# Machine Learning en la detección de fraudes de comercio electrónico aplicado a los servicios bancarios (*Machine Learning en la detección de fraudes de comercio electrónico aplicado a los servicios bancarios)*

Fredi Alvarez[1]

## Resumen

Uno de los principales riesgos a los que están sometidas las entidades financieras son los ataques de fraudes electrónicos. Billones de dólares en pérdidas son absorbidas cada año por las entidades financieras debido a transacciones fraudulentas.

Este artículo plantea un modelo que considera los principales retos en el diseño de un sistema de detección de fraudes: a) clases altamente desequilibradas, b) distribución de estacionaria de los datos y c) la incorporación en línea de la retroalimentación de los investigadores de fraude ante las transacciones etiquetadas como sospechosas. La implementación del modelo en un conjunto de datos de prueba permitió predecir exitosamente la mayoría de casos de transacciones fraudulentas con un mínimo porcentaje de falsos negativos.

**Palabras clave:** fraude electrónico, detección de fraude, sistema de detección de fraudes, servicios bancarios, machine learning, random forest, herramientas big data.

[1] Softwaresocial Consultores. eddybucle@gmail.com

## Abstract

One of the main risks to which financial institutions are subject are electronic fraud attacks. Billions of dollars in losses are absorbed each year by financial institutions due to fraudulent transactions.

This article presents a model that considers the main challenges to design a fraud detection system: a) highly unbalanced classes, b) stationary distribution of data and c) incorporation of online feedback from fraud investigators on transactions labeled suspicious. The implementation of the model in a test dataset allowed to successfully predict the majority of cases of fraudulent transactions with a minimum percentage of false negatives.

**Keyword**: electronic fraud, fraud detection, frauds detection systems, banking services, machine learning, random forest, big data tools.

## I. Introducción

La necesidad de las empresas y entidades bancarias que sus usuarios gestionen operaciones y transacciones por internet eleva el potencial riesgo que dichos procesos sean manipulados por hackers poniendo en peligro la integridad de los datos de los usuarios y la información reservada de las entidades bancarias (Morales, 2018). El fraude contra el sistema financiero constituye un gran problema para los bancos ya que ocasionan pérdida económica, pérdida de imagen y desconfianza de los clientes.

Cuando no se puede prevenir un fraude, es necesario poder detectarlo lo antes posible. El problema de la prevención y detección de un fraude, se magnifica por una serie de características y limitaciones. En primer lugar, se debe tener cuidado de no procesar demasiadas transacciones legítimas o bloquear incorrectamente tarjetas o cuentas genuinas. En segundo lugar, las instituciones financieras procesan un gran número de transacciones, de las cuales sólo un pequeño porcentaje es fraudulento (cerca del 0,1%) (Juszczak, Adams, Hand, Whitrow, & Weston, 2008), (Dal Pozzolo A., 2015). En tercer lugar, sólo un número limitado de transacciones pueden ser revisados por investigadores de fraudes, por lo que las instituciones financieras necesitan automatizar su proceso de detección.

Una vez identificados los riesgos de fraude a los que se enfrentan las entidades financieras, es necesario contar con herramientas informáticas, que permitan identificar dentro del gran número de registros de transacciones, patrones de comportamiento que no son usuales y/o que corresponden a actividades potencialmente fraudulentas. La detección a tiempo de un fraude permite evitar daños, proteger la reputación, los activos corporativos e incrementar la confianza por parte de los clientes.

El presente artículo plantea un modelo para detección de fraudes en transacciones financieras. Debido a la falta de información pública de transacciones, el modelo fue planteado utilizando datos de transacciones de autorización de tarjetas de crédito, pero puede ser fácilmente adaptado para otro tipo de transacciones financieras.

El resto del trabajo se organiza de la siguiente manera. La Sección 2 realiza una revisión del concepto y los tipos de fraude electrónico, la Sección 3 presenta los principales conceptos sobre Machine Learning y la clasificación principal de sus algoritmos. La Sección 4 introduce los principales trabajos relacionados en el área. La Sección 5 describe el enfoque propuesto para implementación de una solución de un Sistema de Detección de Fraudes para transacciones del sector bancario. En la Sección 6 se realiza un análisis de los resultados obtenidos con un conjunto de datos de prueba. Finalmente, la Sección 7 menciona las conclusiones y futuros trabajos a realizar con enfoque propuesto.

## II. Fraude Electrónico

El fraude electrónico o delito informático es una actividad indebida basada en la manipulación fraudulenta de elementos informáticos y sistemas de comunicación, para obtener un beneficio no autorizado (Téllez, 2004).

Las cifras publicadas por el APWG (Anti-Phishing Working Group) indican que esta actividad ha aumentado un 5,7% en los últimos 12 años, siendo los países de Latinoamérica los más afectados (Group, 2019). Pese a la recesión económica que varios países han sufrido en los últimos años, los mercados de seguridad de TI han crecido considerablemente, empresas en Latinoamérica han aumentado su concientización sobre las amenazas virtuales y consecuentemente han realizado mayores inversiones en Seguridad de TI en los últimos años. Una de las exigencias de regulaciones internacionales a partir del 2010 es el estándar de seguridad de la información para la industria de tarjetas de pago o PCI DSS (Security Standards Counsil, 2018).

### II – A. Tipos de Fraudes Electrónicos

Phishing. Es sinónimo de suplantación de identidad, es la obtención fraudulenta de información crítica de los clientes por medio de una página web similar a la auténtica, los principales datos críticos a ser obtenidos pueden ser nombres de usuario, contraseñas, números de tarjetas de crédito (Aite, 2014).

**Malware.** Es la abreviatura de **"***Malicious Software"*; este término abarca a todo programa o código malicioso cuyo objetivo es dañar el sistema operativo o causar un mal funcionamiento (Koshrow-Pour). No importa qué tan amplia sea la protección técnica, los estafadores usan malware inteligente para mantenerse al día, esto significa que las amenazas seguirán aumentando, y no solo en los dispositivos móviles.

**Keyloggers.** Son pequeños programas maliciosos utilizados para capturar cualquier tipo de actividad realizada en el computador, de esta manera se podría conocer todo lo digitado por la víctima, esta actividad se puede grabar en un registro en el mismo ordenador o a su vez enviarlo a un equipo remoto que fue configurado por el atacante (Pathak, Pawar, & Patil, 2015).

**Spyware.** Comúnmente conocido como software espía, para este ataque se utiliza un software que es instalado sin autorización en la máquina de la víctima para monitorear las actividades que esta realiza desde un acceso remoto (Koshrow-Pour).

**Spam.** Involucra el envío de correo no deseado donde los correos electrónicos no solicitados o las publicaciones de los grupos de noticias no deseados se envían sin el consentimiento del receptor, con frecuencia son maliciosos y en ocasiones los delincuentes pretenden ser instituciones o compañías financieras solicitando información personal o credenciales de acceso a sus cuentas (Mendez, Fdez-Riverola, Díaz, & Corchado, 2007).

**Hacking.** El hacking malicioso o más conocido como cracking es uno de los delitos informáticos más antiguos relacionados con el acceso ilegal a un sistema informático (Seguridad en Sistemas y Técnicas de Hacking, 2011). Los piratas informáticos obtienen acceso no autorizado a grandes cantidades de datos confidenciales con el objetivo de robar información, causar daños monetarios y de reputación a la entidad objetivo (Donaldson, Williams, & Siegel, 2019).

## III. Machine Learning

Machine Learning (ML) es una disciplina científica que maneja sistemas inteligentes, es decir que aprenden automáticamente al identificar ciertos patrones presentes en los datos. Para este aprendizaje ML usa algoritmos que se encargan de revisar datos mediante ejemplos o instrucciones predefinidas para así predecir comportamientos futuros permitiendo además la incorporación de información adicional y reajustar el resultado. ML maneja conocimiento inductivo obteniendo un enunciado general en base a enunciados que describen casos particulares (Mohri, Rostamizadeh, & Talwalkar, 2018). Los algoritmos de ML se clasifican generalmente en:

**Aprendizaje supervisado.** La máquina aprende no sólo de los propios datos finales (inputs), sino que es posible darle modelos o datos adicionales ya categorizados (outputs) para que el aprendizaje sea mucho más fiable (Murphy, 2012).

**Aprendizaje no supervisado.** Sólo se dan los datos finales (inputs) a la máquina para que encuentre patrones interesantes a partir de esos datos (Murphy, 2012). A diferencia del aprendizaje supervisado, el no supervisado utiliza procedimientos inductivos, extrayendo conocimiento sólo de los datos, como en el caso del análisis de clusters para la clasificación.

**Aprendizaje por refuerzo.** En el aprendizaje por refuerzo (AR) el agente no cuenta con los datos de entrada y la respuesta esperada. En éste caso el algoritmo intenta obtener la mayor recompensa posible ante un determinado estado y una acción tomada para dicho estado. El agente debe descubrir que acciones le brindan la mayor recompensa ante determinado estado, una medida numérica, un número alto representa un mayor nivel de recompensa (Merino, 2019).

## IV. Trabajos Relacionados

La minería de datos y el aprendizaje automático son métodos populares para estudiar y combatir los casos de fraude con tarjetas de crédito. Existe una gran cantidad de estudios que explotaron la fuerza de la minería de datos y el aprendizaje automático para prevenir las actividades fraudulentas con tarjetas de crédito.

Por su parte, en (Adewumi, 2016), los autores han realizado una revisión de las técnicas mejoradas de detección de fraudes con tarjetas de crédito. Precisamente,

el trabajo se centró en las recientes técnicas de detección de fraudes de tarjetas de crédito basadas en el aprendizaje automático e inspiradas en la naturaleza propuestas en la literatura. Los autores han proporcionado una imagen de la tendencia reciente en la detección de fraudes con tarjetas de crédito. Además, esta revisión ha descrito algunas limitaciones y contribuciones de las técnicas de detección de fraude con tarjetas de crédito existentes, y también ha proporcionado la información básica necesaria para los investigadores en este dominio. Como conclusión, el trabajo podría servir como guía y trampolín para las instituciones financieras y las personas que buscan técnicas nuevas y efectivas de detección de fraude con tarjetas de crédito.

En (Dhankhad, 2018), los autores proponen diferentes algoritmos de aprendizaje automático supervisados para detectar transacciones fraudulentas con tarjetas de crédito utilizando un conjunto de datos del mundo real. Además, han empleado estos algoritmos para implementar un superclasificador utilizando métodos de aprendizaje por conjuntos. Identificaron las variables más importantes que pueden conducir a una mayor precisión en la detección de transacciones fraudulentas con tarjetas de crédito. Por su parte, en (Awoyemi, 2017), los autores han propuesto analizar el rendimiento de Naive Bayes, KNN y regresión logística en datos de fraude de tarjetas de crédito muy sesgados. Los autores desarrollaron una técnica híbrida de submuestreo y sobremuestreo de los datos asimétricos. Las tres técnicas se aplican a los datos sin procesar y pre-procesados. Con una implementación realizada en Python, los resultados muestran una precisión óptima para Naive Bayes, KNN y regresión logística de 97,92%, 97,69% y 54,86% respectivamente.

En (Yee, 2018), los autores emplearon técnicas de aprendizaje automático para predecir las transacciones sospechosas y no sospechosas automáticamente mediante el uso de clasificadores. La combinación de técnicas de aprendizaje automático y minería de datos pudo identificar las transacciones genuinas y no genuinas al aprender los patrones de los datos. Los autores analizaron la clasificación basada en supervisión que utiliza clasificadores de redes bayesianas, como, K2, Tree Augmented Naïve Bayes (TAN) y Naïve Bayes, logística y clasificadores J48. Después de preprocesar el conjunto de datos mediante la normalización y el análisis de componentes principales, todos los clasificadores lograron una precisión superior al 95% en comparación con los resultados obtenidos antes de preprocesar el conjunto de datos.

En esta línea, en (Campus, 2018), los autores han investigado el rendimiento del árbol de decisiones, Random Forest, SVM y regresión logística en datos de fraude de tarjetas de crédito muy sesgados. El conjunto de datos de transacciones con tarjetas de crédito proviene de titulares de tarjetas europeos que contienen 284.786 instancias. Estas técnicas se han aplicado a los datos sin procesar y pre-procesados. El rendimiento de las técnicas se evaluó en función de la precisión, sensibilidad, y especificidad. Los resultados indicaron que la precisión óptima para la regresión logística, el árbol de decisión, Random Forest y SVM con valores de 97,7%, 95,5%

y 98,6%, 97,5% respectivamente. Como se puede ver, Random Forest obtuvo un desempeño superior al resto de los algoritmos.

Finalmente, en (Sailusha, 2020), los autores implementaron algoritmos de random forest y el Adaboost para la detección de fraude en tarjetas de crédito. Los resultados de los dos algoritmos se basan en la accuracy, precision, recall y F1-score. Se compararon los algoritmos Random Forest y Adaboost y el algoritmo que ha demostrado el mejor desempeño para detectar el fraude fue Random Forest.

## V. Enfoque propuesto

Los Sistemas de detección de Fraudes se basan en operaciones manuales y automáticas. Las operaciones manuales son realizadas por personal de la entidad financiera denominadas investigadores de fraude, mientras que los componentes automáticos se implementan mediante algoritmos que funcionan en tiempo real y casi tiempo real. Las operaciones en tiempo real tienen lugar antes de que se autorice el pago, mientras que las operaciones casi en tiempo real se ejecutan después de que se produjo el pago.

La propuesta planteada se implementa para las operaciones cuya autorización ha sido emitida y propone un modelo de ML corriendo sobre componentes basados en herramientas estándar del ecosistema de Apache: Kafka, Spark Streaming, Cassandra. Se escogieron estas herramientas ya que corren en un mismo ecosistema y presentan similares sistemas de tolerancia a fallos y tareas distribuidas. Además, los componentes planteados permiten escalamiento horizontal cuando se tenga una gran cantidad de transacciones a ser procesadas.

### V – A. Construcción del Modelo

Para proceder a la construcción de un correcto modelo de aprendizaje automático supervisado se deben seguir los siguientes pasos tal como se muestra en la Figura 1:

- Realizar el proceso de *feature engineering* (ingeniería de características) para transformar los datos históricos en un conjunto de características y etiqueta de clasificación para ser utilizado en un algoritmo de aprendizaje automático supervisado.
- Dividir el conjunto de datos disponible en dos partes, una para construir el modelo y otra para probar el modelo.
- Construir el modelo con las características y etiquetas de entrenamiento.
- Probar el modelo con el conjunto de datos de prueba para obtener predicciones y luego comparar las predicciones obtenidas del modelo con las etiquetas del conjunto de pruebas.
- Ajustar el modelo hasta obtener un índice de precisión deseado.

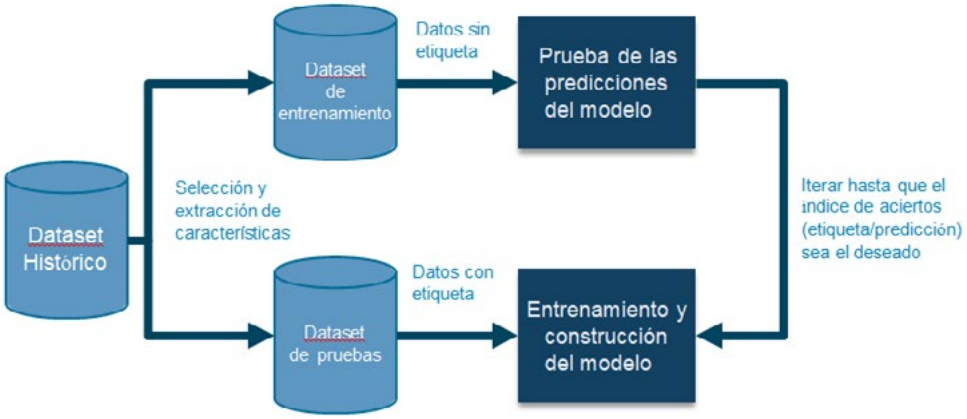


**Figura 1.** Flujo construcción de un modelo de aprendizaje automático.

## V – B. Feature Engineering

Feature Engineering hace referencia al proceso de transformar datos sin procesar en entradas para un algoritmo de aprendizaje automático. La ingeniería de características depende en gran medida del tipo de caso de uso y las posibles fuentes de datos.

*Para el caso de fraudes en tarjetas de crédito las características (atributos) de cada transacción se pueden dividir* en:

- Características asociadas a la transacción: fecha y hora de la transacción, comercio adquiriente, categoría de comercio, monto de la transacción, si es corriente o diferido, ciudad de origen y país de origen.
- Características asociadas a la tarjeta/cuenta y su dueño: número de tarjeta/cuenta, tipo de tarjeta/cuenta, edad del dueño, si es principal o adicional, género del dueño, y nacionalidad del dueño de la tarjeta.
- Características obtenidas a partir del histórico de transacciones: monto mínimo, monto máximo, sumatoria de montos, cantidad de transacciones, promedio de los montos, monto mínimo por cada categoría del comercio, monto máximo por categoría del comercio, suma de los montos por categoría del comercio, cantidad de transacciones por categoría del comercio y promedio de los montos por categoría del comercio.

## V– C. Implementación del Modelo

La Figura 2 muestra un diagrama de la arquitectura a alto nivel de la solución propuesta. Los eventos de las transacciones son entregados a través del sistema de mensajería de Kafka. Spark streaming procesa y verifica las transacciones en busca de características fraudulentas utilizando MLib con el modelo implementado y asigna una probabilidad de fraude. Luego se utiliza Cassandra para almacenar las transacciones enriquecidas con los datos agregados y almacena la tabla con el top N de las transacciones con mayor índice de probabilidad de fraude. Esta tabla es luego tomada por el sistema monitor que muestra los datos al personal encargado de comunicarse con el dueño de la tarjeta y etiquetar la transacción como genuina o fraudulenta que a posterior será utilizado en el siguiente periodo de entrenamiento como feedback.

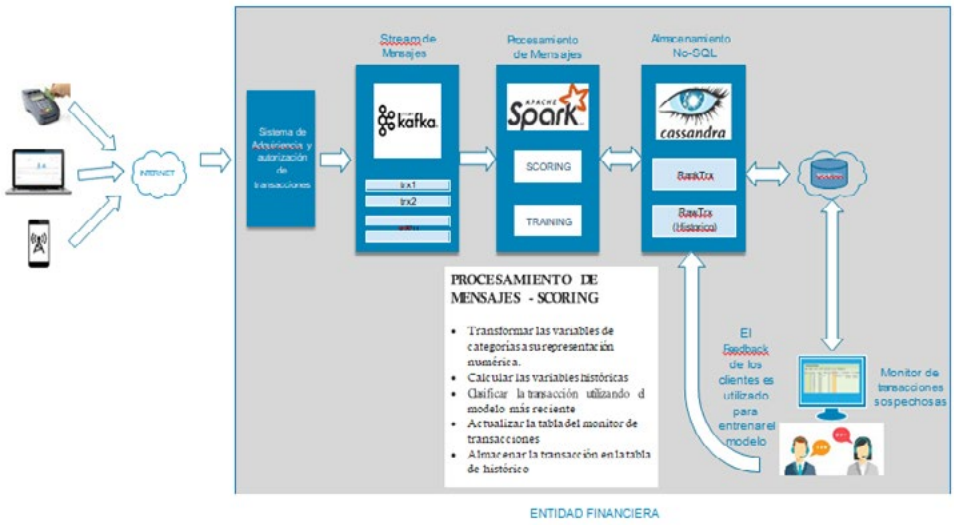


**Figura 2.** Arquitectura implementación del modelo.

La solución propone un Sistema de Detección de Fraudes donde la herramienta permite la interacción con los investigadores de fraude. El papel de los investigadores de fraude es centrarse en las transacciones más sospechosas y contactar a los titulares de tarjetas.

Para esto el sistema automático recibe una retroalimentación (legítimo o fraude) de solo un pequeño subconjunto de transacciones que activaron una alerta, para el resto de las transacciones no se reciben comentarios a menos que el titular de la tarjeta informe un fraude. Esto significa que se puede suponer que las transacciones no alertadas son genuinas solo después de un tiempo. Ésta estrategia permite integrar los datos de las transacciones para las que se tiene el feedback de los investigadores

y las reportadas por el cliente con cierto retraso.

Como su puede ver en la Fig. 2, el pre-procesamiento de la entrada que se hace es la siguiente, correspondiente al procesamiento de mensajes realizado por el módulo de Spark:

- Transformar las variables de categorías a su representación numérica;
- Calcular las variables históricas;
- Clasificar la transacción utilizando el modelo más reciente;
- Actualizar la tabla del monitor de transacciones;
- Almacenar la transacción en la tabla de histórico.

La clasificación se basa en el algoritmo de Random Forest, que ha demostrado ser particularmente eficaz en escenarios de detección de fraude (Whitrow, Hand, Juszczak, Weston, & Adams, 2009), (Bhattacharyya, Jha, Tharakunnel, & Westland, 2011), (Correa Bahnsen, Stojanovic, Aouada, & Ottersten, 2013), (Viola & Jones, 2001).

La estimación del riesgo final de una transacción está dada en función de los resultados de la probabilidad de una transacción positiva (fraudulenta) para cada uno de los 2 clasificadores:

- Un clasificador basado en Random Forest *($F_t$)* que considera las observaciones generadas en los últimos días y para el cual los investigadores devolvieron un feedback.
- Un clasificador *"Delayed" (Dt)* basado en Balanced Random Trees (BRTs) entrenado con las transacciones anteriores para los cuales luego de n días se puede asumir razonablemente que tienen etiquetas conocidas. Este clasificador necesita un conjunto de datos mucho más grande para su entrenamiento.

## VI. Resultados y Análisis

Para la implementación del prototipo de la solución planteada, se tomó como datos las transacciones de tarjetas de crédito en un período de 21 días con una muestra de 80.792 transacciones, como se ilustra en la Tabla 1.

Para correr los experimentos, se utilizó el lenguaje Python, sobre un sistema operativo Windows 10 en una arquitectura Intel Core i3 con 4GB de RAM.

| | Transacciones | Transacciones Fraudulentas | Transacciones Legales |
|---|---|---|---|
| Valor | 80792 | 286 | 80686 |
| Porcentaje | 100% | 0.35% | 99.64% |

**Tabla 1.** Distribución de las transacciones.

## VI – A. Resultados

Los resultados presentados en la Tabla 2 son calculados a partir del scoring generado para cada transacción que es obtenido a partir de la combinación de los scorings de los 2 clasificadores presentados en la sección *IV – C. En la Figura 3 se muestra la distribución de los casos positivos, detectados y no detectados por cada día de* ocurrencia.

| | Transacciones sospechosas | Transacciones fraudulentas detectadas | Transacciones fraudulentas no detectadas | Transacciones Sospechosas Legítimas | Transacciones Legítimas no sospechosas |
|---|---|---|---|---|---|
| Valor | 640 | 264 | 22 | 376 | 80332 |
| Porcentaje | 100% | 92.30% | 7.70% | 58.75% | 99.21% |

**Tabla 2.** Distribución de transacciones detectadas y no detectadas

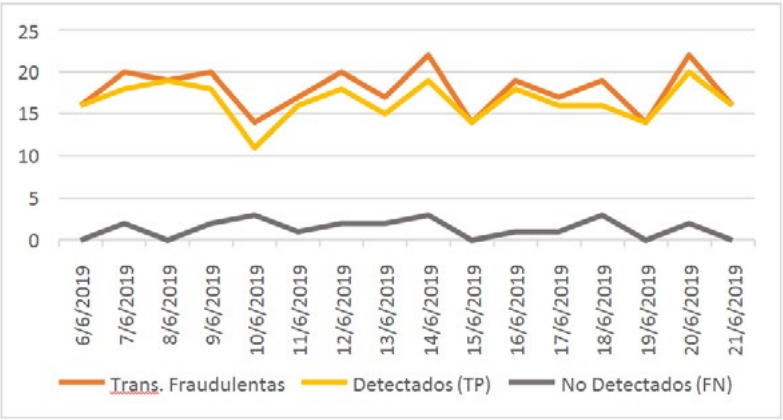


**Figura 3.** Distribución de transacciones detectadas y no detectadas por día.

Para el cálculo de los resultados en la matriz de confusión se tomaron las transacciones reportadas en el monitor como transacciones que se predijeron como fraudulentas (ver Tabla 3).

| Fraude (+) | | Clase Real | |
|---|---|---|---|
| | | Legítimo (-) | |
| Predicción | Fraude (+) | TP | FP |
| | Legítimo(-) | FN | TN |
| Fraude (+) | | Clase Real | |
| | | Legítimo (-) | |
| Predicción | Fraude (+) | 264 | 376 |
| | Legítimo(-) | 22 | 80332 |

**Tabla 3.** Matriz de confusión para el conjunto de datos de muestra.

Con los valores obtenidos de la matriz de confusión se procedió a obtener las métricas presentadas en la Tabla 4.

| Métrica | Valor |
|---|---|
| Exactitud | 0.995 |
| Precisión | 0.412 |
| TPR (Sensibilidad) | 0.923 |
| TNR (Especificidad) | 0.999 |

**Tabla 4.** Métricas obtenidas para el modelo planteado.

El valor de exactitud indica que el 99.5% de los registros fueron clasificados exitosamente, pero al ser un conjunto de datos desequilibrado se procede a obtener el valor de exactitud equilibrada.

$$Balanced\ Accuracy = \frac{TPR+TNR}{2} = \frac{0.923+0.999}{2} = 0.961$$

Por tanto, se puede deducir que el nivel de exactitud de modelo planteado tiene un valor de 96.14%.

Finalmente, cabe destacar los tiempos de ejecución de cada uno de los 2 algoritmos utilizados en el trabajo:

- Clasificador Random Forest *($F_t$):* 2500 segundos.
- Clasificador *"Delayed" ($D_t$) basado en Balanced Random Trees (BRTs): 1650 segundos.*

## VII. Conclusiones y trabajos futuros

La detección de eventos fraudulentos es una tarea particularmente desafiante y compleja, al ser eventos raros y al estar en constante evolución son difíciles de modelar. El aumento del volumen de transacciones que ocurren todos los días exige el uso de herramientas automáticas para apoyar la detección y los recursos humanos destinados a apoyar ésta tarea deben estar enfocados a investigar los casos más sospechosos. Generalmente los sistemas tradicionales utilizan una pequeña muestra de transacciones históricas de cada tarjeta para crear variables a nivel de cuenta. Debido a que es computacionalmente exigente calcular los agregados, estas características generalmente son calculadas fuera de línea y luego se agregan al vector de características que representa la transacción cuando está autorizada. La introducción de tecnologías de Big Data permite superar estos problemas, es decir, calcular los agregados en tiempo real y utilizar un conjunto más amplio de transacciones históricas. La implementación de la solución con Kafka, Spark y

Cassandra puede proporcionar escalabilidad fácil y tolerancia a fallas, para recibir, agregar y clasificar transacciones a una alta tasa de procesamiento.

Como trabajo futuro, se planifica que los resultados del modelo planteado sean revisados por los investigadores expertos en fraude para obtener sus observaciones y mejorar las predicciones del modelo. Por otro lado, se trabajará en aumentar los datasets para nuevos experimentos. Asimismo, se compararán los algoritmos descritos en la sección de trabajos relacionados utilizando el dataset del experimento realizado. Finalmente, sería deseable agregar al prototipo presentado una interfaz que se encargue de comunicar el sistema de adquiriencia y autorización de la entidad financiera con el sistema de mensajería de Apache Kafka.

## Referencias

Aite. (2014). Financial Institutions, Merchants, and the Race Against Cyberthreats.

Adewumi, A. O., & Akinyelu, A. A. (2017). A survey of machine-learning and nature-inspired based credit card fraud detection techniques. International Journal of System Assurance Engineering and Management, 8(2), 937-953.

Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis. In 2017 International Conference on Computing Networking and Informatics (ICCNI) (pp. 1-9). IEEE.

Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. (2011). Data mining for credit card fraud: A comparative study. Decision Support Systems.

Breiman, L. (2001). Universidad de California. Obtenido de Random Forest: https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf

Campus, K. (2018). Credit card fraud detection using machine learning models and collating machine learning models. International Journal of Pure and Applied Mathematics, 118(20), 825-838.

Correa Bahnsen, A., Stojanovic, A., Aouada, D., & Ottersten, B. (2013). Cost sensitive credit card fraud detection using bayes minimum risk. Machine Learning and Applications (ICMLA). 2013 12th International Conference.

Dal Pozzolo, A. (2015). Adaptive Machine Learning for Credit Card Fraud Detection.

Dhankhad, S., Mohammed, E., & Far, B. (2018, July). Supervised machine learning algorithms for credit card fraudulent transaction detection: a comparative study. In 2018 IEEE International Conference on Information Reuse and Integration (IRI) (pp. 122-125). IEEE.

Donaldson, S. E., Williams, C. K., & Siegel, S. (2019). Undertanding Security Issues.

Group, A.-P. W. (2019). Phishing Activity Trends Report - 2nd Quarter 2019.

Juszczak, P., Adams, N., Hand, D., Whitrow, C., & Weston, D. (2008). Off- the-peg and bespoke classifiers for fraud detection. En Computational Statistics & Data Analysis - Volume 52 (págs. 4521- 4532). Elsevier.

Koshrow-Pour, M. (s.f.). Encyclopedia of Criminal Activities and the Deep Web. IGI Global.

Mendez, J., Fdez-Riverola, F., Díaz, F., & Corchado, J. (2007). Sistemas inteligentes para la detección y filtrado de correo spam: una revisión. Inteligencia Artificial. Revista Iberoamericana.

Merino, M. (01 de 2019). Conceptos de inteligencia artificial: qué es el aprendizaje por refuerzo. Obtenido de https://www.xataka.com/inteligencia-artificial/ conceptos- inteligencia-artificial-que-aprendizaje-refuerzo

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). Foundations of machine learning . Second edition. Cambridge, MA : The MIT Press.

Morales, L. (2018). BIGDATA: 5 MÉTODOS PARA DETECTAR

POSIBLES FRAUDES. Obtenido de https://www.grupo- novatech.com/bigdata-5-metodos-para-detectar-posibles-fraudes/

Murphy, K. (2012). Machine learning. A probabilistic perspective. London: The MIT Press.

Pathak, N., Pawar, A., & Patil, B. (2015). A Survey on Keylogger: A malicious Attack. International Journal of Advanced Research in Computer Engineering & Technology.

Sailusha, R., Gnaneswar, V., Ramesh, R., & Rao, G. R. (2020). Credit Card Fraud Detection Using Machine Learning. In 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1264-1270). IEEE.

Security Standards Counsil. (2018). Requisitos y procedimientos de evaluación de seguridad. Version 3.2.1. Obtenido de https://es.pcisecuritystandards.org/ minisite/env2/

Seguridad en Sistemas y Técnicas de Hacking. (2011). ¿Qué es el Hacking?

Obtenido de https://thehackerway.com/about/

Téllez, J. (2004). Derecho Informático. México: McGraw-Hill, 3ra edición.

Viola, P., & Jones, M. (2001). Fast and robust classification using asymmetric adaboost and a detector cascade. Advances in Neural Information Processing System.

Whitrow, C., Hand, D., Juszczak, P., Weston, D., & Adams, M. (2009). Transaction aggregation as a strategy for credit card fraud detection. Data Mining and Knowledge Discovery.

Yee, O. S., Sagadevan, S., & Malim, N. H. A. H. (2018). Credit card fraud detection using machine learning as data mining technique. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 10(1-4), 23-27.