UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES
INSTITUTO DE POSTGRADO

TITULO DEL TRABAJO DE TITULACIÓN

APRENDIZAJE DE MÁQUINA PARA DETECTAR FRAUDE EN TARJETAS DE DÉBITO DE LA COOPERATIVA DE AHORRO Y CRÉDITO LUCHA CAMPESINA DE LA CIUDAD DE CUMANDÁ.

AUTOR

Ing. Pedro Antonio Borbor Balón

TRABAJO DE TITULACIÓN

Previo a la obtención del grado académico en
MAGISTER EN TECNOLOGÍAS DE LA INFORMACIÓN

TUTOR

Ing. Shendry Rosero Vásquez, Mgtr.

Santa Elena, Ecuador

Año 2024

UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES
INSTITUTO DE POSTGRADO

TRIBUNAL DE SUSTENTACIÓN

Firmado electrónicamente por:
ALICIA GERMANIA ANDRADE VERA

Ing. Alicia Andrade Vera, Mgtr.
COORDINADORA DEL PROGRAMA

SHENDRY BALMORE ROSERO VASQUEZ

Ing. Shendry Rosero Vázquez, Mgtr.
TUTOR

Firmado electrónicamente por:
DELIA ISABEL CARRION LEON

Ing. Delia Carrión León, Mgtr.
DOCENTE
ESPECIALISTA 1

Firmado electrónicamente por:
JUAN PABLO AMON SALINAS

Ing. Juan Pablo Amón Salinas, Mgtr.
DOCENTE
ESPECIALISTA 2

Firmado electrónicamente por:
MARIA MARGARITA RIVERA GONZALEZ

Abg. María Rivera, Mgtr.
SECRETARIA GENERAL
UPSE

UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA FACULTAD DE SISTEMAS Y TELECOMUNICACIONES INSTITUTO DE POSTGRADO

CERTIFICACIÓN

Certifico que luego de haber dirigido científica y técnicamente el desarrollo y estructura final del trabajo, este cumple y se ajusta a los estándares académicos, razón por el cual apruebo en todas sus partes el presente trabajo de titulación que fue realizado en su totalidad por PEDRO ANTONIO BORBOR BALÓN, como requerimiento para la obtención del título de Magister en Tecnologías de la Información.

TUTOR

Ing. Rosero Vásquez Shendry Balmore, Msc

Santa Elena, 22 de marzo del 2024

UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES
INSTITUTO DE POSTGRADO

DECLARACIÓN DE RESPONSABILIDAD

Yo, **PEDRO ANTONIO BORBOR BALÓN**

DECLARO QUE:

El trabajo de Titulación, Aprendizaje de máquina para detectar fraude en tarjetas de débito de la cooperativa de ahorro y crédito Lucha Campesina de la ciudad de Cumandá previo a la obtención del título en Magister en Tecnologías de la Información, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

Santa Elena, 22 de marzo del 2024

EL AUTOR

Firmado electrónicamente por:
PEDRO ANTONIO BORBOR BALON

Ing. Pedro Borbor Balón

UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA FACULTAD DE CIENCIAS DE LA INGENIERÍA INSTITUTO DE POSTGRADO

CERTIFICACIÓN DE ANTIPLAGIO

Certifico que después de revisar el documento final del trabajo de titulación denominado Aprendizaje de máquina para detectar fraude en tarjetas de débito de la cooperativa de ahorro y crédito Lucha Campesina de la ciudad de Cumandá, presentado por el estudiante, PEDRO ANTONIO BORBOR BALÓN fue enviado al Sistema Antiplagio COMPILATIO, presentando un porcentaje de similitud correspondiente al 07%, por lo que se aprueba el trabajo para que continúe con el proceso de titulación.

TUTOR

SHENDRY BALMORE ROSERO VASQUEZ

Ing. Rosero Vásquez Shendry Balmore, Msc

UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES
INSTITUTO DE POSTGRADO

AUTORIZACIÓN

Yo, **Pedro Antonio Borbor Balón**

Autorizo a la Universidad Estatal Península de Santa Elena, para que haga de este trabajo de titulación o parte de él, un documento disponible para su lectura consulta y procesos de investigación, según las normas de la Institución.

Cedo los derechos en línea patrimoniales de artículo profesional de alto nivel con fines de difusión pública, además apruebo la reproducción de este artículo académico dentro de las regulaciones de la Universidad, siempre y cuando esta reproducción no suponga una ganancia económica y se realice respetando mis derechos de autor

Santa Elena, 22 de marzo del 2024

EL AUTOR

Firmado electrónicamente por:
PEDRO ANTONIO BORBOR BALON

Ing. Pedro Borbor Balón

AGRADECIMIENTO

Elevo mi más profundo agradecimiento a Dios por su infinita bondad y por permitirme alcanzar una nueva meta en mi vida profesional. Su guía y protección fueron fundamentales para superar los desafíos que se presentaron durante este camino.

Al Ing. Shendry Rosero, expreso mi más sincero agradecimiento por su invaluable apoyo y orientación durante el desarrollo de este trabajo investigativo. Su conocimiento y experiencia fueron pilares fundamentales para la culminación exitosa de este proyecto. A las ingenieras Marjorie Coronel y Alicia Andrade mi más sincero agradecimiento por su invaluable apoyo y colaboración en la estructuración de este trabajo. Su calidad como docentes y su disposición para compartir sus conocimientos fueron de gran utilidad para lograr un buen resultado final.

Agradezco profundamente a todos los docentes que me acompañaron en esta nueva etapa de mi vida. Sus enseñanzas, consejos y dedicación fueron esenciales para mi crecimiento profesional. A mis amigos y compañeros de clase, gracias por su amistad y apoyo durante todo este proceso. Compartimos momentos de aprendizaje, alegría y también de dificultades, y su compañía fue un pilar fundamental para seguir adelante. En especial, agradezco a la ingeniera Mercedes Soriano, Carlos Mendoza y Joa Rodríguez por su invaluable apoyo en los momentos más desafiantes.

A la Cooperativa de Ahorro y Crédito Lucha Campesina, por brindarme la oportunidad de realizar este trabajo de titulación y por compartir conmigo los recursos y el apoyo necesarios para alcanzar mis objetivos. A la ingeniera Alexandra Bone mi más sincero agradecimiento por su constante y desinteresada ayuda. su apoyo fue fundamental para alcanzar la meta final.

Pedro Antonio, Borbor Balón

DEDICATORIA

A mi ser interior, a esa fuerza que reside en lo más profundo de mi alma, dedico este trabajo. A la tenacidad que me impulsó a seguir adelante incluso en los momentos más difíciles. A la pasión que me motivó a sumergirme en este proyecto con entusiasmo y dedicación. A la disciplina que me permitió mantener el enfoque y la constancia durante meses de investigación y trabajo duro.

A las noches de desvelo, a las horas de estudio y análisis, a los sacrificios y renuncias que fueron necesarios para alcanzar este logro. A la confianza inquebrantable en mis capacidades, a la fe en mi potencial y en mi capacidad para superar cualquier obstáculo.

A las lágrimas de frustración y a las sonrisas de satisfacción, a los momentos de duda y a las experiencias de aprendizaje. A la montaña rusa de emociones que acompañó este viaje, a la transformación personal que experimenté en el camino.

A la persona que soy hoy. A la versión futura de mí mismo, que encontrará en este trabajo una fuente de inspiración y un recordatorio de lo que puedo lograr cuando me lo propongo.

Pedro Antonio, Borbor Balón

ÍNDICE GENERAL

ÍNDICE DE TABLAS

ÍNDICE DE ECUACIONES

ÍNDICE DE FIGURAS

RESUMEN

Las instituciones financieras usan estrategias de prevención que le den seguimiento a las transacciones *e-commerce*, por tal motivo, el presente trabajo respecto al "*Aprendizaje de máquina para detectar fraude en tarjetas de débito de la cooperativa de ahorro y crédito Lucha Campesina de la ciudad de Cumandá",* tiene como objetivo aplicar un algoritmo basado en aprendizaje de máquina que analice las transacciones financieras de *e-commerce* de la Cooperativa de ahorro y crédito Lucha Campesina y detecte el fraude ocurrido entre los meses de abril y junio el 2023, Empleando un enfoque no experimental, de tipo transversal, con un diseño descriptivo cuantitativo. En base a revisión literaria se utilizó los siguientes modelos: Regresión Logística, Máquina de Vector de Soporte y Bosque Aleatorio, se entrenó el modelo con un conjunto de datos de 11000 transacciones legítimas y 303 fraudulentas, en donde el bosque aleatorio tuvo los mejores resultados, un f1-score del 100%.

Palabras claves: Fraude, E-commerce, Bosque aleatorio

ABSTRACT

Financial institutions use prevention strategies that follow up on e-commerce transactions, for this reason, the present work regarding "Machine learning to detect fraud in debit cards of the Lucha Campesina savings and credit cooperative of the city of Cumandá", aims to apply an algorithm based on machine learning that analyzes the e-commerce financial transactions of the Lucha Campesina Savings and Credit Cooperative and detects the fraud that occurred between the months of April and June 2023, Using a Non-experimental cross-sectional approach with a quantitative descriptive design. Based on a literary review, the following models were used: Logistic Regression, Support Vector Machine and Random Forest, the model was trained with a data set of 11,000 legitimate transactions and 303 fraudulent ones, where the random forest had the best results. an f1-score of 100%.

Keywords: Fraud, E-commerce, Random forest

INTRODUCCIÓN

La tecnología ha ayudado a las entidades financieras a brindar sus servicios por los diferentes medios electrónicos que existen (Soley, 2015), pero el mayor temor para estas instituciones sean estos bancos o cooperativas de ahorro y crédito es el fraude en las transacciones electrónicas, los países más afectados por los delitos de internet son: China, Alemania, Reino Unido, en donde los clientes pierden aproximadamente 31.000 dólares por transacciones realizadas en línea (Rohall, 2022), esto conlleva un alto nivel de riesgo en sus operaciones online, el fraude en transacciones financieras es una preocupación constante debido a la creciente sofisticación de las técnicas utilizadas por los estafadores. El nivel de fraude puede variar dependiendo de las condiciones socioeconómicas, la infraestructura tecnológica y las regulaciones de cada región. El fraude a nivel de Latinoamérica entre los años de 2019 y 2021 creció en un 23% en relación con el crecimiento del comercio electrónico en los países de México, Panamá, Chile, Argentina y Colombia (Duque, 2022)

De acuerdo a la superintendencia de economía popular y solidaria(SEPS), el 28% de los clientes de las cooperativas de ahorro y crédito realizan transacciones electrónicas (Superintendencia de Economía Popular y Solidaria, 2021), por tal motivo las entidades financieras buscan mantener sus servicios bancarios en línea, a disposición las 24 horas día para que sus clientes pueden acceder a sus servicios ya sea desde las ventanillas en las sucursales, cajeros, aplicaciones móviles, páginas web, botones de pago o por los dispositivos electrónicos ubicados en los negocios en donde el usuario puede realizar transacciones electrónicas, estos dispositivos son llamado Point of Sale (POS).

La portabilidad del dinero del socio es unos de los servicios que brindan las entidades financieras con ayuda de las tarjetas de débito o crédito, otorgándole al usuario la posibilidad de comprar en cualquier establecimiento que permita el pago con tarjeta o incluso compras en línea (*e-commerce*).

Las compras por internet o llamadas *e-commerce* se realizan sin necesidad de tener la tarjeta físicamente, es decir solo se necesita la información de la tarjeta, es por este medio *e-commerce* en donde el fraude tiene origen, los criminales cibernéticos con técnicas de ingeniería social buscan obtener información de las tarjetas de débito de los socios, para posteriormente cometer el delito.

Es fundamental que las cooperativas en Ecuador mantengan una sólida imagen ante sus clientes y generen confianza. Una cooperativa que se vea involucrada en problemas relacionados con el

fraude electrónico no podrá cumplir con estos objetivos ante sus socios, lo que representa un problema que muchas cooperativas de ahorro y crédito buscan resolver en la actualidad.

La cooperativa de ahorro y crédito Lucha Campesina tiene aproximadamente 58.000 socios entre las 9 agencias, Cumandá, Bucay, Naranjito, Simón Bolívar, Baba, El Triunfo, La Troncal, Milagro y Vinces (Lucha Campesina, 2023), el estimado de transacciones realizadas al día son de 28.000, el 40% es realizado por ventas en internet(*e-commerce*), el principal riesgo para nuestros socios es que sus transacciones se vean afectados por los delincuentes cibernéticos.

El fraude bancario representa una seria amenaza para la seguridad financiera y se cuenta entre las actividades delictivas más destacadas en todo el mundo, como señaló (Diners Club, 2022). Las instituciones financieras se enfrentan al desafío de mitigar esta problemática, no solo para evitar pérdidas económicas, sino también para salvaguardar su reputación y mantener la confianza de sus clientes.

Planteamiento de la investigación (Fundamentación de la investigación)

Las instituciones financieras buscan constantemente herramientas tecnológicas (Cuenca Jiménez et al., 2022), que les permitan controlar y gestionar eficientemente las compras electrónicas de sus socios en la cooperativa. Las compras en línea, o *e-commerce*, pueden ser propensas al fraude financiero si no se les da un seguimiento adecuado.

Con el surgimiento del aprendizaje de máquina identificar patrones y anomalías orientados a grandes volúmenes de datos es más sencillo (Gutiérrez Portela et al., 2023). En las instituciones financieras se realizan transacciones bancarias todos los días dicha transacciones están formadas por varios campos: información del adquiriente, comercio, el autorizador, del cliente. El aprendizaje de máquina puede manejar eficazmente estos grandes volúmenes de información y detectar fraudes financieros oportunamente, el margen de error es menor en el aprendizaje de máquina cuando se ha tenido data considerable para su entrenamiento.

Dependiendo del problema se puede utilizar algoritmos supervisados o no supervisados que a su vez necesitaran un conjunto de datos etiquetados o no etiquetados para realizar el proceso de aprendizaje. Uno de los desafíos al llevar a cabo las investigaciones es acceder a información relacionada con el problema, y esto se vuelve aún más complicado cuando se trata de datos relacionados con servicios bancarios.

En este estudio, la ventaja radica en que el investigador trabaja en una institución de ahorro y crédito y tiene acceso a la información necesaria para formar el dataset que se utilizara para llevar a cabo la investigación de manera viable y efectiva.

Los beneficiarios de la investigación serán los socios de las 9 agencias, Cumandá, Bucay, Simón Bolívar, Naranjito, Baba, Vinces, Milagro, La Troncal y El Triunfo, con los que cuenta la cooperativa de ahorro y crédito Lucha Campesina (Lucha Campesina, 2023).

El análisis de estudios previos sobre la aplicación de aprendizaje automático en la detección de fraudes financieros establece la base inicial para las recomendaciones que deben seguirse en esta novedosa área de investigación. Estos análisis ofrecen conclusiones metodológicas que deben considerarse en futuros estudios.

La implementación de un sistema de detección temprana de fraudes en las instituciones financieras conlleva beneficios significativos. Permite tomar acciones oportunas para prevenir transacciones fraudulentas, evitando así el impacto económico negativo en los socios propietarios de las cuentas.

Formulación del problema de investigación

¿Cómo se busca detectar los fraudes electrónicos que afectan a los socios que realizan transacciones *e-commerce* en la cooperativa de ahorro y crédito Lucha Campesina en el 2023?

Objetivo General:

Seleccionar un modelo de aprendizaje de máquina que en base al análisis de transacciones financieras de tipo e-commerce realizadas en la Cooperativa de ahorro y crédito Lucha Campesina en la ciudad de Cumandá, permita la detección de transacciones consideradas como fraude.

Objetivos Específicos:

- Adquirir un dataset de datos históricos de transacciones financieras que incluya información sobre casos de fraude y transacciones legítimas, para entrenar el modelo de aprendizaje de máquina
- Identificar en el dataset las características más relevantes que pueden ayudar a distinguir patrones o comportamientos anómalos relacionadas al fraude en el canal *e-commerce*.
- Entrenar 3 algoritmos de aprendizaje de máquina para detectar fraudes en transacciones de *e-commerce*.
- Evaluar la efectividad de los 3 algoritmos de aprendizaje de máquina.

Planteamiento hipotético

Hipótesis de la investigación

El uso de algoritmos de aprendizaje de máquina aplicado a transacciones financieras permite detectar los fraudes electrónicos.

CAPÍTULO 1. MARCO TEÓRICO REFERENCIAL

1.1. Revisión de literatura

En Perú, el artículo de (Dávila-Morán et al., 2023) "Aplicación de modelos de aprendizaje automático en la detección de fraudes en transacciones financieras", el investigador tenía como objetivo identificar en tiempo real el desempeño de las técnicas de aprendizaje automático utilizando los modelos de *Random Forest*, Redes Neuronales, *Naive Bayes* y Regresión Logística, utilizando técnicas de procesamientos de datos que incluían la transformación de variables logro tener una puntuación de f1-score superior al 95% en los modelos descritos, concluyendo que los modelos de *Random Forest* y las Redes Neuronales tienen la mejor eficacia para la detección de fraudes en tiempo real.

En Argentina, (Álvarez, 2020)en su propuesta de "*Machine Learning* en la detección de fraudes de comercio electrónico aplicado a los servicios bancarios", utilizo una metodología cuantitativa y comparativa, el investigador contaba con un dataset de 80,792 datos que contenía solo transacciones de autorización de tarjetas de crédito en donde 286 fueron etiquetadas con fraude, propone un sistema de detección de fraudes en donde la clasificación basada en el algoritmo de *Random Forest* predominara particularmente por su eficacia, para la construcción del modelo siguió los siguientes pasos: Revisión literaria, Selección de características, División de datset en pruebas y entrenamiento, Construcción, Pruebas y Ajustes del modelo, en la sección de validación del modelo obtuvo un nivel de exactitud del 96,14%.

En Colombia, (Alberto & Arcos, 2022) en su trabajo de titulación de "Selección de una Técnica de Aprendizaje de Máquina para la Detección de Fraude Financiero Digital Enfocado a Transacciones no Autorizadas o Consentidas", tenía como objetivo utilizar un modelo de aprendizaje de máquina para detectar fraude en transacciones bancarias utilizando los modelos de *Logistic Regression, Random Forest, Support Vector Machine y Neural Network* con una metodología de: compresión de negocio, estudio y compresión de los datos, entendimiento y obtención de los datos, análisis y selección de características, modelado y evaluación. Obteniendo como resultado que, en la detección efectiva de fraudes el modelo *Support Vector Machine* alcanzo el 38% más en detección. Concluyendo que SVM es el modelo más apropiado para resolver problemas de clasificación.

En Cuba, (Ameijeiras Sánchez et al., 2021) realizó la siguiente investigación "Algoritmos de detección de anomalías con redes profundas. Revisión para detección de fraudes bancarios", el

investigador tenía como objetivo analizar los principales algoritmos de detección de anomalías basados en aprendizaje profundo, utilizó una metodología de revisión literaria respecto a la aplicación de los modelos redes profundas en la detección de fraudes financieros en tarjetas de crédito/débito, los modelos que estudio fueron: Redes neuronales artificiales, Convulucionales, *Autoencoders*, detallando que el último modelo descrito es el que tiene mayor efectividad para rastrear transacciones bancarias fraudulentas, concluyendo que el aprendizaje profundo es la solución para la detección de fraudes bancarios.

En España, el trabajo de titulación de (Dueñas Quesada, 2020) fue "Aplicación de técnicas de *Machine Learning* a la ciberseguridad: Aprendizaje supervisado para la detección de amenazas web mediante clasificación en árboles de decisión", tenía como objetivo clasificar peticiones http entre normales y anómalas utilizando el modelo de Árbol de Decisión, la metodología usando fue cuantitativa, utilizo un dataset con 97,065 datos, para la elección de las características más relevantes utilizo la selección automática basada en varianza de los datos, concluyendo que el modelo de árbol de decisión ayudo a detectar amenazas web en un 100% de efectividad.

En Colombia, (Pérez González, 2021)realizó un artículo llamado "Detección de transacciones fraudulentas en tarjetas de crédito mediante el uso de modelos *Machine Learnin*g" en donde realizó una comparación de precisión exactitud entre los modelos de *Oneclass Support Vector Machine, Isolation Forest* y *Autoencoder*, utilizo una metodología cuantitativa, el dataset utilizado en este artículo fue uno Europeo público de 284,807 transacciones reales realizadas en el 2013,en donde 492 estaban con etiqueta de fraude, la secuencia del desarrollo fueron: Revisión Literaria, Selección de algoritmos a utilizar, Validación de los modelos seleccionados. Concluyendo que el modelo que mejor se desempeña minimizando los falsos positivos es el *Isalation Forest*, la investigación se realiza a partir del trabajo de titulación de (Fernández Khatiboun, 2019), en donde concuerda que *Isalation Forest*, tiene mayor efectividad, el articulo proporciona los mecanismos a tomar en cuenta para poder evaluar la eficiencia de los modelos que propongamos.

En España, (Fernández Khatiboun, 2019) realizó el siguiente trabajo de titulación "*Machine Learning* en la Ciberseguridad" donde indica en unos de sus objetivos la utilización de algoritmo *Isolation Forest* para el entrenamiento del modelo, utilizo un dataset de 284,807, las fases que utilizo para realizar el trabajo fueron: Obtención de datos, Procesamiento de datos, Selección de características, Entrenamiento del modelo, Realización de pruebas, Análisis de resultado. Concluyendo que *Isolation Forest* es mejor en comparación con *Random Forest* teniendo una tasa de precisión de 90% vs el 86%.

1.2. Desarrollo teórico y conceptual

Sistema financiero

El Sistema Financiero es básicamente un mecanismo sano de distribución de fondos. La función básica de las entidades financieras es la intermediación entre usuarios– superavitarios y deficitarios- a través de un trabajo técnico que permita administrar, de forma correcta, los riesgos inherentes a esta actividad (Banco Internacional, 2021).

Servicios financieros

Los servicios financieros comprenden todo servicio de esa naturaleza, bien sea servicio de banca, seguros, valores, factoraje, arrendamiento financiero y finanzas; y cualquier otro servicio conexo o auxiliar de uno financiero (Quintana Adriano, 2004).

Inclusión financiera

La inclusión financiera se refiere al acceso que tienen las personas y las empresas a diversos productos y servicios financieros útiles y asequibles que atienden sus necesidades —transacciones, pagos, ahorro, crédito y seguros— y que se prestan de manera responsable y sostenible (Banco Mundial, 2022).

Canales electrónicos

Los canales electrónicos son las vías o formas por las que los clientes y/o usuarios pueden hacer transacciones con las entidades controladas, usando elementos o dispositivos electrónicos o tecnológicos, utilizando tarjetas. Principalmente, son canales electrónicos: los cajeros automáticos (ATM), dispositivos de puntos de venta (POS y PIN Pad), sistemas de audio respuesta (IVR), banca electrónica, banca móvil, u otros mecanismos electrónicos similares (Superintendencia de Bancos, 2021).

Medios electrónicos

Los Medios Electrónicos son los elementos de la tecnología que tienen características digitales, magnéticas, inalámbricas, ópticas, electromagnéticas u otras similares (Superintendencia de Bancos, 2021).

Medios de pago

Los medios de pago son los activos que pueden servir para cancelar una deuda pendiente (Corporación Financiera Nacional, 2018).

Tarjeta de debito

Las tarjetas de débito son tarjetas bancarias que sirve para manejar el dinero disponible en la cuenta a la que está vinculada, a través de retiros desde cajeros automáticos y pagos en establecimientos (Banco Pichincha, 2022).

Transacción

Las transacciones son flujo económico que refleja creación, transformación, intercambio, transferencia o extinción de un valor económico y entraña traspasos de propiedad de bienes o activos financieros, prestación de servicios o suministro de mano de obra y capital (Instituto Nacional de estadísticas y Censo, 2019).

E-commerce

El *e-commerce* o comercio electrónico consiste en la distribución, venta, compra, marketing y suministro de información de productos o servicios a través de Internet (Visa, 2023a). Jeffey Rayport en su libro de *e-commercer,* define al comercio electrónico como "intercambios mediados por la tecnología entre diversas partes (individuos, organizaciones o ambos), así como las actividades electrónicas dentro y entre organizaciones que facilitan esos intercambios" (referenciar).

Fraude

La real academia española define al fraude como una "*acción contraria a la verdad, que perjudica a la persona contra quien se comete*" (Real Academia Española, 2023), como también a personas jurídicas, con el fin de obtener beneficio injusto, en el ámbito financiero ocurre un tipo de fraude relaciona a transacciones electrónicas no consentidas.

Fraude bancario

El fraude bancario afecta a miles de personas en todo el mundo. Causa pérdidas a diario a personas y empresas por igual (Diners Club, 2022).

Fraude electrónico

Fraude electrónico se realiza utilizando el internet, teléfono, equipos informáticos o sistemas de comunicación para obtener un beneficio no autorizado (Álvarez, 2020), con el fin de recabar la mayor cantidad de información personal de los usuarios, en donde el *phishing* es el tipo de fraude electrónico más utilizado.

Fraude en comercio electrónico

Fraude en comercio electrónico se produce cuando el usuario realiza compra – venta de bienes o servicios por medio de internet utilizando sus tarjetas de crédito o débito (Gobierno de México, 2017); en sitios donde no se tiene la información de contacto del vendedor o medios que validen la legitimidad de la página web.

Phishing

Phishing consiste en él envió de correos electrónicos o mensajes de texto con enlaces a páginas web con virus, en donde proceden a obtener información personal o bancaria de los usuarios que accidentalmente den clip a dichos enlaces (Diners Club, 2022)

skimming

Hurtar información de las tarjetas de crédito o débito para clonarlas y realizar compras a comercios electrónicos por medio de transacciones, este delito es conocido como fraude de tipo *skimming* (Diners Club, 2022).

Algoritmo

Secuencia de pasos ordenados de operaciones, funciones o procedimientos que permiten hallar la solución a un problema (Real Academia Espanola, 2023).

Machine Learning

El aprendizaje de máquina es la ciencia que hace que los ordenadores aprendan a partir de datos (Bobadilla, 2021), dependiendo del tipo de problema que se desea abordar se puede utilizar unos de las clasificaciones de machine learning: Aprendizaje supervisado, no supervisado o aprendizaje reforzado (Mirjalili & Rasch, 2020).

Aprendizaje Supervisado

En el aprendizaje supervisado se utiliza datos etiquetados (Bobadilla, 2021), por ejemplo, el producto de dos variables proporciona un valor “y, es valor de “y” se conoce y se lo etiqueta, los algoritmos que se utilicen se entrenaran con los datos etiquetados y buscan patrones en base a esos datos.

Aprendizaje no supervisado

El aprendizaje no supervisado es lo contrario al aprendizaje anterior, se utiliza para encontrar patrones y estructuras en conjunto de datos sin necesidad de disponer de las etiquetas (Borràs García & Caballero, 2023)

Tipos de algoritmos de *Machine Learning*

Regresión Logística

La regresión logística según (Dietterich et al., 2022) es un modelo de clasificación discriminativo ampliamente utilizado

$$p(y|x;\theta)$$

Ecuación 1 Probabilidad de que ocurra el evento

- donde x ∈ R^D es un vector de entrada de dimensión fija
- y ∈ {1, ..., C} es la etiqueta de clase
- θ son los parámetros del modelo.

Si C = 2, esto se conoce como regresión logística binaria, y si C > 2, se conoce como regresión logística multinomial. regresión logística o, alternativamente, regresión logística multiclase.

Regresión Logística Binaria

(Dietterich et al., 2022) comenta que, la regresión logística binaria corresponde al siguiente modelo

$$p(y|x;\theta) = Ber(y|\sigma(wTx + b))$$

Ecuación 2 Regresión Logística Binaria

- donde σ es la función sigmoidea
- w son los pesos
- b es el sesgo y θ = (w, b) son todos los parámetros.

En otras palabras

$$p(y = 1 \mid x;\theta) = \sigma(a) = \frac{1}{1 + e^{-a}}$$

Ecuación 3 Función Sigmoide σ

- donde a = wTx + b son las probabilidades logarítmicas,
- log(p/1 − p)
- donde p = p(y = 1|x; θ),

Máquina de Vector de Soporte (SVM)

SVM es un algoritmo para aprender semiespacios con un cierto tipo de conocimiento previo, es decir, preferencia por un margen grande (Hastie et al., 2008).

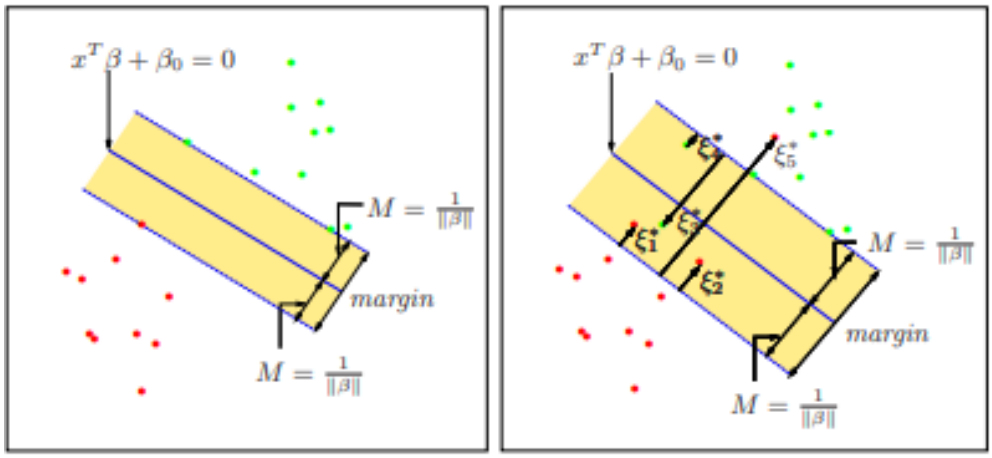

Ilustración 1 Clasificadores de Vectores de Soporte (Hastie et al., 2008)

Con base a la ilustración 1 (Hastie et al., 2008) define la formula siguiente:

$$yi(x_i^T\beta + \beta_0) \geq M(1 - \xi_i)$$

Ecuación 4 Clasificador de Vectores de Soporte "Estándar"

Donde:

- yi => Representa la etiqueta o clase. En clasificación binaria, suele ser 1 para la clase positiva y 0 para la clase negativa.
- xTi => Representa el producto punto entre el vector de características xi y el vector de pesos β.
- β => Representa el vector de pesos del modelo que se aprende durante el entrenamiento.
- β0 => Representa el término de intercepción del modelo, que es un valor constante añadido a la combinación lineal de las características.
- M => Representa el margen de separación entre las clases positiva y negativa. (llamados vectores de soporte).
- ξi => Representa la holgura. Es una variable no negativa que permite cierta flexibilidad en la clasificación, ya que puede haber puntos que se encuentren cerca o incluso dentro del margen de separación.

Bosque Aleatorio *(Ramdon forest)*

Un bosque aleatorio está definido por (Shai Shalev & Shai Ben, 2014) como un clasificador que consta de una colección de árboles de decisión, donde cada árbol se construye aplicando un algoritmo A en el conjunto de entrenamiento S y un vector aleatorio adicional, θ.

La predicción del bosque aleatorio se obtiene por mayoría de votos sobre las predicciones de los árboles individuales. (Shai Shalev & Shai Ben, 2014) comenta que, para especificar un bosque aleatorio particular, se necesita definir el algoritmo A y la distribución sobre θ.

(Shai Shalev & Shai Ben, 2014) genera θ de la siguiente manera:

- Primero, escoge una submuestra aleatoria. de S con reemplazos; es decir, toma muestras de un nuevo conjunto de entrenamiento S' de tamaño m' usando la distribución uniforme sobre S.
- En segundo lugar, construye una secuencia $I_1, I_2, \ldots$, donde cada I_t es un subconjunto de [d] de tamaño k, que se genera mediante muestreo uniforme en elementos aleatorios de [d]. Todas estas variables aleatorias forman el vector θ.
- Luego, el algoritmo A genera un árbol de decisión basado en las muestras S', donde en cada etapa de división, el algoritmo es restringido a elegir una característica que maximice la ganancia del conjunto I_t. Intuitivamente, si k es pequeño, esta restricción puede evitar el sobreajuste.

Otros Algoritmos

Según la naturaleza del problema se puede utilizar los siguientes tipos de algoritmos de *Machine Learning*.

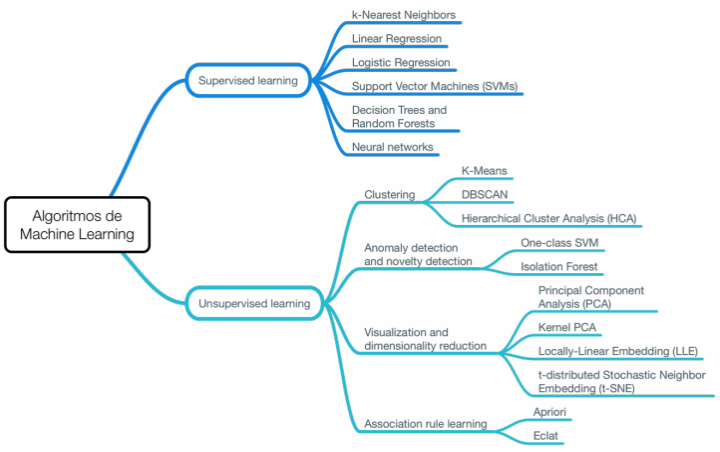

Ilustración 2 Clasificación de algoritmos de Machine Learning (Dueñas Quesada, 2020)

Evaluación de Algoritmos de Aprendizaje de Máquina

Para la evaluación de los algoritmos de *Machine Learning* se utiliza la matriz de confusión en donde (Pérez González, 2021) detalla lo siguiente:

TN (Verdaderos Negativos)	**FP** (Falsos Positivos)
FN (Falsos Negativos)	**TP** (Verdaderos Positivos)

Tabla 1 Matriz de Confusión

Si la métrica de sensibilidad o *Recall* tiene su valor lo más cercano a 1 indica que se detectaron correctamente las transacciones fraudulentas, si su valor es cercano a 0 indica que no las detecta correctamente, su fórmula es la siguiente:

$$sensibilidad(R) = \frac{TP}{(TP + FN)}$$

Ecuación 5 Fórmula de Sensibilidad

La métrica de precisión indica el porcentaje de transacciones fraudulentas detectadas que efectivamente son de esta categoría, la fórmula es la siguiente:

$$\text{precisión(P)} = \frac{TP}{(TP + FP)}$$

Ecuación 6 Fórmula de Precisión

La métrica de exactitud o *Accuracy* indica que tanto las transacciones fraudulentas como las legítimas fueron clasificadas correctamente, la fórmula es la siguiente:

$$exactitud = \frac{(TN + TP)}{(TN + FP + TP + FN)}$$

Ecuación 7 Fórmula de Exactitud

La métrica de especificidad indica que se identificó correctamente las transacciones legítimas, la fórmula se muestra a continuación.

$$\text{especificidad} = \frac{TN}{(TN + FP)}$$

Ecuación 8 Fórmula de Especificidad

La métrica de F1-score indica una media armónica entre la sensibilidad y la precisión, esta es la métrica que se tomara en cuenta al momento de seleccionar un modelo, debido a que hay un porcentaje mínimo de transacciones fraudulentas en comparación con las transacciones legítimas y se necesita tener equilibrio entre detectar correctamente tanto las transacciones legítimas como las fraudulentas.

$$f1score = 2 * \frac{P * R}{(P + R)}$$

Ecuación 9 Fórmula de F1-Score

Marco normativo

Actualmente el Ecuador impone pena máxima a las personas que realicen fraudes financieros, Código Orgánico Integral Penal en el artículo 186 literal 1 y 2 detalla

1. *Defraude mediante el uso de tarjeta de crédito, débito, pago o similares, cuando ella sea alterada, clonada, duplicada, hurtada, robada u obtenida sin legítimo consentimiento de su propietario*. (Ecuador, 2018)

2. *Defraude mediante el uso de dispositivos electrónicos que alteren, modifiquen, clonen o dupliquen los dispositivos originales de un cajero automático para capturar, almacenar, copias o reproducir información de tarjetas de crédito, débito, pago o similares*. (Ecuador, 2018)

Una gestión temprana en las variables que pueden causar riesgos operativos en las instituciones financieras puede evitar caídas de los servicios financieros, la Organización Internacional de Normalización (ISO) en la norma 31000:2018 referente a la gestión de riesgos (Grupo ISO/TC 262/STTF, 2017), determina los lineamientos a seguir ante los posibles riesgos operativos que puedan aparecer en las instituciones públicas o privadas.

La Superintendencia de Economía Popular y Solidaria (SEPS) en la resolución No. SEPS-IGT-IGS-INR-INGINT-2022-0211 indica que la administración del riesgo operativo de las instituciones financieras debe considerar las pérdidas derivadas de la ocurrencia de eventos externos como los fraudes, manteniendo procedimientos a seguir con la finalidad de garantizar la operatividad continua y sin interrupciones del negocio (Superintendencia de Economía Popular y Solidaria, 2022). La SEPS en la resolución 103 de la sección II sobre las medidas tecnológicas de seguridad en el uso de transferencias electrónicas en el artículo 4 ítem, Sistemas de transferencia electrónica estable medidas a seguir para maximizar la seguridad en los servicios electrónicos como mecanismo que permitan reconocer la validez de las transferencias realizadas (Superintendencia de Economía Popular y Solidaria, 2017).

CAPÍTULO 2. METODOLOGÍA

2.1. Contexto de la investigación

Como caso de estudio se escogió a la cooperativa de ahorro y crédito Lucha Campesina que se encuentra ubicada en la provincia de Chimborazo, la sede principal se encuentra cantón Cumandá de la misma provincia, tiene ocho sucursales en Bucay, Naranjito, Milagro, El Triunfo, Simón Bolívar en la provincia del Guayas, La Troncal en la provincia del Cañar y Baba, Vinces en la provincia de los Ríos. Aproximadamente tiene 58.000 socios entre las 9 agencias, Cumandá, Bucay, Naranjito, Simón Bolívar, Baba, El Triunfo, La Troncal, Milagro y Vinces (Lucha Campesina, 2023), el estimado de transacciones realizadas al día son de 28.000, el 40% es realizado por compra - ventas en internet(*e-commece*r), con la autorización de gerente general de la cooperativa(ver anexo autorización), se logró obtener un histórico de las transacciones realizadas en el año 2023, con el cual se obtuvo el dataset que se va a utilizar para entrenar el modelo de aprendizaje de máquina para detectar fraudes financieros.

Ilustración 3 Agencias Lucha Campesina – Ecuador

2.2. Diseño y alcance de la investigación

Para la detección de fraudes en transacciones electrónicas *e-commerce* en la Cooperativa de Ahorro y Crédito Lucha Campesina, se adoptará un enfoque de investigación no experimental de tipo transversal con un diseño descriptivo. Al tener acceso a datos históricos para formar un dataset con transacciones electrónicas fraudulentas y legítimas se puede realizar una investigación no experimental debido a que no estamos construyendo ninguna situación, partimos el estudio desde los datos obtenidos, observando los fenómenos tal y como ocurren en su contexto natural (Hernández Sampieri et al., 2006), analizando las variables o casusa que desencadenaron el fraude. El alcance de la investigación abarcará un análisis exhaustivo de datos históricos de transacciones electrónicas realizadas entre los meses de abril y junio del 2023. El diseño transversal permitirá examinar las características de las transacciones fraudulentas y no fraudulentas en el periodo mencionado, y mediante técnicas estadísticas descriptivas, se analizará variables como el monto de la transacción, la frecuencia de compra, el tipo de comercio y la ubicación geográfica. Los hallazgos de esta investigación permitirán informar el desarrollo de estrategias en prevención de fraudes con la utilización de modelos de aprendizaje de máquina para detectar de una forma temprana los posibles fraudes en transacciones electrónicas *e-commerce*.

2.3. Tipo y métodos de investigación

La detección de fraudes financieros en el ámbito del comercio electrónico es un desafío creciente, además, los estafadores utilizan cada vez métodos más sofisticados para realizar transacciones fraudulentas, lo que dificulta su identificación en el proceso de análisis.

En este estudio, se propone un enfoque cuantitativo para la detección de fraudes financieros basado en el análisis de las transacciones electrónicas. Este enfoque implica la recopilación y análisis de datos, de las transacciones realizadas por los clientes de la cooperativa de ahorro y crédito Lucha Campesina entre los meses de abril y junio del 2023.

El análisis de datos se realizará utilizando las siguientes técnicas:

Análisis descriptivo: Se utilizará para describir las características básicas de las transacciones, como el monto, la frecuencia, el tipo de transacción, el lugar de la transacción y el código del comercio.

Análisis de regresión: Se utilizará para identificar las variables que están relacionadas con la probabilidad de fraude. Por ejemplo, se puede analizar si existe una relación entre el monto de la transacción y la probabilidad de fraude.

2.4. Población

La población a utilizar para detectar fraudes en tarjetas de débito de la Cooperativa de Ahorro y Crédito Lucha Campesina consistiría en todas las transacciones realizadas por el canal *e-commerce* en el periodo 2023 por las 8 agencias de la cooperativa.

Tipo	**Cantidad**
Legítimas	11000
Fraudulentas	303
Total	11303

Tabla 2 Total de transacciones fraude y no fraude

Es importante indicar que no se calcularán muestras, debido a que todos los datos serán tomados en consideración para el estudio a realizar.

2.5. Técnicas e instrumentos de recolección de datos

Las nueve agencias de la cooperativa de ahorro y crédito Lucha Campesina ofrecen a sus clientes la posibilidad de realizar compras online por medio de aplicaciones web utilizando sus tarjetas de débito. Estos procesos generan registros que la cooperativa almacena en sus bases de datos, creando una fuente de información para el análisis y la detección de fraudes.

En esta investigación, se utilizó la técnica de recolección de datos basada en el histórico de transacciones bancarias. Se extrajeron los registros con fecha de corte de abril a junio del 2023, conformando un conjunto de datos (dataset) que contiene variables como:

- Número de la cuenta
- Código de transacción
- Monto de la transacción
- Código del comercio
- Fecha y hora de la transacción
- Monto de la transacción
- Identificación del cliente
- País de transacción
- Lugar de la transacción
- Identificación del cliente
- Tipo de tarjeta (PosEntryMode)
- Tipo de transacción (PosConditionCode)

El dataset se utilizará como instrumento para entrenar los modelos de aprendizaje automático con el objetivo de detectar transacciones fraudulentas en las tarjetas de débito de la cooperativa Lucha Campesina.

2.6. Procesamiento de la evaluación: Validez y confiabilidad de los instrumentos aplicados para el levantamiento de información.

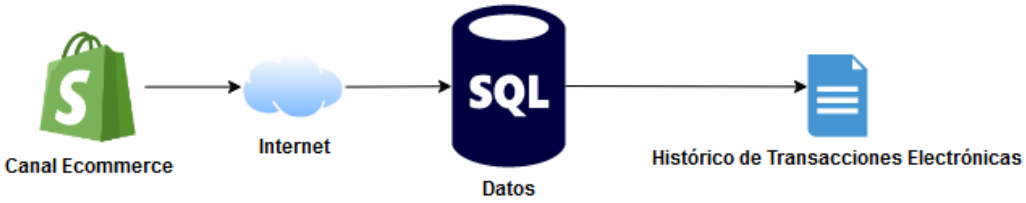

Ilustración 4 Transacciones canal E-commerce (Elaboración Propia)

Los clientes de las nueve agencias que tiene cooperativa de ahorro y crédito Lucha Campesina realizan compras *e-commerce*, dicho proceso generan transacciones que son almacenadas en un histórico, la información bancaria de la cooperativa es confidencial y esta custodiada por el departamento de seguridad de la información en conjunto con el área de tecnología, para que una persona o entidad tenga acceso a la información debe pasar por un proceso de autorización por parte de la Gerencia General, para el estudio a realizar se solicitó autorización (Ver anexo 1 Autorización), para acceder a las transacciones realizadas en los meses de abril a junio del 2023.

Cabe indicar que, las transacciones que ingresan a la cooperativa por el canal del aplicativo móvil y que tiene como objetivo realizar un débito a la cuenta del cliente deben pasar por un proceso de validación automática donde el cliente autoriza dicho proceso, en cambio, las transacciones que tiene origen de compras por internet, para realizar el proceso solo necesitan información de la tarjeta de débito del cliente en donde la cooperativa almacena el registro de dicha operación, si el cliente indica que no ha realizado dicha operación se inicia un proceso de validación con el objetivo de determinar si la transacción es legítima o fraudulenta. La cooperativa de Ahorro y Crédito Lucha Campesina nos entrega un dataset de transacciones electrónicas que contiene operaciones legítimas y fraudulentas.

CAPÍTULO 3. RESULTADOS Y DISCUSIÓN

Procesamiento de Datos

El dataset de transacciones *e-commerce* proporcionado por la cooperativa Lucha Campesina, tiene los siguientes campos:

```
Index(['numeroCuenta', 'codigoTransaccion', 'valorTransaccion', 'mensaje',
       'codigoComercio', 'ChannelId', 'RealDate', 'TerminalId', 'MsgType',
       'AccountId1', 'ChannelId.1', 'EntidadOrigen', 'ChannelName',
       'esCompraPos', 'redOrigen', 'TipoComercio', 'NumIdentificacion', 'Pais',
       'LugarTran', 'TieneCosto', 'ValorParcial', 'Impreso', 'ActionCode',
       'PosEntryMode', 'PosConditionCode', 'isFraude'],
      dtype='object')
```

Ilustración 5 Campos del Dataset de transacciones E-commercer

Al conjunto de datos se procedió a eliminar las columnas que contenían valores categóricos de una solo dimensión, debido a que no aportan información relevante para el modelo y podrían causar la detección de fraudes erróneos o inferir al rendimiento del modelo.

```
numeroCuenta          1512
codigoTransaccion        2
valorTransaccion       952
mensaje                  9
codigoComercio          91
ChannelId                1
RealDate             11301
TerminalId             366
MsgType                  1
AccountId1               1
ChannelId.1              1
EntidadOrigen            1
ChannelName              1
esCompraPos              1
redOrigen                1
TipoComercio            56
NumIdentificacion     1512
Pais                    36
LugarTran             1222
TieneCosto               1
ValorParcial             1
Impreso                  1
ActionCode               2
PosEntryMode             5
PosConditionCode         1
isFraude                 2
```

Ilustración 6 Campos con categoría única

En la imagen anterior se observa las columnas que tiene una solo categoría

```
#Eliminacion de columnas categóricas
df.drop(['codigoTransaccion','ChannelId','ActionCode','MsgType','AccountId1','ChannelId.1',
         'EntidadOrigen','ChannelName','esCompraPos','redOrigen','TieneCosto','ValorParcial',
         'Impreso','PosConditionCode'],axis=1,inplace=True)
```

Ilustración 7 Eliminación de variables Categóricas

```
numeroCuenta          int64
valorTransaccion     object
mensaje              object
codigoComercio       object
RealDate             object
TerminalId           object
TipoComercio          int64
NumIdentificacion     int64
Pais                  int64
LugarTran            object
PosEntryMode          int64
isFraude              int64
dtype: object
```

Ilustración 8 Campos sin Categoría única

Se calculó en términos de porcentaje la cantidad de transacciones fraudulentas y legítimas que tenía el dataset, el 98,3% eran transacciones legítimas y el 2.68% estaban etiquetadas como fraudulentas.

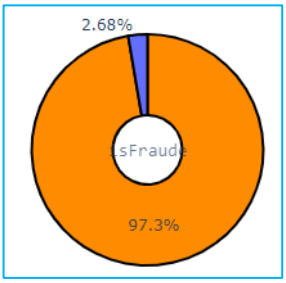

Ilustración 9 Porcentaje de transacciones Legítimas y Fraudulentas

En la ilustración 7 se observa que algunas columnas tienen como tipo de datos Objeto, las columnas en cuestión tienen valores categóricos, se procede a codificar en valores numéricos.

```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
var = ['mensaje','codigoComercio','TerminalId','TipoComercio','Pais','LugarTran','PosEntryMode','isFraude']
for i in var:
    df[i] = le.fit_transform(df[i])
```

Ilustración 10 Codificación de Variables Categóricas

La columna RealDate también tiene como tipo de datos Objeto, se procede a transformar a tipo DateTime, y la columna Valor de transacción a tipo float.

```
# Convertir a tipo DateTime
df['RealDate'] = pd.to_datetime(df['RealDate'])
# Convierte la columna 'Valor_Transaccion' de tipo string a tipo float
df['valorTransaccion'] = df['valorTransaccion'].str.replace(',', '.').str.replace('$', '').str.replace('€', '')
df['valorTransaccion'] = df['valorTransaccion'].astype(float)
```

```
numeroCuenta                 int64
valorTransaccion           float64
mensaje                      int32
codigoComercio               int32
RealDate            datetime64[ns]
TerminalId                   int32
TipoComercio                 int64
NumIdentificacion            int64
Pais                         int64
LugarTran                    int32
PosEntryMode                 int64
isFraude                     int64
dtype: object
```

Ilustración 11 Estandarización de los campos

Selección de Características

Variables Cualitativas

Luego de realizar una limpieza de los datos como eliminación de valores null, en el campo valor de transacción se cambió el separador de decimal de coma por el punto, se procedió a realizar un análisis descriptivo de las variables respecto a la columna fraude, pero antes se muestra los campos a utilizar con la información que almacenan.

Campo	**Descripción**
mensaje	El campo indica si la transacción fue aprobada, fondos insuficientes, error del sistema o cuenta bloqueada.

CódigoComercio	Indica el código del comercio electrónico, a este campo se le coloco la palabra "No información" a los que tenían valor null.
TerminalId	Es un identificador del dispositivo desde donde se está realizando la transacción.
TipoComercio	Indica el tipo de comercio
LugarTran	Indica el nombre o una descripción del lugar de la transacción.
PosEntryMode	Indica la forma que se ha utilizada la tarjeta para realizar la transacción.
País	Indica el país hacia donde se está realizando la transacción.

Tabla 3 Campos Categóricos

La siguiente descripción muestra que los mensajes con aprobado, cuenta bloqueada y fondos insuficientes tiene relación con la etiqueta fraude. En donde el mensaje con "Fondos Insuficientes" tiene el 58.81% del total de los fraudes y "Cuenta bloqueada" tiene un 25.79% y "Aprobada" solo un 20.48%.

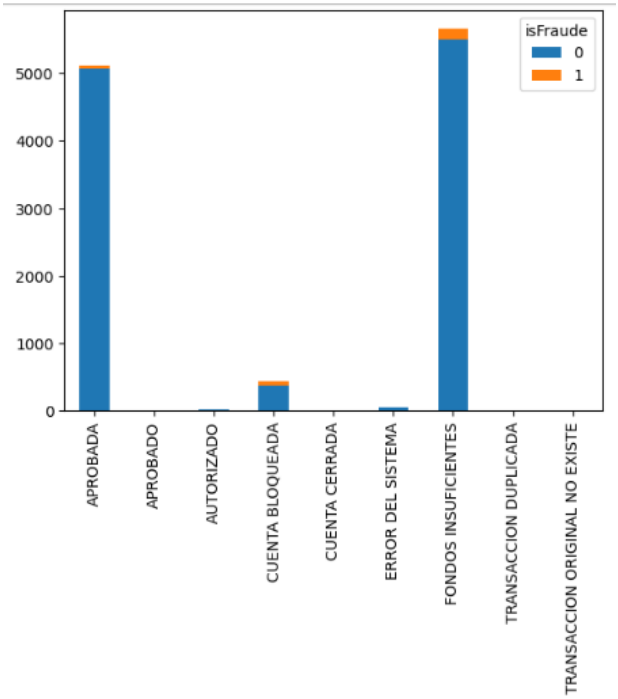

Ilustración 12 Fraude respecto a la columna Mensaje

El análisis descriptivo respecto a la columna "PosEntryMode" nos indica que fraude se realizó por el código 1000 que es de tipo tarjeta no presente

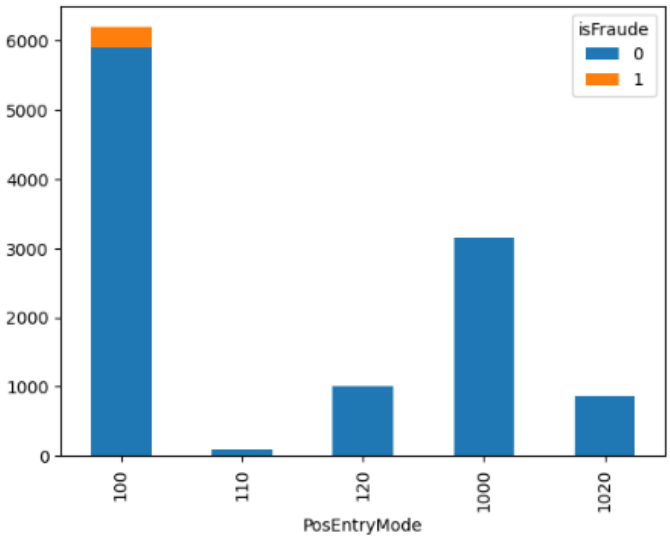

Ilustración 13 Fraude respecto a columna PosEntryMode

El análisis respecto a la columna "TipoComercio" se puede observar que se realizó con los códigos de comercio 4900 y 4814 que según (Visa, 2023b) nos indica que el código 4900 proviene de comerciantes dedicados a la generación, transmisión y/o distribución de energía eléctrica o gas u otros servicios públicos, mientras que el código 4814 proviene de comerciantes que brindan servicios de telecomunicaciones, incluido llamadas telefónicas locales o de larga distancia.

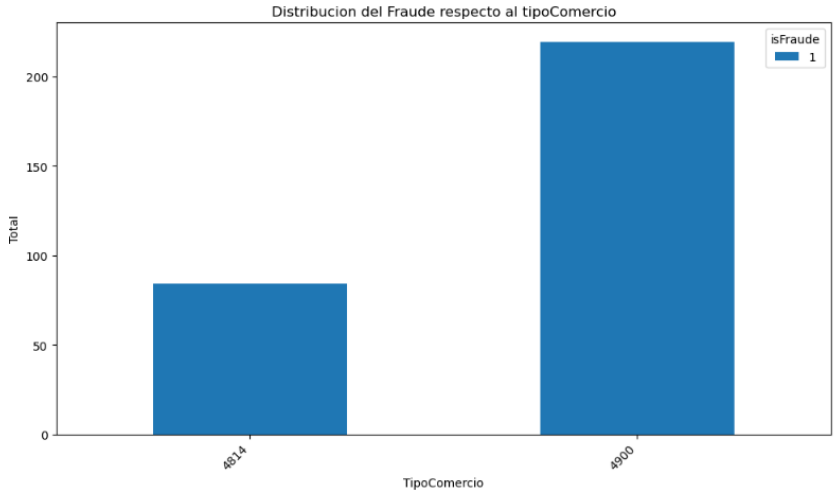

Ilustración 14 Fraude respecto a columna Tipo Comercio

Respecto a la columna "CódigoComercio" y la información anterior, el código de comercio MOVISTARRECBP, MovistarBP, ClaroBP entran en la categoría se servicios telefónicos, en

donde PAGAQUI que es un comercio de servicios públicos tiene un 2.28% de participación respecto a la columna fraude, le sigue ClaroBP con un 13.86% y los dos últimos códigos de comercio con un 6.43%.

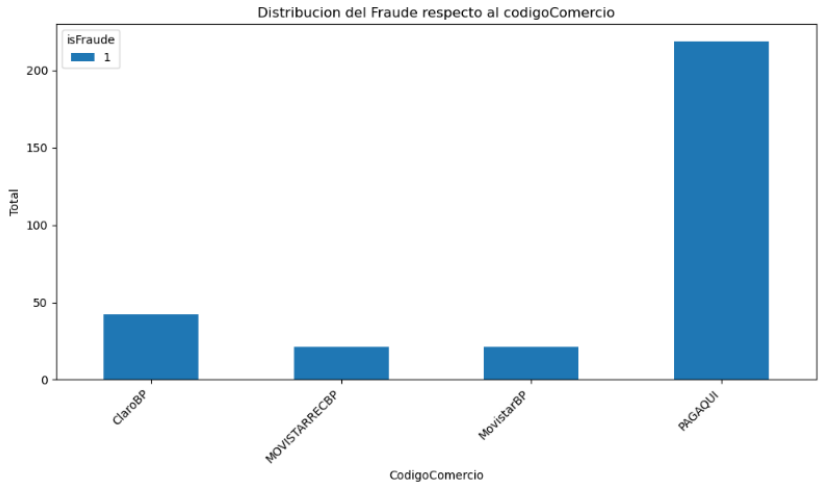

Ilustración 15 Fraude respecto a columna Código Comercio

Respecto a la columna "LugarTran" que nos indica el nombre o una descripción del comercio tiene indicio de porvenir de Santiago de Chile, se puede observar que respecto a la columna "País" que tiene código 152 en donde el (The World Bank, 2010) indica que la etiqueta pertenece al país de Chile, corroboramos y afirmamos que las transacciones de fraude provienen de dicho país.

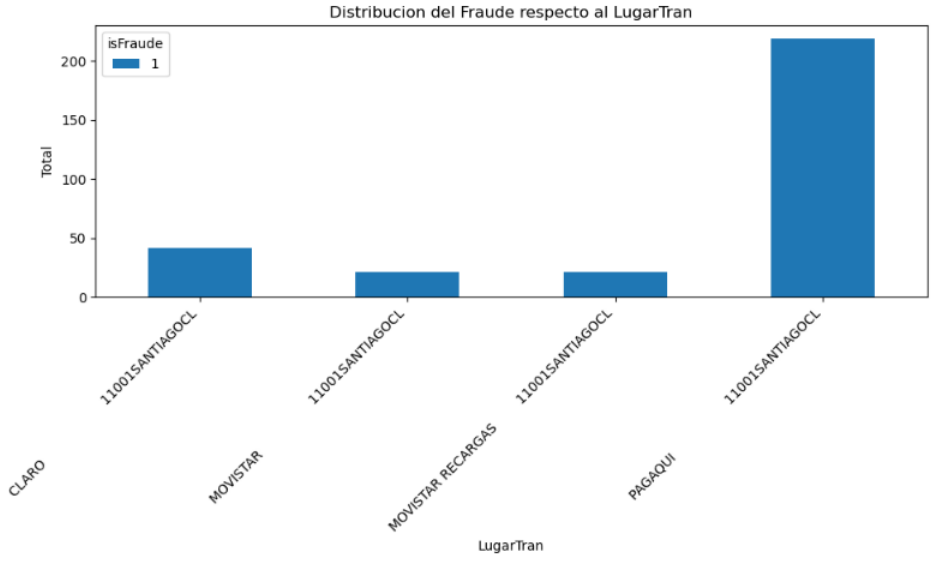

Ilustración 16 Fraude respecto a columna LugarTran

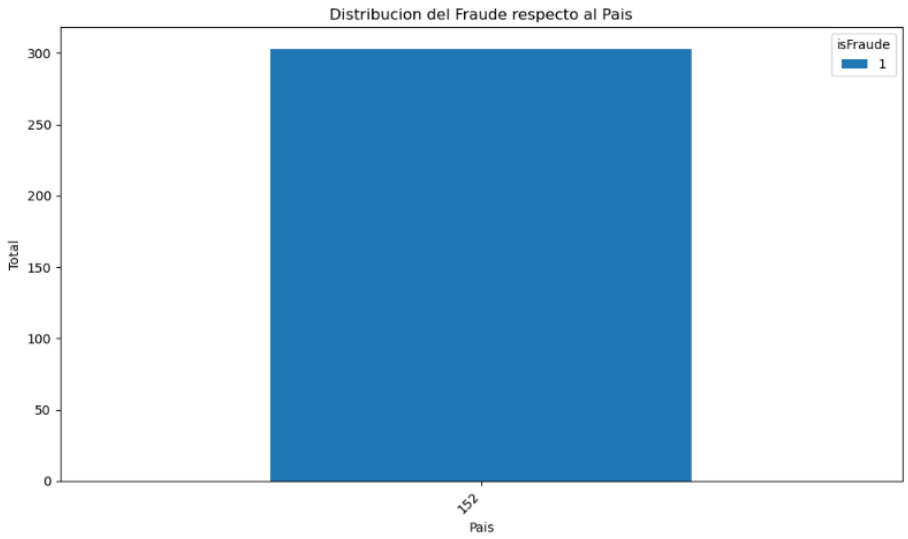

Ilustración 17 Fraude respecto a columna País

La columna "TerminalId" indica un valor unico del dispositivo desde donde se realiza la transaccion, si es realizada desde un cajero local, el valor es Cajero01, al tener un valor por defecto de 999999999, nos indica que el medio transaccional no esta especificado.

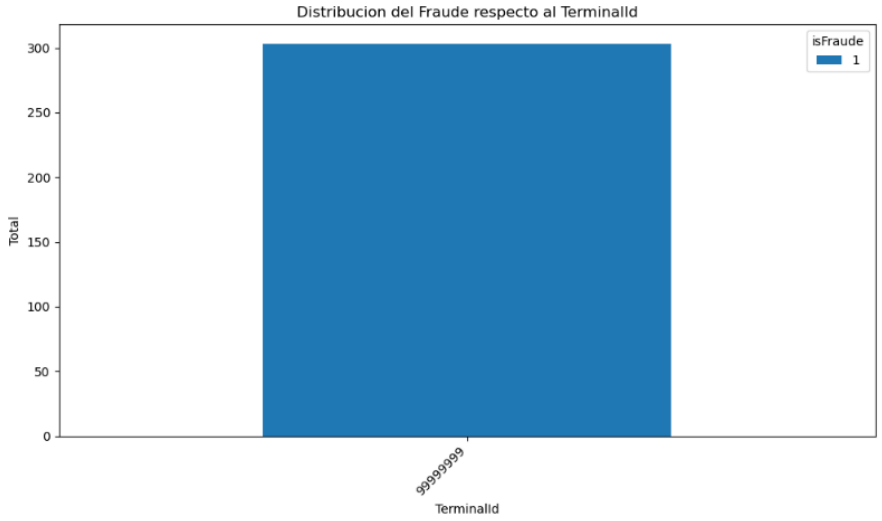

Ilustración 18 Fraude respecto a columna TerminalId

Variables Cuantitativas

Para las variables cuantitativas se utilizó el valor de la transacción y partir de la fecha de la operación se calculó año, mes, día, hora, minuto y segundo de la transacción, posteriormente se realizó un procedimiento para calcular las frecuencias de las transacciones en base al historial del cliente, calculando la frecuencia de operaciones que realiza mensualmente, frecuencia al mismo comercio y la frecuencia al mismo comercio en menos de un minuto, al terminar los cálculos se

procedió a eliminar la columna "RealDate", campo que indicaba fecha y tiempo de cuando se realizó la transacción.

```
from datetime import timedelta

# Extracción de año, mes, día, hora, minuto y segundos de la columna RealDate
df['transactionDateTime_year'] = df['RealDate'].dt.year
df['transactionDateTime_month'] = df['RealDate'].dt.month
df['transactionDateTime_day'] = df['RealDate'].dt.day
df['transactionDateTime_hour'] = df['RealDate'].dt.hour
df['transactionDateTime_minute'] = df['RealDate'].dt.minute
df['transactionDateTime_second'] = df['RealDate'].dt.second

df['frecuencia_trans_mes'] = df.groupby(['numeroCuenta', df['RealDate'].dt.month])['RealDate'].transform('count')
df['frec_trans_mism_comerc'] = df.groupby(['numeroCuenta', 'TipoComercio', df['RealDate'].dt.year, df['RealDate'].dt.month,
                                           df['RealDate'].dt.day])['RealDate'].transform('count')

# Ordenar el conjunto de datos por 'numeroCuenta', 'TipoComercio' y 'RealDate'
df.sort_values(by=['numeroCuenta','RealDate'])
dff = df.groupby(['numeroCuenta', 'TipoComercio', df['RealDate'].dt.day])['RealDate'].diff()
menos_de_un_minuto = dff < timedelta(minutes=1)
df['Transacciones_Mismo_Dia_Menos_1_Min'] = df.groupby(['numeroCuenta','TipoComercio', df['RealDate'].dt.year, df['RealDate'].dt.month,
                                                        df['RealDate'].dt.day])['TipoComercio'].transform(lambda x: x[menos_de_un_minuto].count())
```

Ilustración 19 Extracción de características de la columna RealDate

La característica calculada para conocer el número de transacciones que el cliente realiza mensualmente y realizar un análisis respecto a la columna fraude, se puede observar que el mayor porcentaje está en los clientes que tenían una frecuencia de una transacción por mes, 77.23%, y para los clientes que tiene una frecuencia de 11 operaciones por mes solo tuvo un caso de fraude.

frecuencia_trans_mes	1
1	234
2	29
3	15
4	12
5	4
6	5
9	3
11	1

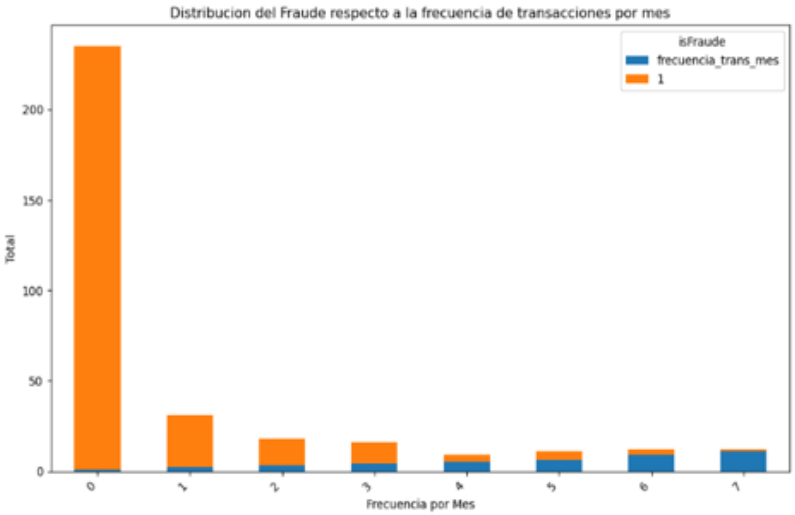

Ilustración 20 Característica Calculada - Frecuencia Transacción por Mes

Se calculó la frecuencia de transacciones que se realizaron al mismo comercio, encontrando que un 91.42% de casos de fraude ocurrió con los clientes que tenían una frecuencia de una operación por mes.

frec_trans_mism_comerc	1
1	277
2	18
4	8

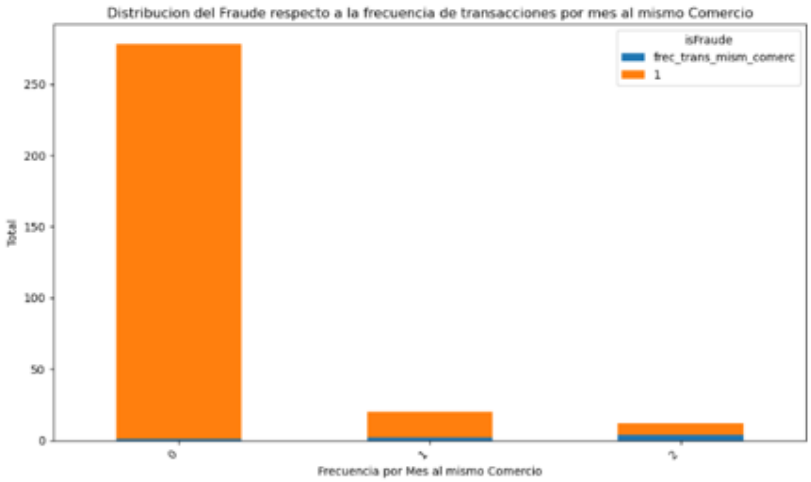

Ilustración 21 Característica Calculada - Frecuencia Transacción por Mes mismo Comercio

La característica calculada referente a las transacciones que se realizaron al mismo comercio en menos de un minuto, nos da conocer que el 95.38% de las operaciones afectados por el fraude no ocurrieron en dicho lapso. Del total de los fraudes, 10 operaciones tuvieron un intento de transaccionar al mismo comercio en menos de un minuto, y solo 4 operaciones tuvieron una frecuencia de 3 operaciones consecutivas en menos de un minuto.

Transacciones_Mismo_Dia_Menos_1_Min	1
0	289
1	10
3	4

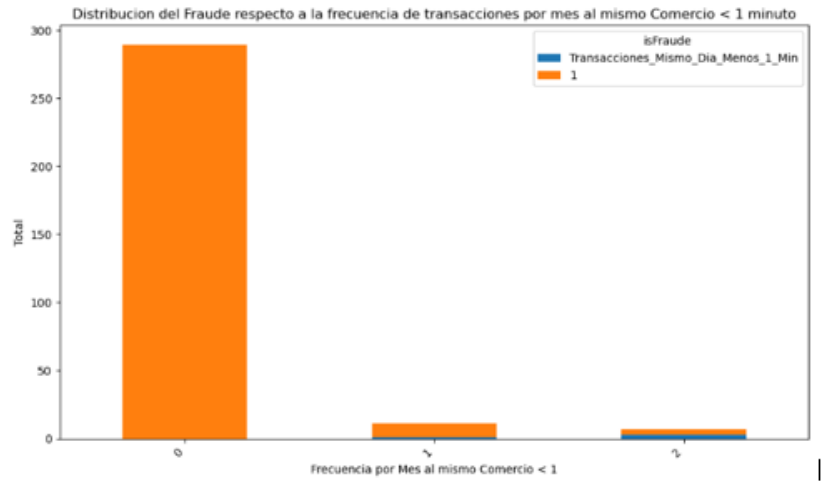

Ilustración 22 Característica Calculada - Frecuencia Transacción por Mes mismo Comercio < 1

El análisis anterior nos permite determinar la selección de las variables cualitativas y cuantitativas a usar en el entrenamiento del modelo. Las variables a utilizar deben pasar por un proceso de codificación.

Variables Cualitativas	• Mensaje • CódigoComercio • TerminalId • País • LugarTran • PosEntryMode

Tabla 4 Características - Variables Cualitativas

Variables Cuantitativas	• Valor de la transacción • Año, Mes, Día, Hora, Minuto, Segundo • Frecuencia_trans_mes • Frec_trans_mism_comerc • Transacciones_Mismo_Dia_Menos_1_Min

Tabla 5 Características - Variables Cuantitativas

Terminado el análisis descriptivo de las columnas del dataset se procedió a realizar buscar las variables que más se correlacionaban con la etiqueta fraude.

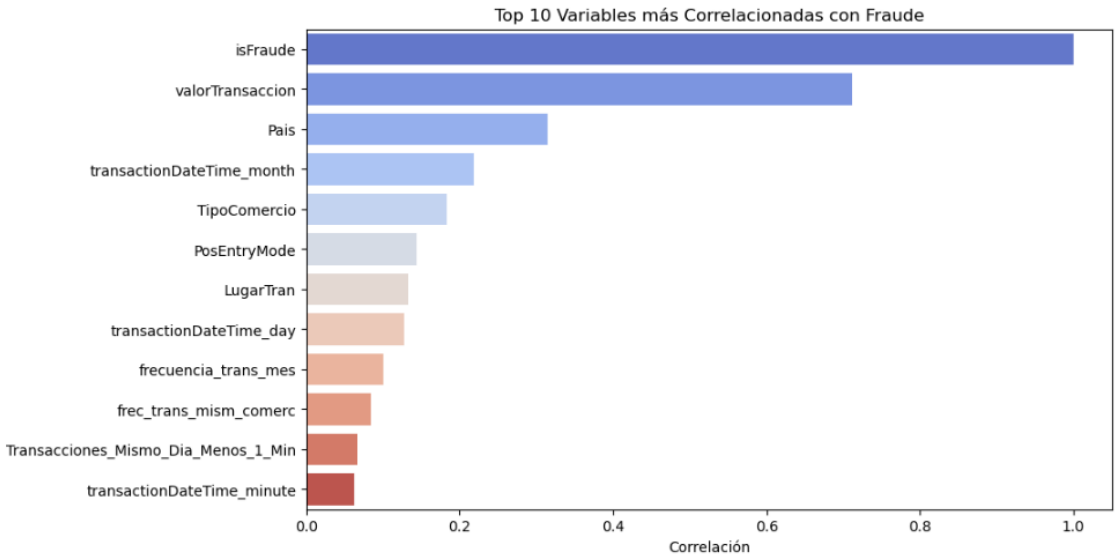

Ilustración 23 Correlación de las variables respecto a la columna Fraude

Entrenamiento de los algoritmos

Con base al análisis de correlación de los campos del dataset respecto a la columna fraude se escogió las siguientes características:

```
# Crear un nuevo dataset con las variables que mas se correlacionan con el fraude
dataTest = pd.DataFrame({
    'valorTransaccion': df['valorTransaccion'],
    'Pais': df['Pais'],
    'transactionDateTime_month': df['transactionDateTime_month'],
    'TipoComercio': df['TipoComercio'],
    'PosEntryMode': df['PosEntryMode'],
    'LugarTran': df['LugarTran'],
    'transactionDateTime_day': df['transactionDateTime_day'],
    'frecuencia_trans_mes': df['frecuencia_trans_mes'],
    'frec_trans_mism_comerc': df['frec_trans_mism_comerc'],
    'Transacciones_Mismo_Dia_Menos_1_Min': df['Transacciones_Mismo_Dia_Menos_1_Min'],
    'isFraude': df['isFraude'],
})
```

Ilustración 24 Características a utilizar para el entrenamiento del modelo

Para el entrenamiento del modelo se crea un conjunto de datos de X y otro conjunto de Y. La variable independiente (X) se elimina la columna isFraude, la variable dependiente (Y) contiene el valor que se va a predecir.

```
# Separar variables
X = dataTest.drop('isFraude', axis=1)
y = dataTest['isFraude']
```

Ilustración 25 Separación de la variable dependiente e independiente

Con el conjunto de datos (X,Y) se procede a dividir y crear conjuntos de datos de entrenamiento y prueba que se utilizaran para el aprendizaje automático y para evaluar el rendimiento del modelo. Para el entrenamiento (X_train, y_train) se escogió el 75% de los datos, para pruebas (X_test, y_test) se escogió el 25% de los datos.

```
# Dividir en conjunto de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)
```

Ilustración 26 Conjuntos de datos de Entrenamiento y Prueba

En la ilustración 8 se observa que se tiene un 97% de transacciones legítimas y un 2.68% de fraudulentas, evidenciando un desbalanceo en el conjunto de datos, se procede a crear un equilibrio entre las transacciones fraudulentas y legitimas usando el módulo SMOTE.

```
# Sobremuestreo con SMOTE
smote = SMOTE(random_state=42)
X_train_smote, y_train_smote = smote.fit_resample(X_train, y_train)
```

Ilustración 27 Balanceo de datos

Con base a una revisión bibliográfica, se seleccionaron tres algoritmos de aprendizaje automático para abordar el problema del fraude en las transacciones electrónica de *e-commerce*: Regresión Logística, Máquina de Vector de Soporte y Bosque Aleatorio. Cada uno de estos algoritmos posee características y ventajas particulares que los hacen idóneos para el problema del fraude.

La Regresión Logística se destaca por su precisión en la predicción de eventos binarios, es decir, en este estudio, detectar si una transacción es fraudulenta o no, mientras que la Máquina de Vector de Soporte es ideal para la clasificación de datos con márgenes amplios, lo que la hace útil para identificar transacciones que se desvían significativamente del comportamiento normal.

Por su parte, el Bosque Aleatorio ofrece una gran robustez y flexibilidad, siendo capaz de manejar conjuntos de datos complejos con alta dimensionalidad como es el caso de las transacciones *e-commerce* que pueden estar formadas por un gran número de características.

Regresión logística

Para determinar los valores óptimos de aprendizaje para el modelo se realizó una búsqueda de hiperparámetros.

```
# Búsqueda en grilla para optimizar hiperparámetros
grid_search = GridSearchCV(LogisticRegression(), param_grid=param_grid, cv=5)
grid_search.fit(X_train_smote, y_train_smote)
```

Ilustración 28 Regresión Logística búsqueda de Hiperparámetros

En la ilustración 25 se determinaron los mejores parámetros a utilizar para el entrenamiento del modelo.

```
# Entrenar el modelo
modelo = LogisticRegression(**grid_search.best_params_)
modelo.fit(X_train_smote, y_train_smote)
```

Ilustración 29 Entrenamiento de Regresión Logística

Evaluación del modelo Regresión Logística

En el trabajo de investigación que se ha realizado se tiene 11302 operaciones con un total de 303 casos de fraudes, al aplicar el algoritmo se obtuvo la siguiente matriz de confusión:

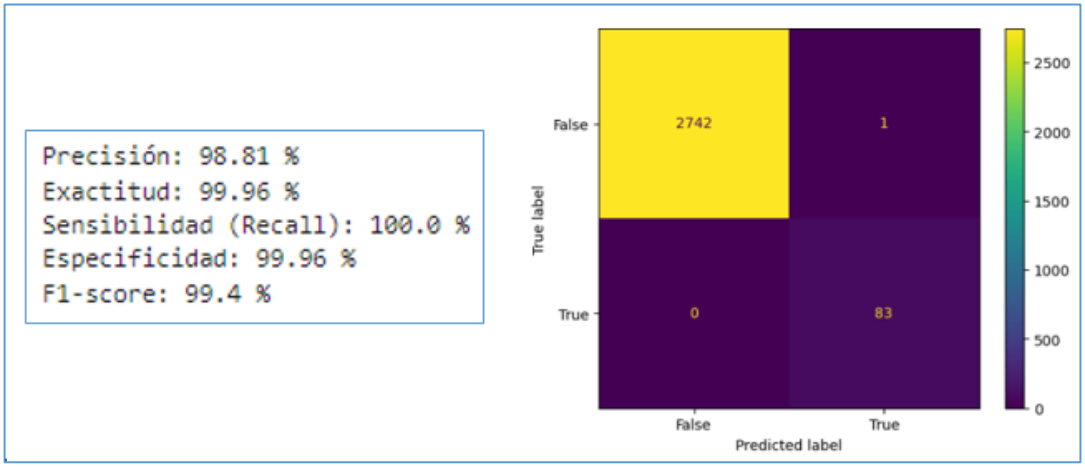

Ilustración 30 Matriz de confusión - Regresión Logística

Precisión

```
precision = precision_score(y_test, modelo.predict(X_test))
```

$$\text{precisión(P)} = \frac{83}{(83+1)} = 0{,}9881$$

Ecuación 10 Regresión logística - Precisión

Sensibilidad

```
recall = recall_score(y_test, modelo.predict(X_test))
```

$$sensibilidad(R) = \frac{83}{(83+0)} = 1$$

Ecuación 11 Regresión logística – Sensibilidad

Exactitud

```
Exactitud = accuracy_score(y_test, modelo.predict(X_test))
```

$$exactitud = \frac{(2742+83)}{(2742+1+\ 83+0)} = 0{,}9996$$

Ecuación 12 Regresión logística – Exactitud

Especificidad

```
specificity = specificity_score(y_test, modelo.predict(X_test))
```

$$\text{especificidad} = \frac{2742}{(2742+1)} = 0{,}9996$$

Ecuación 13 Regresión logística - Especificidad

F1- Score

```
f1 = f1_score(y_test, modelo.predict(X_test))
```

$$f1score = 2 * \frac{0{,}9881 * 1}{(0{,}9881+1)} = 0.994$$

Ecuación 14 Regresión logística - F1-Score

En donde se puede observar que tiene una precisión del 98,81% con una sensibilidad del 100%, al buscar un equilibrio entre la precisión y la sensibilidad se tiene un f1-score de 99.4%, con el modelo de Regresión Logística clasifica correctamente las transacciones fraudulentas y la sensibilidad del modelo nos indica que, de las transacciones fraudulentas el 100% son realmente fraudulentas.

Máquina de Vector de Soporte (SVM)

Búsqueda de hiperparametros para el modelo SVM

```
random_search = RandomizedSearchCV(SVC(kernel='rbf'), param_dist, n_iter=10, cv=5, scoring='accuracy', random_state=42)
random_search.fit(X_train_smote, y_train_smote)
best_params_random = random_search.best_params_
```

Ilustración 31 SVM búsqueda de Hiperparámetros

```
# Crear el modelo SVM
modelo_svm = SVC(kernel='rbf', C=21, gamma=0.01, random_state=42)
# Entrenar el modelo
modelo_svm.fit(X_train_smote, y_train_smote)
```

Ilustración 32 Máquina de Vector de Soporte - Entrenamiento

Evaluación del modelo Máquina de Vector de Soporte

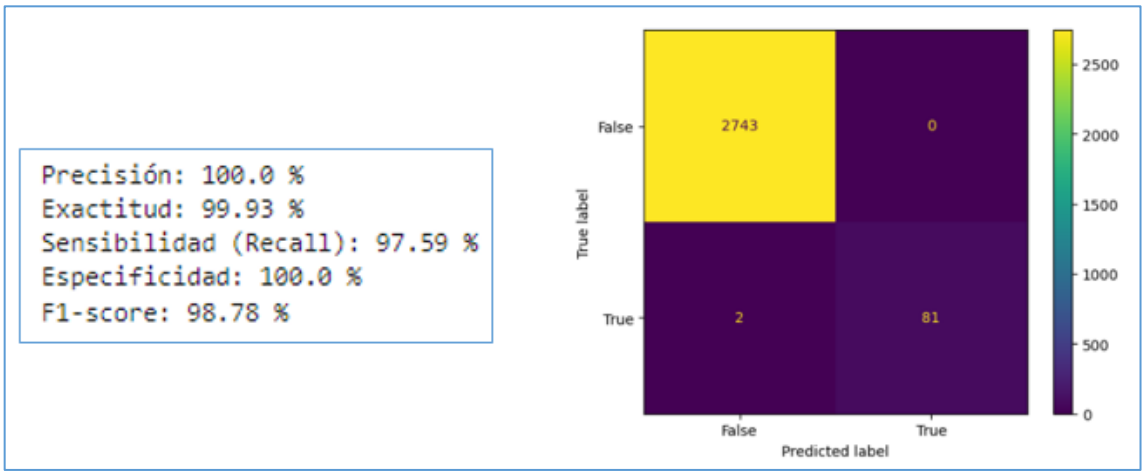

Ilustración 33 Matriz de confusión - SVM

Precisión

```
precision = precision_score(y_test, modelo_svm.predict(X_test))
```

$$\text{precisión(P)} = \frac{81}{(81+0)} = 1$$

Ecuación 15 SVM – Precisión

Sensibilidad

```
recall = recall_score(y_test, modelo_svm.predict(X_test))
```

$$sensibilidad(R) = \frac{81}{(81+2)} = 0{,}9759$$

Ecuación 16 SVM - Sensibilidad

Exactitud

```
Exactitud = accuracy_score(y_test, modelo_svm.predict(X_test))
```

$$exactitud = \frac{(2743+81)}{(2743+0+\ 81+2)} = 0{,}9993$$

Ecuación 17 SVM - Exactitud

Especificidad

```
specificity = specificity_score(y_test, modelo_svm.predict(X_test))
```

$$\text{especificidad} = \frac{2743}{(2743 + 0)} = 1$$

Ecuación 18 SVM - Especificidad

F1- Score

```
f1 = f1_score(y_test, modelo_svm.predict(X_test))
```

$$f1score = 2 * \frac{1 * 0{,}9759}{(1 + 0{,}9759)} = 0.9878$$

Ecuación 19 SVM - F1-Score

Al aplicar el modelo SVM se puede observar que tiene una precisión del 100% es decir, está detectando correctamente las transacciones legítimas, en la sensibilidad se tiene un 97.59%, es decir, tiene un margen de error del 3% para detectar correctamente las transacciones fraudulentas, al buscar un equilibrio entre la precisión y la sensibilidad se tiene un f1-score de 98.78% inferior al modelo anterior.

Bosque Aleatorio (*Random Forest*)

Se procede a crear y entrenar el modelo

```
# Crear el modelo RandomForestClassifier
rf_model = RandomForestClassifier(n_estimators=100, random_state=42)
# Entrenar el modelo con los datos de entrenamiento
rf_model.fit(X_train_smote, y_train_smote)
```

Ilustración 34 Entrenamiento de Random Forest

Evaluación del modelo *Random Forest*

```
Precisión: 98.81 %
Exactitud: 99.96 %
Sensibilidad (Recall): 100.0 %
Especificidad: 99.96 %
F1-score: 99.4 %
```

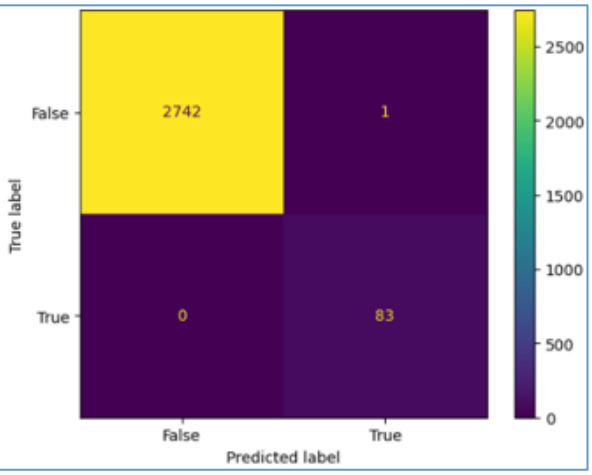

Ilustración 35 Matriz de confusión - Random Forest

Precisión

```
precision = precision_score(y_test, rf_model.predict(X_test))
```

$$\text{precisión(P)} = \frac{83}{(83+1)} = 0{,}9881$$

Ecuación 20 Random Forest – Precisión

Sensibilidad

```
recall = recall_score(y_test, rf_model.predict(X_test))
```

$$sensibilidad(R) = \frac{83}{(83+0)} = 1$$

Ecuación 21 Random Forest – Sensibilidad

Exactitud

```
Exactitud = accuracy_score(y_test, rf_model.predict(X_test))
```

$$exactitud = \frac{(2742+83)}{(2742+1+\ 83+0)} = 0{,}9996$$

Ecuación 22 Random Forest - Exactitud

Especificidad

```
specificity = specificity_score(y_test, rf_model.predict(X_test))
```

$$\text{especificidad} = \frac{2742}{(2742+1)} = 0{,}9996$$

Ecuación 23 Random Forest - Especificidad

F1- Score

```
f1 = f1_score(y_test, rf_model.predict(X_test))
```

$$f1score = 2 * \frac{0{,}9881 * 1}{(0{,}9881+1)} = 0{,}994$$

Ecuación 24 Random Forest - F1-Score

Al aplicar el modelo *Random Forest* se puede observar que tiene una precisión del 98,81% es decir, está detectando correctamente las transacciones legítimas, en la sensibilidad se tiene también un 100%, es decir, se está detectando correctamente las transacciones fraudulentas, al buscar un equilibrio entre la precisión y la sensibilidad se tiene un f1-score de 99,4%, al compararlos con los modelos anteriores se puede observar que el f1-score de *Random Forest* es igual al fi-score de Regresión Logística, la bibliografía consultada indica que el modelo más utilizado para problemas de detección de fraude es el modelo de *Random Forest.*

Los modelos de aprendizaje de máquina que se utilizaron en este estudio fueron: Máquina de Vector de Soporte, Regresión Logística y Bosque Aleatorios, con un conjunto de datos en donde 11000 transacciones son legítimas y 303 tienen etiqueta de fraude, obteniendo los siguientes resultados:

Modelo	**Precisión**	**Sensibilidad**	**Exactitud**	**Especificidad**	**F1-score**
Regresión Logística	98.81%	100%	99.96%	99.96%	99.4%
Máquina de vector de Soporte	100%	97.59%	99.93%	100%	98.78%
Bosque Aleatorio	98.81%	100%	99.96%	99.96%	99.4%

Tabla 6 Comparación de Modelos de Aprendizajes

CONCLUSIONES

- Este estudio demuestra que, entre los algoritmos de aprendizaje de máquina evaluados para la detección de fraude en transacciones electrónicas, el Bosque Aleatorio se destaca por su precisión, al alcanzar un f1-score del 100%. Se realizó una búsqueda de los mejores parámetros de aprendizaje para los modelos de Regresión Logística y Máquina de Vector de Soporte, pero no lograron superar los resultados del Bosque Aleatorio. De 303 transacciones fraudulenta se dividió el 75% de los registros en entrenamiento y el 25% para evaluación, el modelo de Bosque Aleatorio logró descubrir al 100% las transacciones con etiqueta de fraude, teniendo un margen de error del 0%
- Las características que se consideraron para que el modelo comenzara su etapa de aprendizaje fueron: la variable cuantitativa valor de la transacción, variables cualitativas país, posEntry mode, código y tipo de comercio, mensaje de la transacción. El análisis de las variables cualitativas reveló que las transacciones fraudulentas se originaron principalmente en Chile, se concentraron en los códigos de comercio 4900 (servicios públicos) y 4814 (servicios de telecomunicaciones), y se asociaron con los mensajes de transacción "Aprobación", "Cuenta Bloqueada" y "Fondos Insuficientes". El análisis de la frecuencia de compra reveló que los clientes afectados por el fraude eran nuevos en el comercio electrónico, ya que solo habían realizado una transacción a la fecha de corte. Esta información es un fuerte indicador de fraude y puede ser utilizada para mejorar la precisión de los modelos de detección.
- La detección de fraude es un desafío continuo que requiere un enfoque proactivo. La implementación de modelos de aprendizaje automático como Bosque Aleatorio, junto con otras medidas de seguridad, puede ayudar a proteger a las instituciones financieras y a sus clientes del fraude. Se puede concluir que los algoritmos de aprendizaje de máquina se pueden utilizar en conjunto con los departamentos de seguridad de las instituciones financieras para analizar y prevenir las transacciones fraudulentas, creando mecanismos que alerten al administrador del área de posibles casos de fraude en las transacciones *e-commerce*.

RECOMENDACIONES

- Combinar el modelo de Bosque Aleatorio con el conocimiento experto de los departamentos de seguridad para mejorar el proceso de detección de fraudes. Para mantener la relevancia y efectividad del modelo de detección de fraude, se recomienda establecer un proceso de iteración continua que incluya el reentrenamiento periódico del modelo con nuevos datos de transacciones. Esto no solo mejorará la capacidad del modelo para adaptarse a las cambiantes tácticas de fraude, sino que también ayudará a incorporar nuevas variables que puedan surgir como indicadores relevantes de actividad fraudulenta.
- Este estudio se basa en un conjunto de datos específico, 11000 transacciones legítimas y 303 fraudulentas y los resultados pueden no ser generalizables a otros conjuntos de datos. Tener características como la ubicación geográfica del dispositivo desde donde se realiza la transacción ayudaría a tener un nuevo parámetro de aprendizaje para el modelo, permitiendo detectar fraude en base a las coordenadas geográficas del cliente. Para mejorar la precisión de los modelos de detección de fraude, se recomienda incluir la frecuencia de compra como una variable y establecer un umbral para identificar transacciones sospechosas. Se necesitan más investigaciones para identificar nuevas variables y desarrollar modelos aún más precisos para la detección de fraude. Estudiar el comportamiento de los estafadores para comprender mejor sus estrategias y desarrollar modelos más robustos
- El archivo .pkl generado por el modelo de Bosque Aleatorio contiene el conocimiento adquirido del modelo a partir de los datos proporcionados. Este archivo puede ser utilizado para crear un servicio web que permita predecir fraudes en tiempo real, lo que facilitaría la toma de decisiones estratégicas que ayuden a prevenir y mitigar el impacto del fraude en las transacciones *e-commerce*. Ante los resultados obtenidos, y dada la creciente complejidad de los esquemas de fraude, se sugiere explorar el uso de técnicas de aprendizaje profundo, como redes neuronales y autoencoders, para la detección de fraude. Estas técnicas, conocidas por su capacidad para capturar patrones complejos y no lineales en grandes volúmenes de datos, podrían ofrecer mejoras significativas en la precisión y la detección temprana de transacciones fraudulentas.

REFERENCIAS

Alberto, C., & Arcos, V. (2022). *Selection of a Machine Learning Technique for the Detection of Digital Financial Fraud Focused on Unauthorized or Consented Transactions*.

Álvarez, F. (2020). *Machine Learning en la detección de fraudes de comercio electrónico aplicado a los servicios bancarios*. https://dspace.palermo.edu/ojs/index.php/cyt/article/view/4310/6143

Ameijeiras Sánchez, D., Valdés Suárez, O., & González Diez, H. (2021). *Algoritmos de detección de anomalías con redes profundas. Revisión para detección de fraudes bancarios*. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2227-18992021000500244&lang=es

Banco Internacional. (2021, February 5). *¿Qué es y cómo funciona el sistema financiero ecuatoriano? - Banco Internacional*. https://www.bancointernacional.com.ec/que-es-y-como-funciona-el-sistema-financiero-ecuatoriano/

Banco Mundial. (2022, March 29). *Inclusión financiera*. https://www.bancomundial.org/es/topic/financialinclusion/overview

Banco Pichincha. (2022, January 25). *Diferencias entre tarjeta de crédito y débito*. https://www.pichincha.com/blog/diferencias-entre-tarjeta-credito-y-debito

Bobadilla, J. (2021). Machine Learning y deep learning usando python, scikit y keras. *Editorial Ra-Ma (España)*, 294.

Borràs García, P., & Caballero, P. B. (2023). *PARA EL ANÁLISIS PREDICTIVO Memoria y Anexos*.

Corporación Financiera Nacional. (2018). *GLOSARIO DE TÉRMINOS FINANCIEROS*.

Cuenca Jiménez, M. J., Calle Oleas, R. B., & Jaramillo Pedrera, C. (2022, April 26). *El Sistema Financiero a través de la Tecnología*. https://fipcaec.com/index.php/fipcaec/article/view/563/999

Dávila-Morán, R. C., Castillo-Sáenz, R. A., Vargas-Murillo, A. R., Dávila, L. V., García-Huamantumba, E., García-Huamantumba, C. F., Cajas, R. F. P., & Paredes, C. E. G. (2023). Application of Machine Learning Models in Fraud Detection in Financial Transactions. *Data and Metadata, 2*. https://doi.org/10.56294/dm2023109

Dietterich, T., Bishop, C., Heckerman, D., Jordan, M., & Kearns, M. (2022). *Adaptive Computation and Machine Learning*. https://lccn.loc.gov/2021027430

Diners Club. (2022). *3 tipos de fraudes bancarios, y cómo evitarlos en Ecuador*. https://www.dinersclub.com.ec/experiencias/diners-club/tipos-fraudes-bancarios

Dueñas Quesada, J. M. (2020). *Aprendizaje supervisado para la detección de amenazas web mediante clasificación basada en árboles de decisión: Aplicación de técnicas de machine learning a la ciberseguridad.*

Ecuador. (2018). *Código Orgánico Integral Penal (modificado hasta el 14 de febrero de 2018)*. www.lexis.com.ec

Fernández Khatiboun, A. (2019). *Machine Learning en Ciberseguridad*.

Gobierno de Mexico. (2017). *¿Ya conoces estos tipos de fraude? | Comisión Nacional para la Protección y Defensa de los Usuarios de Servicios Financieros | Gobierno | gob.mx*. https://www.gob.mx/condusef/articulos/ya-conoces-estos-tipos-de-fraude

Grupo ISO/TC 262/STTF. (2017). *ISO 31000:2018(es), Gestión del riesgo — Directrices*. https://www.iso.org/obp/ui/es/#iso:std:iso:31000:ed-2:v1:es

Gutiérrez Portela, F., Rodríguez Cárdenas, S., Patiño Ospina, L. P., & Hernandez Aros, L. (2023). Estudio de la prevención y detección de fraudes financieros a través de técnicas de aprendizaje automático. *CAFI*, *6*(1), 77–101. https://doi.org/10.23925/CAFI.V6I1.58372

Hastie, T., Tibshirani, R., & Friedman, J. (2008). *Springer Series in Statistics The Elements of Statistical Learning Data Mining, Inference, and Prediction*.

Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2006). *Metodologia de la Investigacion Cuarta Edicion*.

Instituto Nacional de estadísticas y Censo. (2019). *CONCEPTOS Y DEFINICIONES.doc | Enhanced Reader*.

Katherine Quintero Acuña, Lady. (2023). *Aplicación de Machine Learning a un modelo tradicional de Prevención y detección de fraude en entidad financiera proyectado a periodos trimestrales*. https://ciencia.lasalle.edu.co/maest_analitica_inteligencia_negocios

Lucha Campesina. (2023, August 20). *Lucha Campesina*. https://luchacampesina.fin.ec/

Mirjalili, V., & Rasch, S. (2020). *Python Machine Learning*. https://books.google.com.ec/books?hl=es&lr=lang_es&id=5EtOEAAAQBAJ&oi=fnd&pg

=PT5&dq=machine+learning&ots=erE3PyYHO9&sig=7v8jGpM1InmP4p3b3OIQh_B0Bes&redir_esc=y#v=onepage&q=machine%20learning&f=false

Pérez González, G. (2021). *Detección de transacciones fraudulentas en tarjetas de crédito mediante el uso de modelos de Machine Learning*.

Quintana Adriano, E. A. (2004, September). *Los servicios financieros en México y la Organización Mundial de Comercio*. https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0041-86332004000300005

Real Academia Espanola. (2023). *fraude | Definición | Diccionario de la lengua española | RAE - ASALE*. https://dle.rae.es/fraude

Shai Shalev, S., & Shai Ben, D. (2014). *Understanding Machine Learning: From Theory to Algorithms*. http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning

Soley, J. (2015). *Banca y tecnología: dos realidades hermanadas*.

Superintendencia de Bancos. (2021). *LIBRO I.-NORMAS DE CONTROL PARA LAS ENTIDADES DE LOS SECTORES FINANCIEROS PÚBLICO Y PRIVADO*.

Superintendencia de Economía Popular y Solidaria. (2017). *Resolución No. SEPS-IGT-IR-ISF-ITIC-IGJ-2017-103*.

Superintendencia de Economía Popular y Solidaria. (2021). *Evaluación de la Inclusión Financiera y los Servicios Financieros Digitales en el Ecuador*.

Superintendencia de Economía Popular y Solidaria. (2022). *Resolución No. SEPS-IGT-IGS-INR-INGINT-2022-0211*.

The World Bank. (2010). *Códigos de países*. https://wits.worldbank.org/wits/wits/WITSHELP-es/content/codes/country_codes.htm

Visa. (2023a). *Qué es e-commerce o comercio electrónico | Visa*. https://www.visa.com.ec/dirija-su-negocio/pequenas-medianas-empresas/notas-y-recursos/tecnologia/que-es-ecommerce-o-comercio-electronico.html

Visa. (2023b). *Visa Merchant Data Standards Manual*.

ANEXOS

Anexo 1: Autorización

DE: Ing. Pedro Borbor
Ingeniero de Desarrollo
COAC LUCHA CAMPESINA.

PARA: Ing. Juan Carlos Zambrano
Gerente General
COAC LUCHA CAMPESINA.

FECHA: 08 de noviembre del 2023

ASUNTO: Permiso de uso de información para entrenamiento de programa de machine learning – Sistema de Prevención Temprana de Fraudes Fase 1

Me dirijo a usted como funcionario de la cooperativa y como estudiante del Programa de Maestría en Tecnología de la Información de la Universidad Estatal Península de Santa Elena. Actualmente, me encuentro desarrollando mi tesis de maestría, que se titula " ***Aprendizaje de máquina para detectar fraude en tarjetas de débito de la Cooperativa de ahorro y crédito Lucha Campesina***", tiene como objetivo contribuir a la mejora de la detección de actividades fraudulentas en transacciones que se hagan con tarjeta de débito en establecimientos e-commerce.

Para lograr esto, es crucial que utilice datos reales y específicos en el proceso de entrenamiento de un programa de machine learning. Por tanto, solicito su autorización para permitirme utilizar los datos de la cooperativa en mi proyecto de tesis.

Estoy comprometido en tratar toda la información con la más estricta confidencialidad y con las debidas seguridades, así como firmar los acuerdos de confidencialidad necesarios y seguir todas las políticas establecidas por la cooperativa para el manejo de los datos. Mi objetivo es asegurarme de que esta colaboración sea beneficiosa para ambas partes y se lleve a cabo de manera ética y legal.

Agradezco su tiempo y consideración en esta solicitud, quedo de ustedes.

Atentamente,

Pedro Borbor, Ing.

Ingeniero de Desarrollo

cha Campesina
ORIZADO
19/11/2023 Hora 11:36
RFNCIA

cc. Carlos Velez; Jefe de Tecnologia

Sadoth Rodriguez; OSI