Revista Cubana de Ciencias Informáticas
Vol. 19, No. 2, Abril-Junio, 2025
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 22-36

Tipo de artículo: Artículos originales
Temática: Inteligencia artificial

Aplicación de Aprendizaje Automático Automatizado en pagos electrónicos

Application of Automated Machine Learning in electronic payments

Lesdaniel Herrera Reyes [1*] https://orcid.org/0009-0006-9848-0314

[1]Universidad Tecnológica de La Habana José Antonio Echeverría (CUJAE), La Habana, Cuba.

*Autor para la correspondencia. (lesdaniel.herrera@gmail.com)

Revista Cubana de Ciencias Informáticas
Vol. 19, No. 2, Abril-Junio, 2025
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 22-36

RESUMEN

El auge de los pagos electrónicos y su desarrollo a través de las pasarelas de pago suponen un reto para la toma de decisiones encaminadas a mitigar anomalías y potenciar pagos de servicios. Esta rápida expansión ha provocado que cada vez más se recurra al uso de modelos de aprendizaje automático, algo para lo cual se requiere de personal experto en el área, que tenga la capacidad de emplear el modelo adecuado en la situación adecuada. Una solución a ello responde al AutoML, esta área del aprendizaje automático se ha visto acaparada por grandes compañías del comercio electrónico, las que han tomado el liderazgo en el desarrollo, resultando todo un reto económico su aplicación en países en vías de desarrollo. El presente artículo muestra los beneficios de esta herramienta en los sistemas nacional, así como diferentes contextos para su aplicación. Además, se presenta un servicio web que integra una herramienta AutoML para su aplicación en pagos electrónicos.

Palabras clave: pasarelas de pago; aprendizaje automático; AutoML; servicio web

ABSTRACT

The rise of electronic payments and their development through payment gateways pose a challenge for decision-making aimed at mitigating anomalies and promoting payments for services. This rapid expansion has led to increased use of machine learning models, something that requires expert personnel in the area and the ability to use the appropriate model in the appropriate situation. One solution to this is AutoML, this area of machine learning has been monopolized by large e-commerce companies, which have taken the lead in development, making its application in developing countries a real economic challenge. This article shows the benefits of this tool in national systems, as well as different contexts for its application. In addition, a web service is presented that integrates an AutoML tool for its application in electronic payments.

Keywords: payment gateways; machine learning; AutoML; web service

Revista Cubana de Ciencias Informáticas
Vol. 19, No. 2, Abril-Junio, 2025
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 22-36

Recibido: 15/02/2025

Aceptado: 18/03/2025

Introducción

El proceso de informatización de la sociedad cubana es una de las prioridades del gobierno para el desarrollo de la nación. Uno de los sectores fundamentales para su implementación es el comercio electrónico, apoyado en la evolución del internet en el país. Esto permite acceder de forma más fácil a bienes y servicios y que se realice de manera más eficiente a través de las pasarelas de pagos. Algunos de los beneficios proporcionados por los sistemas de pago en línea incluyen transacciones garantizadas, costos reducidos, aumento de la protección en la información confidencial y de pago a proveedores (Lowry et al., 2006; Perdigon et al., 2018). Una pasarela de pagos es un sistema que representa un servicio intermediario entre una página de comercio electrónico y un banco cuando se ejecutan transacciones bancarias online (Fonseca et al., 2013). Se integran a la tienda virtual y almacenan información del banco que maneja las cuentas de compradores y vendedores. En el pago con tarjeta, la pasarela de pagos valida la veracidad de la tarjeta y organiza la transferencia del dinero de la cuenta del comprador a la cuenta del vendedor. Actualmente existe gran diversidad en todo el planeta en cuanto a pasarelas de pagos se refiere, cada una de ellas se diferencia entre sí por los requisitos que cumplen, el costo que cobran por transacciones, y los países con los cuales trabajan, entre las cuales podemos mencionar Paypal, Amazon Pay o Apple Pay (Hassan et al., 2020).

En Cuba con el auge de las Tecnologías de la Información y las Comunicaciones (TICs) se han sumado pasarelas como Transfermovil (ETECSA, 2016/ 2023) y Enzona (Xetid, 2019/ 2023) que han permitido el desarrollo del comercio electrónico en el país. Sin embargo, el aumento exponencial de las operaciones de estos sistemas a raíz de la Covid-19, el desarrollo de nuevas plataformas para adquisición de bienes de manera más expedita y la expansión de servicios de conectividad genera un marcado interés en el análisis de las operaciones que operan estas plataformas y como pueden prepararse las mismas para la toma de decisiones encaminadas a mitigar anomalías y potenciar pagos de servicios.

Dado el tamaño y la complejidad de estas nuevas estructuras de datos, los métodos estadísticos estándar a menudo no serán adecuados o factibles. Se impone el uso de técnicas con capacidades mejoradas para detectar efectos de interacción, no lineales y de orden superior donde el aprendizaje automático ofrece soluciones (Chakraborty and Joseph, 2017). El aprendizaje automático se ha convertido en un campo central de la ingeniería computacional, con énfasis en los algoritmos plug-and-play disponibles a través de kits de herramientas de aprendizaje automático de código abierto. Las áreas de finanzas centradas en algoritmos, como el comercio algorítmico, han sido los principales adoptantes de esta tecnología. (Dixon et al., 2020).

El rendimiento de un método de aprendizaje automático dado en un conjunto de datos particular depende de la calidad del algoritmo, en función de diferentes criterios, como la precisión, el uso de la memoria y el tiempo de ejecución, así como su parametrización (es decir, la configuración de parámetros). En el análisis de datos comúnmente nos enfrentamos al desafío de una gran cantidad de opciones. Por ejemplo, la selección entre una amplia gama de posibles algoritmos en función de la tarea de aprendizaje automático, además de ajustar los diferentes parámetros del algoritmo seleccionado para llegar a la configuración de mejor rendimiento. Sin embargo, se necesitan evaluar varios algoritmos y ajustar sus diferentes parámetros para seleccionar el modelo de mejor rendimiento y fijar los valores de parámetro óptimos correspondientes, en función de sus suposiciones y experiencia. Esta es de hecho una tarea costosa y tediosa. Una solución a ello son los métodos de aprendizaje automático automatizado o AutoML (por sus siglas en inglés) (Bahri et al., 2022; Laptev et al., 2015)

La mayoría de las herramientas de AutoML siguen un proceso común de tres etapas ilustrado en la Figura 1. En general, estos tres componentes se optimizan de forma iterativa para obtener el mejor resultado.

Fig. 1 - Flujo de trabajo de AutoML.

El preprocesamiento de datos suele ser la primera tarea en los procesos de aprendizaje automático. Por el momento, ninguna de las herramientas de AutoML realiza muy bien esta tarea y aun requiere una considerable intervención humana. En el siguiente paso son extraídas las características del paso anterior y se utilizan para entrenar múltiples modelos, cada uno con muchos conjuntos diferentes de parámetros (optimización de hiperparámetros), luego se selecciona el mejor modelo (o un conjunto de modelos) como modelo final. Por último, se proporciona una representación detallada de los resultados a través de paneles de modelos y diferentes métodos de visualización (Truong et al., 2019; Zoller and Huber, 2015).

Las grandes compañías en la industria del software que permiten la aplicación de aprendizaje automático se han sumado al desarrollo de servicios de pago, que provean soporte al desarrollo de proyectos con AutoML; la tónica ha sido emplear las bondades de la nube para implementar estos servicios. La Tabla 1 resume la comparación entre algunas de estas herramientas AutoML evaluando el tipo de licencia y ejecución local o no, pues estos dos factores resultan capitales en la selección de entorno de trabajo, ya sea por la barrera que representan las costosas licencias propietarias o las limitaciones de acceso, dígase a internet o a los servidores de las herramientas (Karmaker et al., 2021; Mustafa and Rahimi, 2021).

Tabla 1 - Tabla comparativa de herramientas de AutoML.

Herramientas	**Licencia**	**Ejecución**	**Nivel de automatización**	**Bibliografía**
Auto-weka	Libre	Local	Requiere de la intervención humana para la construcción del flujo	Asequible
Auto-sklearn	Libre	Local	Requiere de la intervención humana para la construcción del flujo	Asequible
H2O-AutoML	Libre, versión de pago	Local	Requiere de la intervención humana para la construcción del flujo	Asequible
KNIME	Libre, pago por KNIME Server	Local, Nube	Nodo AutoML: automatiza el proceso en su totalidad	Asequible
Amazon SageMaker Autopilot	Pago	Nube	Automatiza el proceso en su totalidad, con la posibilidad de adaptar la mayor parte del mismo	Limitada
Apple CreateML	Pago	Local, Nube	Automatiza el proceso en su totalidad	Limitada

Google Cloud AutoML	Pago, libre los primeros 12 meses	Nube	Automatiza el proceso en su totalidad	Limitada
ML.NET	Libre	Local	Requiere de la intervención humana para la construcción del flujo	Asequible

En este trabajo se abordan los aspectos generales del AutoML, específicamente su posible aplicación enfocada a los pagos electrónicos a través de pasarelas de pagos. También se ofrece un análisis sobre las características, ventajas y desventajas de herramientas AutoML como solución de código abierto. Además, se presenta el desarrollo de una solución que permite el uso de esta herramienta en un sistema de pago.

Métodos o Metodología Computacional

La aplicación de AutoML en un contexto de pagos electrónicos en donde intervienen las pasarelas de pago están basadas en un enfoque financiero y para mejorar la experiencia del usuario de cara a estos sistemas intermediarios. Es por ello que el uso del aprendizaje supervisado en torno a técnicas de predicción y clasificación son imprescindible para la aplicación de AutoML en donde proponemos los siguientes escenarios que podrían beneficiar al sistema informático.

1. Predicción de anomalías en el flujo de pagos: Los datos almacenados diariamente presentan un patrón sobre su tiempo de ejecución, monto y cantidad de los mismos, basado en tipo de servicio o proveedor. Estos datos pueden clasificarse bajo un criterio binario los cuales pueden usarse con el objetivo de entrenar un modelo de regresión que permitan predecir cuándo un determinado comportamiento cumple o no las características a analizar (Nassif et al., 2021; Elmrabit et al., 2020).
2. Recomendación de los principales servicios de pago atendiendo a característica en los mismos: Los accesos rápidos a los servicios de mayor demandan para usuarios atendiendo a sus edad o uso permitiría a las pasarelas agilizar y facilitar la experiencia del usuario. Estos datos pueden clasificarse basado en la selección de cada usuario y recomendar hasta una determinada cantidad de servicios (Aslam et al., 2021; Kou et al., 2019).

3. Aplicación de bonos y descuentos selectivos: El análisis de los pagos de cada usuario facilita clasificar el perfil del mismo, ya que puede ser aprovechado para instaurar sistemas de cashback que incentiven al uso de las pasarelas de pagos a los clientes (Moscato et al., 2021; Allen et al., 2022).

Existen varias herramientas de código abierto que permiten implementar AutoML. Algunas de estas son Auto-Weka (Kotthoff et al., 2017) y Auto-Sklearn (Feurer et al., 2015) implementadas en Java y Phyton respetivamente. Sin embargo, no son los únicos; en los últimos a nos han surgido soluciones como ML.NET (Microsoft, 2017/ 2023) que ofrece la posibilidad de insertar aprendizaje automático y AutoML a soluciones en .NET como se muestra la Figura 2.

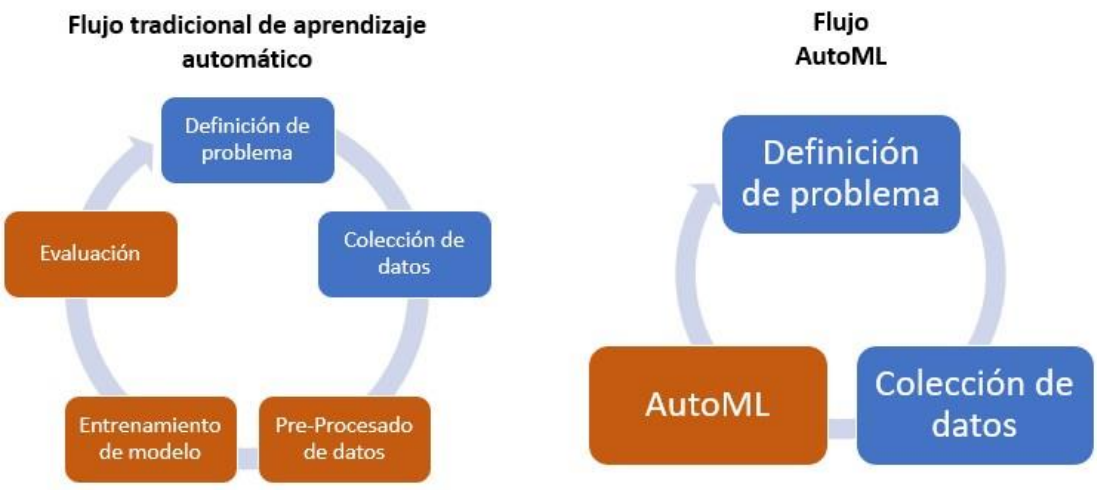

Fig. 2 - Flujo AutoML con ML.NET.

Para demostrar el uso de AutoML de decidió optar por ML.NET debido a que presenta un alto rendimiento y precisión como se muestra en la Figura 3. Además, representa una ventaja para su futuro uso en pasarelas de pagos nacionales, las cuales se pueden ver beneficiadas por su entorno de desarrollo, lo que agiliza la comprensión de la solución y permite un mayor soporte en los posibles escenarios donde estará a prueba.

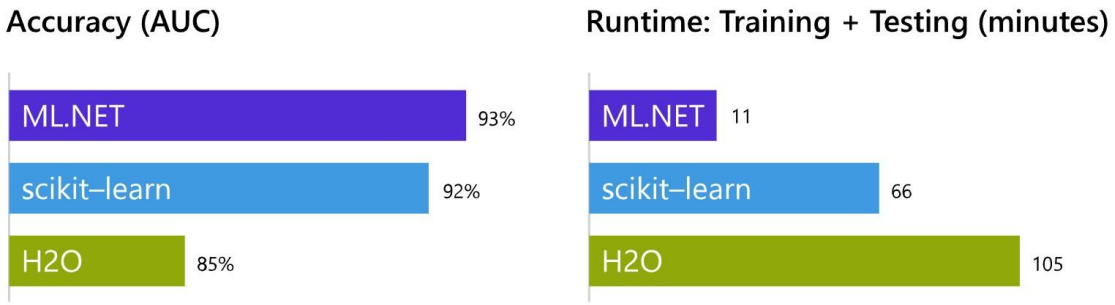

Fig. 3 - Comparación de ML.NET con otras bibliotecas AutoML. Una mayor precisión y un menor tiempo de ejecución son mejores (Ahmed et al., 2019)

API AutoML con ML.NET

Para demostrar el uso de AutoML es posible emplear una API REST en .NET que permite la integración de ML.NET al trabajo. Este sistema carga los datos de entrenamiento de un archivo *csv* y debe seleccionarse una propiedad dentro del mismo sobre la cual se realizará la predicción y categorizada como *Label*. Además, se incorpora al proceso AutoML el control del tiempo de ejecución del entrenamiento como se muestra en la Tabla 2 y el nivel de confianza para el entrenamiento.

Tabla 2 - Tiempo promedio recomendado para entrenamiento.

Tamaño de archivo de datos	**Tiempo de entrenamiento[a]**
0 - 10 MB	10 sec
10 - 100 MB	10 min
100 - 500 MB	30 min
500 - 1 GB	60 min
1 GB +	3+ hours

[a] La configuración de la API solo soporta este tiempo en segundos y aquellos casos en minuto debes ser recalculados

Flujo de trabajo AutoML

ML.NET ofrece algunas funciones preconfiguradas en su catálogo para aplicar regresión y clasificación basado en el flujo de AutoML. El proceso de entrenamiento como parte del flujo AutoML, tiene implícito el

preprocesado, donde se realiza la detención automática de los tipos de datos y una limpieza de los mismos. Se incorporó al proceso la división en datos de entrenamiento y prueba para la ejecución del experimento. El proceso de entrenamiento, que varía de acuerdo al volumen de datos, se ejecuta bajo una serie de pruebas automáticas con el fin de seleccionar el modelo más ajustado a la técnica seleccionada.

En las pruebas para la tarea de clasificación, el modelo se evalúa midiendo qué tan bien una categoría predicha coincide con la categoría real y para la agrupación, la evaluación se basa en qué tan cerca están los elementos agrupados entre sí y cuanta separación hay entre los grupos. La métrica para los problemas de predicción valora fundamentalmente el coeficiente de determinación R^2, mientras que para los problemas de clasificación se utiliza la precisión, que define la proporción de predicciones correctas que realiza el modelo sobre el conjunto de datos de prueba. Los valores de estos oscilan entre 0 y 1 y cuanto más cerca a este último, mejor será la evaluación. Otras métricas informadas, como área bajo la curva, la puntuación F1 y el error absoluto y cuadrático medio, son usadas para complementar el resultado de las pruebas.

Resultados y discusión

La API incorpora a la herramienta ML.NET, un conjunto de métodos que permiten cargar disímiles formatos de ficheros en una clase genérica. Para ello almacena la referencia al atributo original y asigna cada uno de éstos a una propiedad llamada *Col*.

El conjunto de datos utilizado para este estudio consta de 6362320 registros sin valores nulos y fue tomado del repositorio Kaggle. Originalmente provienen de datos generados utilizando el simulador llamado PaySim, el cual utiliza un conjunto de datos privado para generar datos sintéticos que se asemeja al funcionamiento normal de las transacciones e inyecta comportamiento malicioso para luego evaluar el rendimiento de los métodos de detección de fraude (PaySim, 2017/ 2023).

Como parte del proceso de AutoML, la inferencia de los tipos de datos y su manejo en la API se genera la

Tabla 3 que representa las estructura de la misma y la asignación de los atributos para el entrenamiento.

Tabla 3 - Inferencia de datos de fichero.

Nombre	Clase Genérica	Categoría	Proposito [a]
isFlaggedFraud	Col10	Single	Numeric
isFraud	Col11	Boolean	Label
type	Col0	String	Categorical
nameOrig	Col1	Single	Categorical
nameDest	Col2	String	Categorical
oldbalanceOrg	Col14	Single	Numeric
newbalanceOrig	Col15	Single	Numeric
amount	Col16	Single	Numeric
oldbalanceDest	Col17	Single	Numeric
newbalanceDest	Col18	Single	Numeric
step	Col19	Single	Numeric

[a] Valor numérico asociado al tipo de servicio.

[b] Cantidad de ocasiones en que usuario interactúa con el servicio

Prueba de la API AutoML

Basado en uno de los ejemplos expuestos para el uso de AutoML en las pasarelas de pagos aplicaremos el proceso AutoML presente en la API tomando como referencia los datos descritos anteriormente. El objetivo será predecir los posibles fraudes en los distintos tipos de pagos presentes en la fuente de datos con el objetivo de alertar al sistema ante las posibles irregularidades. Para descubrir las anomalías presentes en los flujos de pagos tomaremos como referencia la variable *isFraud* y será entrenada por un período de 30 minutos ajustando el nivel de confianza en 97.

El fujo AutoML expone como modelo más ajustado el *FastTreeBinary* para el conjunto de datos proporcionados. Los detalles de los resultados arrojados por el mismo se presentan en la Figura 4 y 5

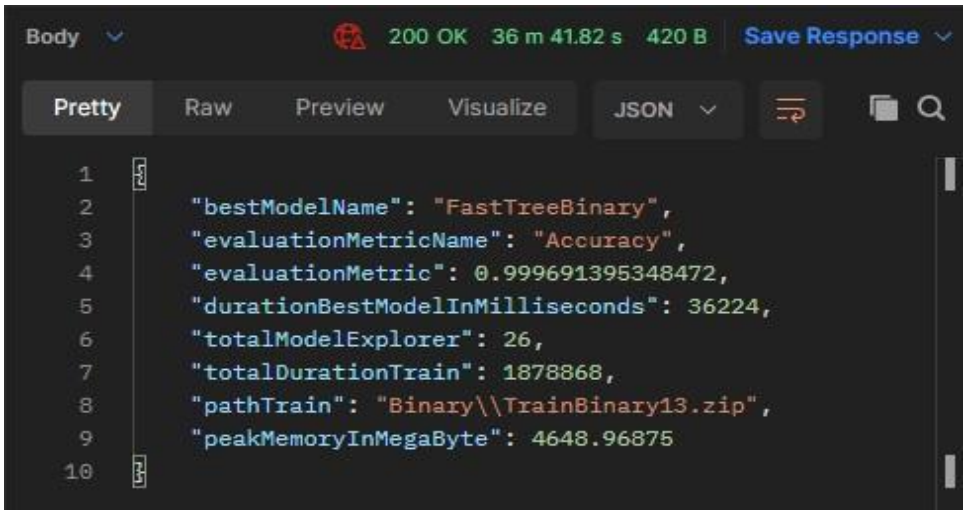

Fig. 4 - Resultado de entrenamiento con API AutoML.

```
Metrics of best model on test data --
Accuracy: 0.999691395348472
AreaUnderPrecisionRecallCurve: 0.9155560983248031
AreaUnderRocCurve: 0.9872267885685553
F1Score: 0.8666440447913133
NegativePrecision: 0.9997060159820402
NegativeRecall: 0.99998506081428
PositivePrecision: 0.9853395061728395
PositiveRecall: 0.7734706238643246
```

Fig. 5 - Detalles de resultado de entrenamiento con API AutoML.

Luego de la ejecución del proceso AutoML la API genera un archivo *zip* que almacena la estructura, el modelo predicho y los resultados de sus métricas. Este archivo resulta imprescindible para predecir muestras aisladas de datos bajo la misma estructura de archivo o para reentrenar el modelo con más datos de muestra. un ejemplo de ello se puede apreciar en la Figura 6 donde se expone el resultado de una predicción de una muestra única.

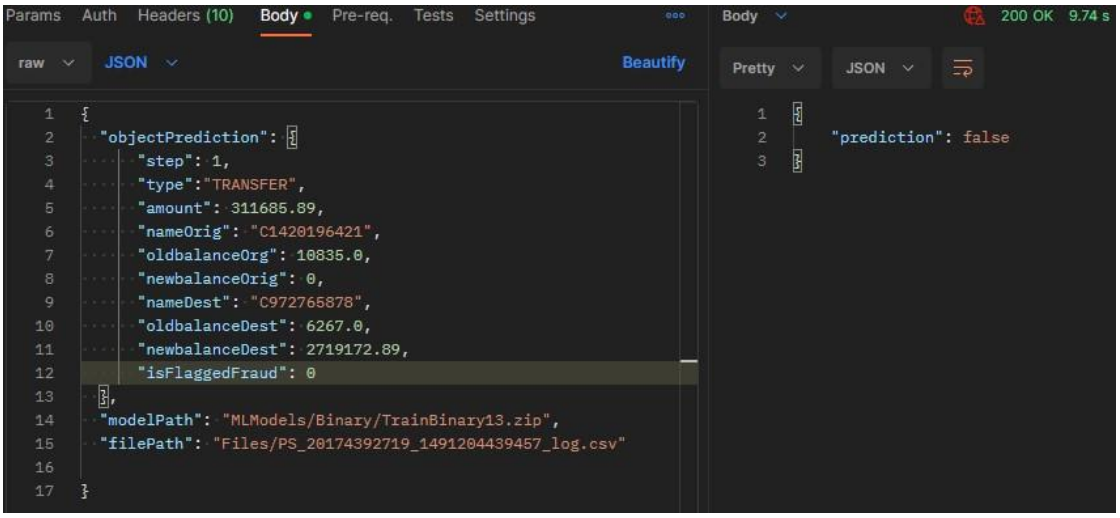

Fig. 6 - Respuesta de predicción para valor de prueba.

Conclusiones

El trabajo de investigación demuestra el uso de las herramientas AutoML en los sistemas de pagos electrónicos, así como posibles escenarios donde aplicarse. Además, muestra que los mismos facilitan la aplicación de los flujos de preprocesado, entrenamiento y evaluación presentes en el aprendizaje automático. El uso de esta herramienta permite lograr agilizar resultados para aquellos profesionales que no cuentan con un gran dominio de técnicas de aprendizaje automático y deseen optar por alternativas de código abierto. El sistema desarrollado permite un acercamiento a esta herramienta y sienta las bases para facilitar su aplicación en pasarelas de pagos electrónicas nacionales.

Referencias

Z. Ahmed, S. Amizadeh, M. Bilenko, R. Carr, and W. Chin. Machine learning at microsoft with ml. net. *In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery and data mining*,

(11):2448–2458, 2019.

J. Allen, S. Carbo-Valverde, S Chakravorti, F. Rodríguez-Fernandez, and O. Pinar Ardic. Assessing incentives to increase digital payment acceptance and usage: A machine learning approach. *Plos one*, 17(11), 2022.

F. Aslam, T. Awan, and T Fatima. Classification of m-payment users behavior using machine learning models. *Journal of Financial Services Marketing*, 1(1):1–12, 2021.

M. Bahri, F. Salutari, A. Putina, and M Sozio. Automl: state of the art with a focus on anomaly detection, challenges, and research directions. *International Journal of Data Science*, 14(2):113–126, 2022.

C. Chakraborty and A. Joseph. Machine learning at central banks. *Bank of England Working Paper*, 1(674): 89, 2017.

M. F. Dixon, I. B. Halperin, and P. Bilokon. *Machine learning in finance*. Springer International Publishing, 2020.

M. Elmrabit, F. Zhou, F Li, and H. Zhou. Evaluation of machine learning algorithms for anomaly detection. *International conference on cyber security and protection of digital services*, (1):1–8, 2020.

ETECSA. Transfermovil. https://www.etecsa.cu/es/transfermovil, 2016/ 2023. Accessed: 02-02-2023.

Matthias Feurer, Aaron Klein, Katharina Eggensperger, Jost Springenberg, Manuel Blum, and Frank Hutter. Efficient and robust automated machine learning. *Advances in neural information processing systems*, 28, 2015.

D. S. Fonseca, W. R. Perez, and M. L Faurés. Pasarela de pagos para la seguridad de transacciones bancariasén línea. *3c Empresa: investigación y pensamiento crítico*, 2(6):8, 2013.

M. Hassan, S. Shukur, and M Hasan. An efficient secure electronic payment system for e-commerce. *Computers*, 9(3):1–66, 2020.

S. K. Karmaker, M. M. Hassan, M. J. Smith, Lei. Xu, and C Zhai. Automl to date and beyond. challenges and opportunities. *ACM Computing Surveys*, 54:1–36, 2021.

Lars Kotthoff, Chris Thornton, Holger H. Hoos, Frank Hutter, and Kevin Leyton-Brown. Auto-weka 2.0: Automatic model selection and hyperparameter optimization in weka. *Journal of Machine Learning Research*, 18(25):1–5, 2017. URL http://jmlr.org/papers/v18/16-261.html.

F. Kou, X. Chao, Y. Peng, F. Alsaadi, and V. Herrera. Machine learning methods for systemic risk analysis in financial sectors. *Technological and Economic Development of Economy*, 25(5):1–27, 2019.

N. Laptev, S. Amizadeh, and I. Flint. Generic and scalable framework for automated time-series anomaly detection. *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, 2:1939–1947, 2015.

P. B. Lowry, T. B. Wells, G. D. Moody, S. Humphreys, and D Kettles. Online payment gateways used to facilitate e-commerce transactions and improve risk management. *Communications of the Association for Information Systems*, 17(6):1–48, 2006.

Microsoft. Ml.net. https://dotnet.microsoft.com/es-es/apps/machinelearning-ai/ ml-dotnet, 2017/ 2023. Accessed: 02-02-2023.

V. Moscato, A. Picariello, and V Sperlí. A benchmark of machine learning approaches for credit score prediction. *Expert Systems with Applications*, 165(8):11–39, 3 2021.

A. Mustafa and A. Rahimi. Automated machine learning for healthcare and clinical notes analysis. *MDPI Open Access Journal*, 10:24, 2021.

A. B Nassif, M. A Talib, and Q. Nasir. Machine learning for anomaly detection: A systematic review. *Computers*, 9(3):78658–78700, 2021.

PaySim. Synthetic financial dataset for fraud detection. https://www.kaggle.com/datasets/ ealaxi/paysim1, 2017/ 2023. Accessed: 02-06-2023.

L. R. Perdigon, S. H. Viltres, and L. I Madrigal. Estrategias de comercio electronico y marketing digital para pequenas y medianas empresas. *Revista Cubana de Ciencias Informaticas*, 12(2):192–208, 2018.

A. Truong, A. Walters, J. Goodsitt, K. Hines, B. Bruss, and R. Farivar. Towards automated machine learning: Evaluation and comparison of automl approaches and toolss. *Iinternational conference on tools with artificial intelligence (ICTAI)*, pages 1471–1479, 2019.

Xetid. Enzona. https://www.enzona.net/inicio, 2019/ 2023. Accessed: 02-02-2023.

M-A. Zoller and M. Huber. Benchmark and survey of automated machine learning frameworks. *Journal of Artificial Intelligence*, 70:409–472, 2015.

Conflicto de interés

El autor autoriza la distribución y uso de su artículo

Contribuciones de los autores

Conceptualización: Lesdaniel Herrera Reyes

Análisis formal: Lesdaniel Herrera Reyes

Investigación: Lesdaniel Herrera Reyes

Metodología: Lesdaniel Herrera Reyes

Administración del proyecto: Lesdaniel Herrera Reyes

Supervisión: Lesdaniel Herrera Reyes

Validación: Lesdaniel Herrera Reyes

Visualización: Lesdaniel Herrera Reyes

Redacción - borrador original: Lesdaniel Herrera Reyes

Redacción - revisión y edición: Lesdaniel Herrera Reyes