se ocuparon la escala Likert, enfocada a recopilar la información sobre las motivaciones expectativas y preocupaciones de los estudiantes al ingresar a la carrera. La encuesta se distribuyo de manera digital garantizando que será de manera anónima.

Las preguntas realizadas fueron:

Motivaciones:
¿Qué te motivó a elegir la carrera de Ingeniería en Ciencia de Datos?
a) Oportunidades laborales.
b) Interés en la tecnología y el análisis de datos.
c) Impacto social de la disciplina.
d) La posibilidad de trabajar en proyectos innovadores y multidisciplinarios.

Expectativas laborales:
¿Cuál es tu expectativa salarial al graduarte?
a) Menos de $8,000 al mes
b) Entre $8,000 a 10,000
c) Más de $10,000

Preocupaciones:
¿Cuál es tu mayor preocupación al estudiar esta carrera?
a) Dificultad académica.
b) Falta de experiencia práctica.
c) Competencia en el mercado laboral.
d) La necesidad de actualizarse constantemente en tecnologías emergentes.

Impacto social:
¿Qué tan importante consideras que es el impacto social de la ciencia de datos?
Escala Likert: 1 (Nada importante) a 5 (Muy importante).

Habilidades técnicas:
¿Qué habilidad técnica consideras más importante para un científico de datos?
a) Programación (Python, R).
b) Análisis estadístico.
c) Visualización de datos.
d) Aprendizaje automático.

Formación académica:

¿Consideras que tu formación académica actual te prepara adecuadamente para el mercado laboral?
Escala Likert: 1 (Muy en desacuerdo) a 5 (Muy de acuerdo).

Áreas de interés:
¿En qué área te gustaría especializarte?
a) Salud.
b) Finanzas.
c) Medio ambiente.
d) Tecnología.

Experiencia previa:
¿Tienes experiencia previa en programación o análisis de datos?
a) Sí.
b) No.

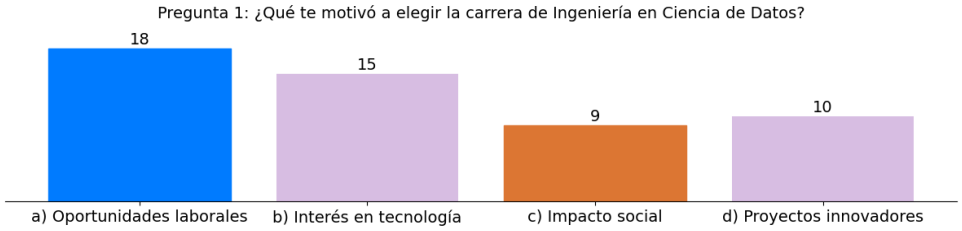
Expectativas de la carrera:
¿Qué esperas lograr al graduarte de esta carrera?
a) Conseguir un empleo bien remunerado.
b) Contribuir a soluciones sociales.
c) Desarrollar habilidades técnicas avanzadas.
d) Emprender y crear mi propia startup basada en ciencia de datos

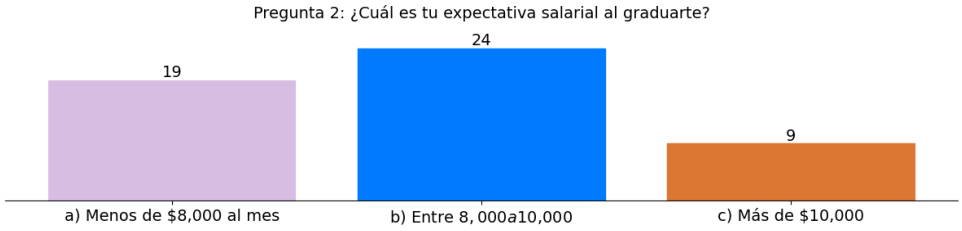
Desafíos:
¿Qué desafíos crees que enfrentarás como científico de datos?
a) Manejo de grandes volúmenes de datos.
b) Actualización constante de conocimientos.
c) La comunicación efectiva de lo que los datos nos quieren contar
d) La ética en el uso de datos y la privacidad de la información.

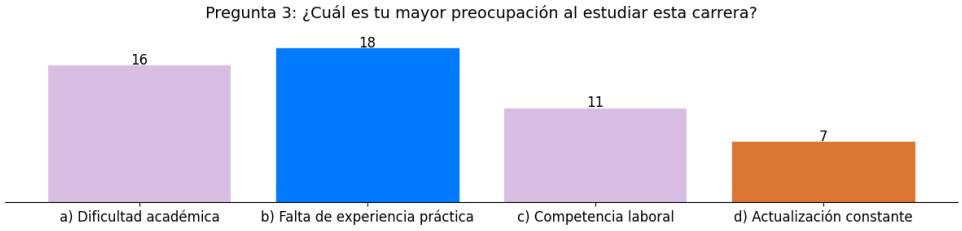
Para el análisis, el estudio se auxilio de técnicas de ciencias de datos, aplicando técnicas de estadística descriptiva e inferencial. Se utilizo el lenguaje de programación Python para procesar la información y generar las visualizaciones necesarias.

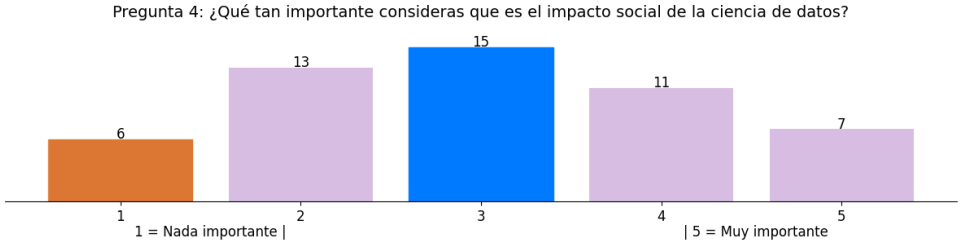
4 Resultados

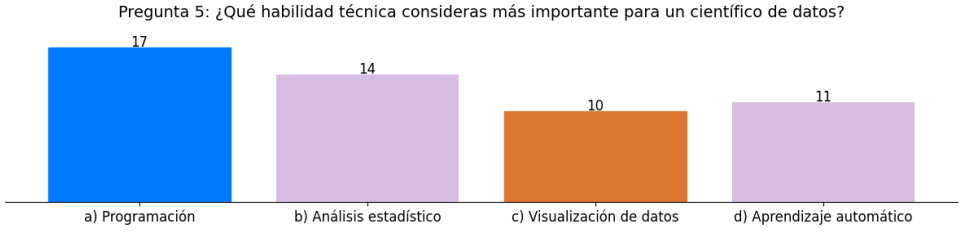
Los resultados obtenidos a través de la encuesta aplicada a 52 estudiantes de la carrera de Ingeniería en Ciencia de Datos revelan tendencias significativas en cuanto a sus motivaciones, expectativas y preocupaciones. A continuación, se presentan los hallazgos más relevantes, organizados en torno a los principales temas de investigación.

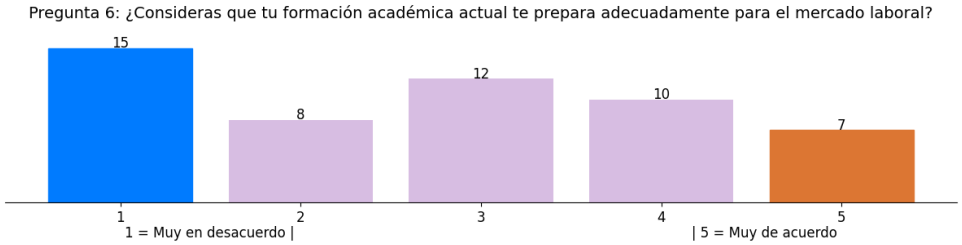

Pregunta 1: ¿Qué te motivó a elegir la carrera de Ingeniería en Ciencia de Datos?
18
15
9
10
a) Oportunidades laborales
b) Interés en tecnología
c) Impacto social
d) Proyectos innovadores

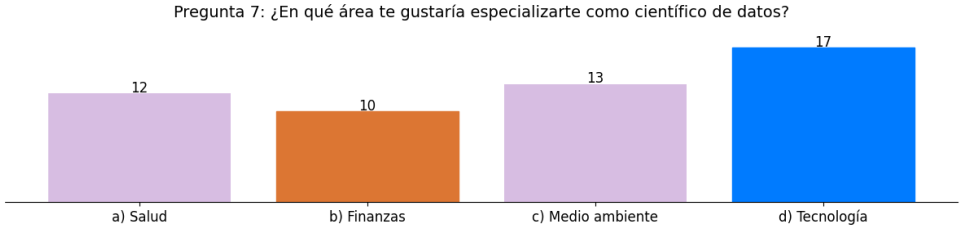

Pregunta 2: ¿Cuál es tu expectativa salarial al graduarte?
19
24
9
a) Menos de $8,000 al mes
b) Entre 8,000a10,000
c) Más de $10,000

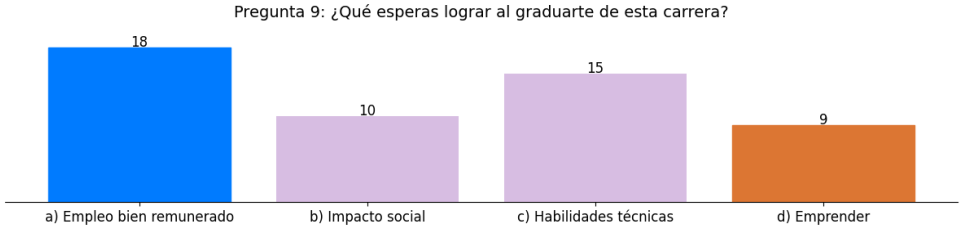

Pregunta 3: ¿Cuál es tu mayor preocupación al estudiar esta carrera?
16
18
11
7
a) Dificultad académica
b) Falta de experiencia práctica
c) Competencia laboral
d) Actualización constante

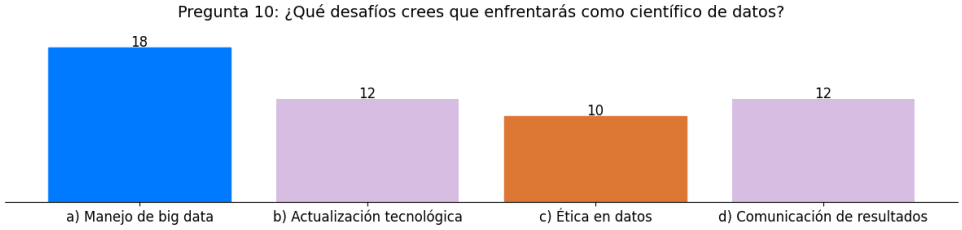

Pregunta 4: ¿Qué tan importante consideras que es el impacto social de la ciencia de datos?
6
13
15
11
7
1
2
3
4
5
1 = Nada importante |
| 5 = Muy importante


Pregunta 5: ¿Qué habilidad técnica consideras más importante para un científico de datos?
17
14
10
11
a) Programación
b) Análisis estadístico
c) Visualización de datos
d) Aprendizaje automático









Ahora dentro de los hallazgos desarrollados en las preguntas, nos encontramos que los estudiantes que estudian ciencia de datos, se enfocan más en una perspectiva económica y tecnológica, para ellos es importante priorizar las oportunidades en el campo laboral sobre

un impacto social. Las expectativas salariales no son tan altas como se podría llegar a pensar, pero como cualquier otra área lo preocupante es llegar a una estabilidad financiera.

Una de las mayores preocupaciones de los mismos, es que la falta de experiencia práctica puede ser, una de las características que más impacten en sus estudios sugiriendo que para ellos, la formación teórica no es suficiente para enfrentar el campo laboral existente. Además, la percepción de que la educación actual, los debe preparar para enfrentarse a las adversidades que se enfrentan en un mundo profesional, lo cual, los hace pensar que es importante que sus conocimientos se vinculan con la industria.

En cuanto al ámbito de las habilidades, los estudiantes observan que la programación y el análisis estadístico son temas fundamentales para su desarrollo profesional, sin embargo, al ser una ingeniería se esperaría que ingresaran con estos conocimientos, pero la realidad es que, muchos de ellos no cuentan con experiencia previa en estas áreas, lo cual implica que se debe abordar generar una base formativa sólida en estos temas en los primeros semestres.

Cabe destacar que hay un interés importante en especializarse en ciencia de datos, en cuanto a las áreas de tecnología y medio ambiente, ya que la ciencia de datos nos ayuda a destacar en estos sectores innovadores. Así como también, el manejo de grandes volúmenes de información llamados el Big data y actualizarse constantemente, parece ser uno de los mayores desafíos que esperan encontrarse estos estudiantes en su futuro profesional.

5 Conclusiones

Como docentes, nosotros tenemos la responsabilidad de enfocarnos en un aprendizaje práctico, tratando de incorporar proyectos reales que simulen su vida laboral en su campo profesional. Uno de los aspectos importantes es enseñar la programación y estadística en los primeros semestres, asegurando que nuestros estudiantes se interesen en el aprendizaje de estos temas como base para asignaturas más especializadas. Además, de que nos debemos enfocar en la formación de habilidades blandas como es: comunicar resultados y la ética en ciencias de datos, estas áreas suelen no tomarse en cuenta en ingeniería, pero son muy esenciales para ejercer su profesión

Finalmente, uno de los rasgos fundamentales que nos ayudarían a desarrollar profesionales en ciencias de datos, es implementar actividades con el sector industrial a través de sus prácticas profesionales, proyectos integradores, así como también concursos, que les ayuden a nuestros estudiantes a probar sus habilidades en el uso de la ciencia de datos. Con todo esto, nosotros podemos preparar a mejores científicos de datos, asegurando que tengan las herramientas suficientes para demostrar sus habilidades y mostrar una confianza necesaria para enfrentar el mercado laboral en un mundo globalizado.

Referencias

Brown, A. (2021). The evolution of data science: From statistics to artificial intelligence. Journal of Data Science, 15(3), 45-60. (Artículo de revista)

Centro Competencia. (2023). Revolución de la ciencia de datos y su impacto social. Recuperado de https://centrocompetencia.com/revolucion-ciencia-de-datos-impacto-social/ (Página de internet)

Chassignol, M., Khoroshavin, A., Klimova, A., y Bilyatdinovac, A. (2018). Artificial Intelligence trends in education: a narrative overview. Procedia Computer Science, vol. 136, pp. 16-24. (Artículo de revista)

Data Universe. (2023). El impacto social de la ciencia de datos: Consideraciones éticas. Recuperado de https://data-universe.org/el-impacto-social-de-la-ciencia-de-datos-consideraciones-eticas/ (Página de internet)

Dickson, B. (2017). How Artificial Intelligence Is Shaping the Future of Education. Recuperado de https://www.pcmag.com/article/357483/how-artificial-intelligence-is-shaping-the-future-of-educati (Página de internet)

García, M., López, R., y Fernández, T. (2019). Big data and its impact on decision-making in organizations. International Journal of Information Management, 47, 123-135. (Artículo de revista)

Lee, H., y Martínez, P. (2018). Data science adoption in Fortune 500 companies: A case study. Proceedings of the International Conference on Data Science, pp. 234-240. (Memoria en extenso)

Pérez, J., y González, L. (2021). Challenges in data science education: Bridging the skills gap. Tech Education Journal, 12(2), 89-102. (Artículo de revista)

Politécnico Grancolombiano. (2023). 10 razones para estudiar la Ingeniería en Ciencia de Datos. Recuperado de https://www.poli.edu.co/blog/poliverso/por-que-estudiar-la-ingenieria-en-ciencia-de-datos (Página de internet)

Smith, R., y Johnson, K. (2020). The data explosion: Trends and implications for the future. Data Science Review, 8(1), 22-35. (Artículo de revista)

Tukey, J. W. (1977). Exploratory data analysis. Addison-Wesley. (Libro)

UNESCO. (2022). The role of data science in addressing global challenges. UNESCO Technical Report Series, 45. Recuperado de https://unesco.org/reports/datascience2022 (Página de internet)

Universidad de Ingeniería y Tecnología. (2023). Perfil del Egresado de Ciencia de Datos. Recuperado de https://www1.utec.edu.pe/carreras/ciencia-de-datos/perfil-de-egresado (Página de internet)

Universidad San Sebastián. (2023). El impacto de la ciencia de datos en la sociedad. Recuperado de https://www.uss.cl/noticias/ciencia-de-datos-sociedad/ (Página de internet)

Capítulo 10

Micro credenciales en el currículo de estudiantes de Ciencia de Datos a nivel superior

Daniel Marcelo González Arriaga, Juan Carlos Conde Ramírez, María Teresa Torrijos Muñoz y Mireya Tovar Vidal

Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla, México

daniel.gonzalezarr@correo.buap.mx, carlos.conder@correo.buap.mx, teresa.torrijos@correo.buap.mx, mireya.tovar@correo.buap.mx

Resumen. El mundo de la Ciencia de Datos cambia muy rápido, y por eso los profesionales siempre tienen que estar aprendiendo cosas nuevas. Para ayudar con esto, han aparecido las micro credenciales, que son como cursos cortos y flexibles que se suman a la carrera universitaria. Este trabajo se mete de lleno a ver qué tal funcionan estas micro credenciales en las carreras de Ciencia de Datos. Queremos saber si de verdad ayudan a los recién egresados a conseguir trabajo y si lo que enseñan es lo que las empresas están buscando. Analizamos los diferentes tipos que hay, como las certificaciones que dan las grandes empresas o los cursos en línea que puedes tomar a tu ritmo. También hablamos de los problemas que tienen, como el hecho de que a veces no se les da un reconocimiento oficial o que es difícil saber si son de buena calidad. Al final, damos algunas ideas para que las universidades puedan incluir estas micro credenciales en sus planes de estudio. La meta es que los estudiantes puedan ir certificando lo que aprenden y así estar mejor preparados para los retos de esta área que nunca deja de cambiar.

Palabras Clave: Micro credenciales, educación superior, Ciencia de Datos, formación profesional, empleabilidad, certificaciones, aprendizaje continuo.

1 Introducción

La Ciencia de Datos avanza muy rápido. Se ha vuelto una disciplina clave para la academia, la ciencia y la industria. Por esta razón, se necesitan profesionales mejor capacitados.

Estos especialistas deben saber adaptarse. El entorno tecnológico cambia constantemente. Además, genera nuevo conocimiento sin descanso (Wheeler & Wong, 2022).

Esto obliga a las universidades a reaccionar para no quedarse anticuadas. Deben ofrecer contenidos más diversos en sus planes de estudio. Sin embargo, las estructuras tradicionales de sus programas a menudo no bastan y se necesita algo más dinámico. Fallan en integrar las nuevas competencias de forma rápida y relevante.

Aquí es donde las credenciales alternativas son muy útiles ya que agregan ese dinamismo en la educación. Permiten a los estudiantes aprender habilidades muy específicas. Ofrecen un aprendizaje flexible, corto y enfocado en la práctica (Oliver, 2019) que las universidades no pueden. Estas credenciales las pueden emitir distintas organizaciones. Por ejemplo, universidades, plataformas en línea o grandes empresas tecnológicas. Todas ofrecen una validación formal del conocimiento y las competencias adquiridas (ver Tabla 1).

Tabla 1 Tipos de microcredenciales utilizadas en la educación superior en Ciencia de Datos

Tipo de microcredencial	**Descripción**	**Ejemplo de institución o plataforma**
Certificados en línea (MOOCs)	Cursos breves de acceso masivo en plataformas digitales, con evaluación y certificación.	Coursera, edX, Udemy
Insignias digitales (Digital Badges)	Insignias electrónicas que representan habilidades específicas adquiridas en módulos o talleres puntuales.	Credly, Badgr, Open Badges
Nano grados o programas especializados	Programas modulares más extensos enfocados en áreas específicas, con reconocimiento industrial.	Udacity Nanodegrees, edX MicroMasters
Certificaciones tecnológicas industriales	Certificaciones enfocadas en habilidades específicas emitidas por empresas tecnológicas líderes y reconocidas mundialmente, especialmente relevantes en Ciencia de Datos.	Google Career Certificates (Data Analytics), IBM Data Science Professional Certificate, Microsoft Azure Data Scientist, AWS Cloud Practitioner, Cisco CCNA, EC-Council CEH

El uso de microcredenciales en estudiantes de educación superior presenta algunos inconvenientes, tales como su reconocimiento dentro de los programas, la homogeneidad de los estándares de calidad, la integración entre sistemas, y la aceptación de los empleadores y las instituciones educativas (UNESCO, 2021). A pesar de las dificultades, varias investigaciones apuntan a la alta tasa de empleabilidad de los egresados, con

aprendizaje continuo y rutas formativas, beneficios que las micro credenciales permiten. Esto demuestra la flexibilidad de estas formaciones (Jorre de St Jorre & Oliver, 2018).

El objetivo de este trabajo es estudiar la integración de microcredenciales en los cursos de nivel universitario, en especial los orientados a Ciencia de Datos. Se examinan modelos de implementación actuales (ver Figura 1), se discuten sus beneficios y limitaciones, y se proponen estrategias para su integración efectiva en el currículo, con el fin de responder de manera pertinente a las exigencias del mercado laboral y la transformación digital de la educación superior.

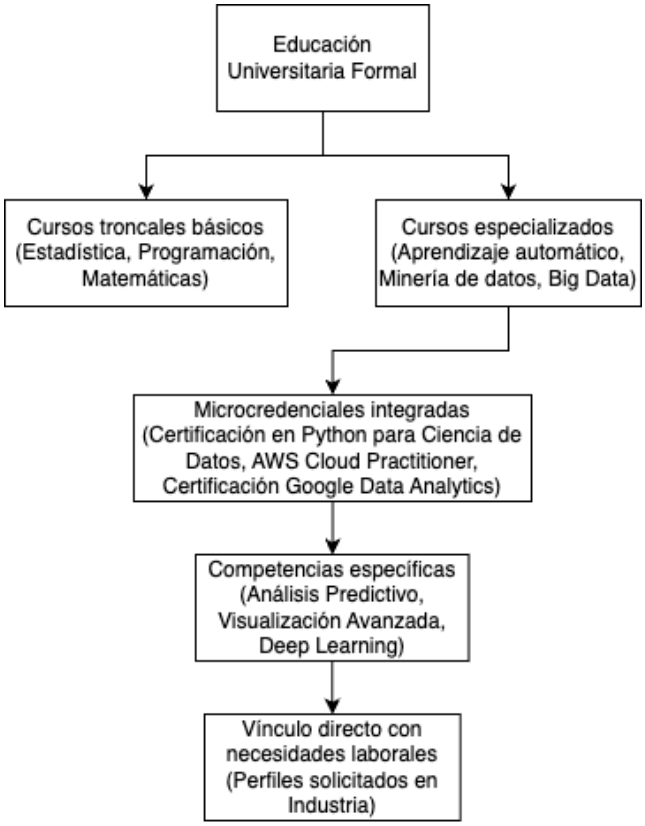


Figura 1 Modelo conceptual ilustrar la integración curricular efectiva de micro credenciales.

2 Preliminares

El esbozo de microcredenciales es reconocido en la última década por su versatilidad. Permiten la verificación de competencias con formaciones de corta duración, que son organizadas y, en la mayoría de los casos, virtuales. Se enfocan en conocimientos y competencias muy específicos y, por ello, ayudan a que los estudiantes forjen una carrera con pasos estratégicos. Esto los orienta a alcanzar metas profesionales de gran relevancia (UNESCO, 2021).

Las microcredenciales son conocidas también como "nano credenciales", "badges", "digital badges" o "certificaciones alternativas" debido a su modularidad, portabilidad y validez externa. Su diseño puede abarcar cursos breves ofrecidos por instituciones académicas o certificaciones expedidas por empresas tecnológicas líderes en el mercado como Google, IBM o Microsoft (Oliver, 2019). Su adopción dentro y fuera del ámbito universitario se ha facilitado debido a la posibilidad de evidenciar logros a través de insignias digitales o certificados verificables en línea.

Dentro de su ágil respuesta a necesidades emergentes, el valor agregado de las microcredenciales se puede evidenciar en sectores dinámicos como la tecnología, la informática y en particular la Ciencia de Datos (Wheelahan & Moodie, 2021). La interdisiplinariedad y el ritmo veloz de cambios en tecnología que son parte de la Ciencia de Datos, representan retos importantes para la educación basada en las estructuras rígidas y curriculares. En contraposición, las microcredenciales del mercado laboral se adaptan rápidamente y su carácter modular contribuye a la competitividad en la educación (Mah et al., 2019).

Aún así, hay ciertas cuestiones pendientes que deben resolverse con respecto a la integración de microcredenciales en los programas académicos de nivel superior. Allá se destacan la falta de requisitos internacionales claros, la heterogeneidad en la calidad de los contenidos, la validación académica formal y la aceptación civil y de la industria (Kato et al., 2020). Estos problemas requieren la colaboración de juntas universitarias, acreditadoras, plataformas educativas y la industria, con el objetivo de definir criterios de evaluación y certificación que puedan ser aceptados de forma transversal.

La aceptación de microcredenciales en el mercado depende también de la relevancia y prestigio de la organización que otorga la certificación. Varias certificaciones de la industria tecnológica son muy valoradas. Esto es especialmente cierto en el campo de la Ciencia de Datos.

Algunas de las más destacadas son de empresas como Google, IBM y Amazon Web Services (AWS). También son importantes las certificaciones de Cisco y las de ciberseguridad como la de EC-Council.

Estas certificaciones hacen más que complementar la formación académica. Proporcionan una credibilidad externa muy importante. Esto amplía de forma significativa las oportunidades profesionales para los egresados (Oliver, 2019; Wheeler & Wong, 2022).

3 Modelos de Implementación de Microcredenciales en Educación Superior

Las microcredenciales se han puesto en práctica de distintas formas. Las instituciones educativas las usan según sus propias consideraciones. Estas pueden ser estratégicas, operativas o pedagógicas. En este análisis, nos enfocamos en los modelos más relevantes para la práctica.

3.1. Integración Curricular Directa

Con este enfoque, las microcredenciales son integradas a la estructura académica de la institución, lo cual posibilita la obtención de créditos que se pueden reconocer en la universidad. Instituciones educativas reconocidas como el MIT han implementado programas de MicroMasters y HarvardX los cursos especializados en edX, lo que confirma este enfoque (Oliver, 2019). Estos programas poseen caminos alternativos que conducen a la obtención de credenciales académicas oficiales.

3.2. Programas Extracurriculares y Educación Continua

Este enfoque otorga a los estudiantes la posibilidad de acceder a determinadas microcredenciales de manera optativa. En su mayoría, los cursos son de corta duración o se tratan de programas modulares, especializados y son ofrecidos a través de Coursera o Udacity (Mah et al., 2019). Estos programas son ofrecidos por instituciones académicas como un aditamento al currículo y buscan permitir al estudiante adquirir habilidades que son altamente valoradas en el mundo laboral, especialmente en el rubro tecnológico.

3.3. Alianzas Estratégicas Universidad-Industria

Las colaboraciones de instituciones académicas con empresas tecnológicas líderes del momento (Google, IBM, AWS, Cisco, EC-Council) suelen ser muy efectivas gracias a la reputación de las certificaciones que se brindan y su impacto en la empleabilidad de los alumnos. Universidades como la Universidad Estatal de Arizona y el Tecnológico de Monterrey las han adoptado con resultados muy positivos (Wheeler & Wong, 2022). En la Tabla 2 se muestra una comparación detallada de estos modelos en términos de ventajas, desventajas y ejemplos representativos.

Tabla 2 Comparación de Modelos de Implementación de Microcredenciales.

Modelo	Ventajas	Desventajas	Ejemplos representativos
Integración curricular directa	Reconocimiento académico formal, créditos oficiales	Rigidez curricular institucional	MIT MicroMasters, HarvardX (edX)
Programas extracurriculares	Flexibilidad, voluntario, acceso masivo	Menor reconocimiento académico formal	Coursera, Udacity Nanodegrees
Alianzas universidad-industria	Alta empleabilidad, reconocimiento externo	Dependencia de terceros (empresas tecnológicas)	AWS Academy, Cisco Networking Academy, Google Certificates

4 Impacto de las Microcredenciales en la Empleabilidad

Este continuo impacto positivo en la empleabilidad es, sin duda, uno de los principales motivantes para que se incorporen microcredenciales en los planes académicos a nivel universidad.

4.1. Validación Externa y Reconocimiento Laboral

La inclusión de microcredenciales proveniente de certificadoras tecnológicas con fuerte reconocimiento a nivel internacional (Google, AWS, Cisco, IBM, EC-Council) elevan el prestigio de un programa y, en consecuencia, aumenta la aptitud del empleador si el egresado está designado a programas con dichas credenciales. Certificados emitidos. Especializados y relevantes para el día de hoy.

4.2. Reducción de la Brecha Formativa-Tecnológica

La microcredencialización bridge la brecha entre la educación universitaria tradicional y las habilidades prácticas necesarias en un entorno laboral. Obtener habilidades específicas con certificados de microcredencialización permite una mejor adaptación inmediata para los estudiantes en los procesos profesionales y de selección (Wheelahan & Moodie, 2021).

4.3. Evidencias Empíricas y Casos de Éxito

Instituciones académicas prestigiosas, como la Universidad Estatal de Arizona, han informado de una notable mejora en los resultados de empleo de sus estudiantes con la

adición de cursos de microcredencialización tecnológica en sus planes de estudio (Wheeler & Wong, 2022). Estos tipos de experiencias institucionales proporcionan evidencia clara del valor del conductor.

5 Desafíos y Limitaciones en la Implementación de Microcredenciales

Aunque los beneficios de la microcredencialización son claros, existen importantes desafíos que deben abordarse para lograr una implementación exitosa a largo plazo.

5.1. Reconocimiento y Equivalencia Académica Institucional

Uno de los problemas principales es la aceptación formal de microcredenciales obtenidas en plataformas externas o a través de certificaciones industriales. Esta aceptación, formal o de otra índole, puede estar relacionada a la inquietud sobre la calidad y la amplitud de los contenidos en cursos de capacitación ofrecidos por instituciones externas. Esta problemática requiere la adopción de políticas definidas y protocolos de verificación de créditos externos transparentes (Kato et al., 2020).

5.2. Estandarización y Calidad Educativa

Por los numerosos y variados proveedores, la calidad de la educación que se ofrece a través de microcredenciales debe ser asegurada. Deben definirse estándares que midan la calidad académica, pedagógica y tecnológica de la credencial antes de ser formalmente incorporada en los currículos académicos de educación superior.

5.3. Actualización Curricular Continua

Frente a los avances tecnológicos y la innovación, los cambios que ocurren en las empresas necesitan ser abordados por los programas académicos a través de estructuras curriculares que sean en perpetuo y rápido cambio. Para la incorporación y continua actualización de las microcredenciales, la flexibilidad de las instituciones es clave (Mah et al., 2019).

Estos desafíos se sintetizan visualmente en la Figura 2, proporcionando un marco conceptual claro sobre las dimensiones críticas involucradas en la integración efectiva de microcredenciales.

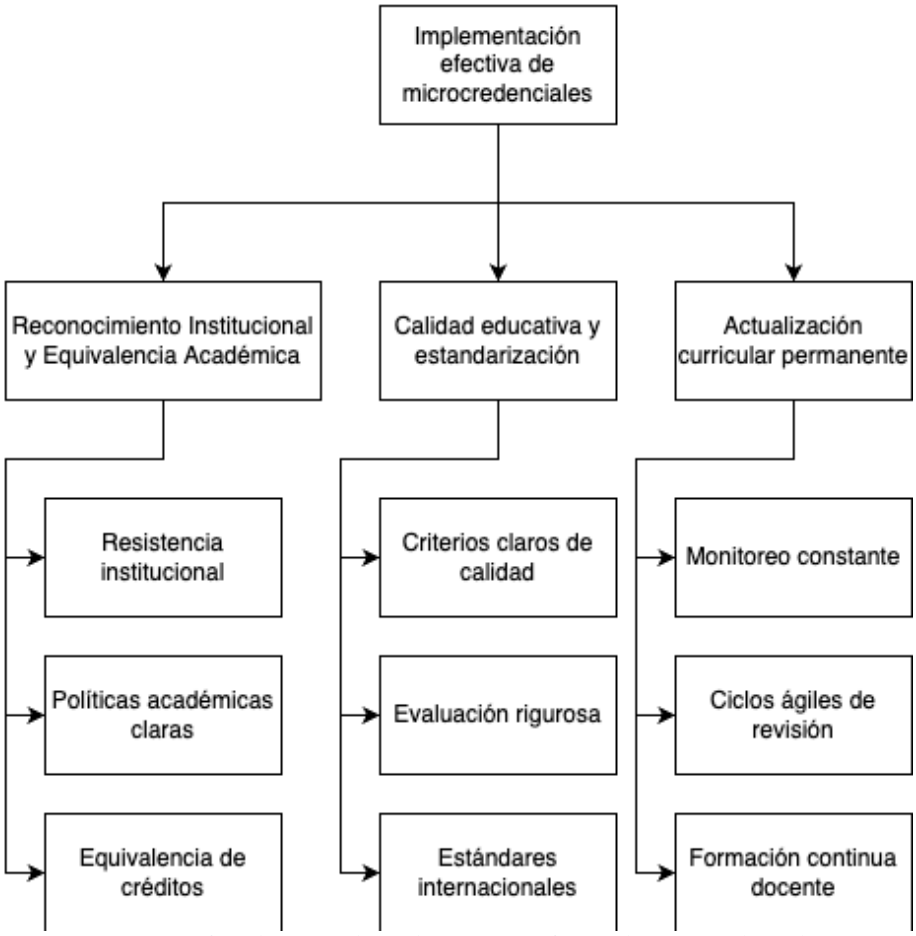


Figura 2 Desafíos clave para la implementación efectiva de microcredenciales.

6 Conclusiones

La inclusión de microcredenciales en la educación universitaria constituye una gran ventaja para adaptarse con mayor precisión a las demandas en constante cambio de la industria en Ciencia de Datos. Estos programas modulares brindan una gran oportunidad, puesto que permiten a los estudiantes obtener habilidades especializadas que son sumamente relevantes en el mercado, en comparación con la educación que brindan las aulas.

En las investigaciones que se han realizado, la "integración curricular directa", el "curso extracurricular", así como "universidad-industria" son las más efectivas para una gran variedad de necesidades familiares en desarrollo institucional. En específico, la innovación vanguardista que Google, IBM, Amazon (AWS), Cisco y EC-Council brindan en otorgar certificaciones acreditadas son un gran avance en la innovación de educación sancionada en la educación superior. Estos, en conjunto, se consideran un gran beneficio en la obtención de la educación y en el inicio de la vida laboral, puesto que especifican el reconocimiento de los mercados acerca de su uso de herramientas científicas.

Sin embargo, en el caso que se desee lograr una mayor efectividad en el uso de microcredenciales, es esencial definir y superar algunos problemas críticos. El reconocimiento académico implica mecanismos institucionales que necesitan ser afinados, desarrollando políticas sobre paralelismos académicos que sean inequívocos, estableciendo políticas que definan estándares de calidad educativa que estén claramente delineados y desarrollando procesos claros para la actualización ágil y continua del currículo para mantenerse al día con las demandas en rápida evolución del sector tecnológico.

Para salvaguardar la integración exitosa de microcredenciales en los planes de estudio universitarios de Ciencia de Datos, se recomienda a las instituciones fortalecer asociaciones estratégicas con la industria tecnológica, capacitar continuamente al personal docente sobre tendencias emergentes y adoptar políticas institucionales que valoren la flexibilidad del currículo así como la innovación en la educación.

Estas recomendaciones están bien posicionadas no solo para mejorar la implementación para enfrentar los desafíos tecnológicos y de fuerza laboral contemporáneos y del futuro inmediato, sino también para mejorar drásticamente la competitividad de los profesionales capacitados.

Referencias

Jorre de St Jorre, T., & Oliver, B. (2018). Want students to engage? Contextualise graduate learning outcomes and assess for employability. Higher Education Research & Development, 37(1), 44-57. https://doi.org/10.1080/07294360.2017.1339183 (Artículo de revista)

Kato, S., Galán-Muros, V., & Weko, T. (2020). The emergence of alternative credentials. OECD Education Working Papers, No. 216. OECD Publishing, Paris. https://doi.org/10.1787/b741f39e-en (Reporte técnico)

Mah, D. K., Yau, J. Y. K., & Ifenthaler, D. (2019). Epilogue: Future Directions on Learning Analytics to Enhance Study Success. En D. Ifenthaler, D. K. Mah, & J. K. Yau (Eds.), Utilizing Learning Analytics to Support Study Success (pp. 313-321). Springer. https://doi.org/10.1007/978-3-319-64792-0_17 (Capítulo de libro)

Oliver, B. (2019). Making micro-credentials work for learners, employers and providers. Deakin University. Recuperado de https://dteach.deakin.edu.au/wp-content/uploads/sites/103/2019/08/Making-micro-credentials-work-Oliver-Deakin-2019-full-report.pdf (Reporte técnico)

UNESCO. (2021). Towards a common definition of micro-credentials. UNESCO Institute for Lifelong Learning. Recuperado de https://unesdoc.unesco.org/ark:/48223/pf0000378179 (Reporte técnico)

Wheelahan, L., & Moodie, G. (2021). Gig qualifications for the gig economy: micro-credentials and the 'hungry mile'. Higher Education, 81(6), 1219–1233. https://doi.org/10.1007/s10734-020-00652-8 (Artículo de revista)

Wheeler, D., & Wong, D. (2022). Micro-credentials and the future of higher education. The Journal of Continuing Higher Education, 70(3), 180–193. https://doi.org/10.1080/07377363.2022.2072065 (Artículo de revista)

Índice de autores

Compiladores

Mireya Tovar Vidal
María Teresa Torrijos Muñoz
Carlos Leopoldo Carreón Díaz de León
Daniel Marcelo González Arriaga
Juan Carlos Conde Ramírez
Jaime Alejandro Romero Sierra

Revisores

Ana Laura Lezama Sánchez
Carlos García Franchini
Carlos Leopoldo Carreón Díaz de León
Claudia Zepeda Cortés
Daniel Marcelo González Arriaga
Erick Barrios González
Etelvina Archundia Sierra
Freddy Palma Mancilla
Gerardo Arizmendi Echegaray
Héctor Saib Marabillo Gómez
Héctor David Ramírez Hernández
Hilda Castillo Zacatelco
Hugo Villanueva Méndez
Jaime Alejandro Romero Siera
José Alejandro Reyes Ortiz
José Laguna Cortés
José Luis Carballido Carranza
José Luis García Cue
José Luis Zechinelli Martini
José Víctor Flores Flores
Juan Carlos Conde Ramírez
Julio Hernandez Torres
María Adelina Cruz
María Patricia Torrijos Muñoz
María Teresa Torrijos Muñoz
Martha Alvarado Arellano
Miguel Ocaña Bribiesca
Milagros Zeballos Rebaza
Mireya Tovar Vidal
Nelva Betzabel Espinoza Hernández
Reyna Carolina Medina Ramírez
Ruben Blancas Rivera
Zobeida Jezabel Guzmán Zavaleta

Editores

Mireya Tovar Vidal
María Teresa Torrijos Muñoz
Carlos Leopoldo Carreón Díaz de León
Daniel Marcelo González Arriaga
Juan Carlos Conde Ramírez
Jaime Alejandro Romero Sierra

Enfoques basados en Ciencia de Datos e Inteligencia Artificial
Editores de la publicación:
Mireya Tovar Vidal
María Teresa Torrijos Muñoz
Carlos Leopoldo Carreón Díaz de León
Daniel Marcelo González Arriaga
Juan Carlos Conde Ramírez
Jaime Alejandro Romero Sierra
A partir de diciembre de 2025
está disposición en PDF en la página
de la Facultad de Ciencias de la Computación
de la Benemérita Universidad Autónoma de Puebla (BUAP)
https://icds.cs.buap.mx/static/LIBRO1EnfoquesCDIA.pdf
Peso del archivo: 7.00 MB