

Fronteras Digitales: IA en la Batalla contra el fraude crediticio en Colombia

Jhoan Sebastian Beltran, John Aldemar Gonzalez, Kevin Andres Quintero

Dpto. de Ciencias Básicas

Universidad Central

Maestría en Analítica de Datos

Curso de Automatización e Integración de Datos

Bogotá, Colombia

jbeltranp3@ucentral.edu.co, jgonzalezl8@ucentral.edu.co, kquinteroc@ucentral.edu.co

April 11, 2024

Contents

1	Introducción	3
2	Características del proyecto de investigación que hace uso de Integración y Automatización de Datos para IA	3
2.1	Titulo del proyecto de investigación	4
2.2	Objetivo general	4
2.2.1	Objetivos especificos	4
2.3	Alcance	4
2.4	Pregunta de investigación	5
2.5	Hipotesis	5
3	Reflexiones sobre el origen de datos e información	6
3.1	¿Cual es el origen de los datos e información?	6
3.2	¿Cuales son las consideraciones legales o éticas del uso de la información?	6
3.3	¿Cuales son los retos de la información y los datos que utilizara en Integración y Automatización de Datos para IA?	7
3.4	¿Que espera de la utilización de Integración y Automatización de Datos para IA para su proyecto?	7
4	Diseño de integración y Automatización de Datos para IA (Diagrama)	8

5	integración de Datos (<i>Segunda entrega</i>)	8
6	Automatización de Datos (<i>Segunda entrega</i>)	8
7	IA (<i>Segunda entrega</i>)	8
8	Proximos pasos (<i>Tercera entrega</i>)	8
9	Lecciones aprendidas (<i>Tercera entrega</i>)	8
10	Bibliografía	8

1 Introducción

El fraude de tarjetas de crédito es un problema que afecta a las entidades financieras en todo el mundo. Este tipo de fraude es un problema creciente en Colombia que genera pérdidas económicas significativas tanto para las entidades financieras como para los consumidores. Es un problema serio que requiere atención por parte del sector financiero, los consumidores y las autoridades. Este proyecto busca comprender las diferentes formas de fraude, sus causas e impacto, para proponer estrategias de prevención y control. A través del análisis realizado a un conjunto de datos de transacciones financieras, extraído de un cliente en particular, se proponen estrategias de control a través de reglas de negocio del proveedor para identificar y validar cuáles son las diferentes acciones que realizan los actores del fraude. Una vez es identificado el proceso o accionar del fraude, podemos implementar las estrategias y acciones necesarias para bloquear aquellas transacciones sospechas y generar las alertas necesarias para dar el tratamiento indicado a cada fraude y poder entrenar un modelo automatizado que pueda responder autónomamente a cada ataque realizado en el proceso de transacciones financieras a través del manejo de tarjetas de crédito de los diferentes clientes

2 Características del proyecto de investigación que hace uso de Integración y Automatización de Datos para IA

Este proyecto de integración y automatización de datos se ha definido de acuerdo con los requisitos previos analizados para la detección del fraude en transacciones con tarjetas de crédito. Hemos definido claramente los objetivos principales y específicos apuntando a una solución automatizada que brinde una estructura eficaz para detectar y reportar el fraude. Se han seleccionado adecuadamente los datos estructurados de una fuente de datos real de control de transacciones anómalas mediante archivos csv. Se extraen los datos de los archivos, luego, se limpian, se transforman, se normalizan y se manejan los valores faltantes o datos erróneos para ser almacenados y procesados mediante la herramienta de motor de base de datos SQL server. Con los datos preparados, se crea un modelo automatizado IA a través de un algoritmo con parámetros específicos que validen y seleccionen transacciones que cumplan con las validaciones configuradas e identificarlas como fraude y poder generar las notificaciones requeridas al área o usuarios encargados de atender dichas anomalías. De igual manera se genera visualización de las transacciones de fraude. Por último, se valida el modelo implementado con los resultados obtenidos y se generan nuevos datos que son ingresados al modelo para que sean validados y obtener posibles nuevas transacciones que pueden ser generadas como anómalas o de fraude.

2.1 Título del proyecto de investigación

Fronteras Digitales: IA en la Batalla contra el fraude crediticio en Colombia

2.2 Objetivo general

Desarrollar e implementar un modelo de inteligencia artificial (IA) para la detección y prevención del fraude con tarjetas de crédito en Colombia, utilizando técnicas de aprendizaje automático y análisis de datos, para identificar patrones y comportamientos sospechosos en tiempo real, generando alertas y reportes visuales.

2.2.1 Objetivos específicos

- Analizar las diferentes formas de fraude de tarjetas de crédito en Colombia, identificando sus características, causas y tendencias, utilizando técnicas de investigación y análisis de datos.
- Desarrollar un modelo de IA capaz de detectar patrones y comportamientos sospechosos en tiempo real, utilizando técnicas de aprendizaje automático supervisado, y un conjunto de datos de transacciones financieras preprocesadas.
- Implementar el modelo de IA en una arquitectura escalable, integrándose con la base de datos transaccional a través de APIs, y desplegándose usando tecnologías como Docker.
- Establecer un proceso de monitoreo y mejora continua del modelo de IA, incluyendo la recopilación periódica de nuevos datos, el reentrenamiento del modelo y el diseño de un tablero de control y sistema de alertas para la toma de acciones oportunas.

2.3 Alcance

El proyecto se centrará en el desarrollo e implementación de un modelo de IA para la detección y prevención del fraude con tarjetas de crédito en Colombia. Incluirá un análisis profundo de las diversas formas de fraude, sus características, causas y tendencias, utilizando un conjunto de datos de transacciones financieras proporcionado por redeban y credibanco. Se desarrollará un modelo de IA basado en técnicas de aprendizaje automático supervisado, capaz de identificar patrones y comportamientos sospechosos en tiempo real. El modelo se integrará en una arquitectura escalable, conectándose con la base de datos transaccional mediante APIs y utilizando tecnologías de contenedores como Docker. Además, se establecerá un proceso de monitoreo y mejora continua, que abarca la recopilación periódica de nuevos datos, el reentrenamiento del modelo y la creación de un tablero de control y sistema de alertas.

2.4 Pregunta de investigación

¿De qué manera un modelo de inteligencia artificial, desarrollado mediante técnicas de aprendizaje automático supervisado y utilizando un conjunto de datos de transacciones financieras preprocesadas, puede contribuir a mejorar la detección y prevención del fraude con tarjetas de crédito en Colombia?

2.5 Hipotesis

La implementación de un modelo de inteligencia artificial, desarrollado a través de técnicas de aprendizaje automático supervisado, reducirá significativamente la tasa de fraudes asociados al fraude con tarjetas de crédito en Colombia. Aumentando la confianza de los usuarios en el sistema financiero colombiano.

3 Reflexiones sobre el origen de datos e información

Reflexionar sobre el origen de los datos e información, específicamente del reporte sobre fraude en las franquicias Visa y Mastercard dirigido al canal de adquierecia en Colombia, resalta la importancia de fuentes focalizadas para entender el fraude financiero. Este origen demuestra la necesidad de datos precisos para abordar los retos de fraude en transacciones electrónicas y la importancia de la colaboración entre entidades financieras, franquicias de tarjetas y comerciantes en la prevención del fraude.

La concentración en el canal de adquierecia indica la relevancia de estos datos para identificar vulnerabilidades y tendencias en fraudes, sugiriendo un esfuerzo conjunto para reforzar la seguridad y las políticas de protección al consumidor. Este enfoque no solo destaca la complejidad del entorno digital financiero, que requiere vigilancia y actualización constante de mecanismos de defensa contra el fraude, sino también la importancia de la transparencia y responsabilidad en la gestión de información relacionada con el fraude para fomentar la confianza entre usuarios y entidades financieras.

En resumen, el análisis del origen de estos datos aporta insights sobre las características del fraude en Colombia y fomenta una discusión sobre los desafíos de proteger la integridad financiera y la seguridad de los consumidores en la era digital, subrayando el valor de la cooperación y la innovación en estrategias de prevención.

3.1 ¿Cual es el origen de los datos e información?

El origen de los datos e información proviene de un reporte detallado sobre el fraude asociado con las franquicias Visa y Mastercard. Este reporte se enfoca específicamente en el canal de adquierecia y está dirigido al contexto colombiano. Dicho canal implica a las entidades que procesan pagos con tarjetas de crédito o débito, lo que indica que la información recabada es crucial para entender las dinámicas y el volumen de actividades fraudulentas dentro del país. Este informe constituye una fuente primaria valiosa para analizar y mitigar los riesgos relacionados con el fraude en transacciones realizadas mediante estas franquicias en Colombia.

3.2 ¿Cuales son las consideraciones legales o éticas del uso de la información?

Las consideraciones legales y éticas del uso de la información del reporte de fraude de Visa y Mastercard en Colombia incluyen garantizar la privacidad y protección de datos personales, cumpliendo con normativas como la Ley de Protección de Datos. Éticamente, es fundamental usar la información de manera responsable, evitando estigmatizaciones y asegurando que las medidas de prevención y corrección no vulneren derechos individuales. Legalmente, se debe respetar el marco jurídico aplicable a fraudes financieros y comercio

electrónico, promoviendo transparencia, justicia y equidad en el manejo y análisis de los datos.

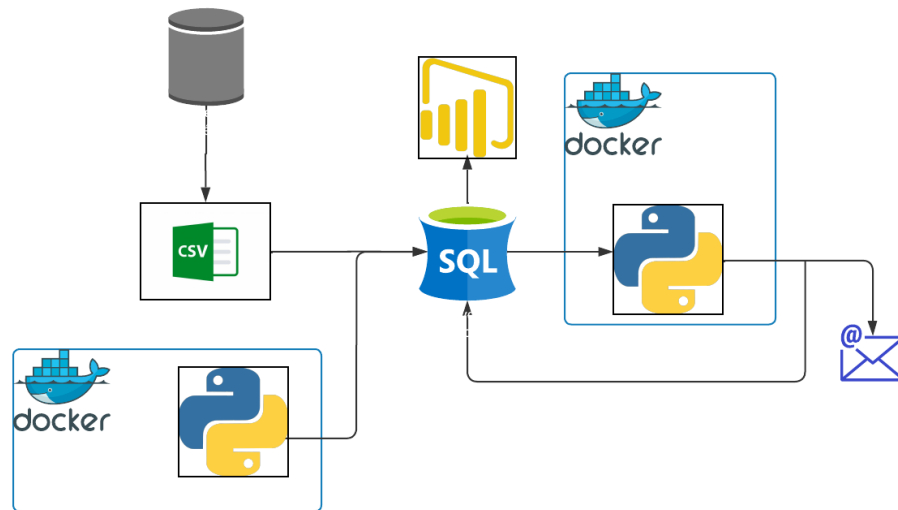
3.3 ¿Cuales son los retos de la información y los datos que utilizara en Integración y Automatización de Datos para IA?

Los retos principales en la integración y automatización de datos para IA incluyen garantizar la calidad y precisión de los datos, enfrentando desafíos como la heterogeneidad de las fuentes y la inconsistencia en los formatos. Es crucial implementar métodos eficientes para la limpieza y preparación de datos, asegurando su relevancia y actualidad. Además, se debe abordar la privacidad y seguridad de la información, cumpliendo con regulaciones legales y éticas, mientras se optimizan los procesos de integración para facilitar el acceso y la interoperabilidad de los datos, permitiendo así su uso efectivo en aplicaciones de inteligencia artificial.

3.4 ¿Que espera de la utilización de Integración y Automatización de Datos para IA para su proyecto?

Se espera que la integración y automatización de datos para IA potencien nuestro proyecto, mejorando la eficiencia en el procesamiento y análisis de datos. Esto debería conducir a insights más precisos y decisiones basadas en datos, con un impacto directo en la optimización de recursos y la mejora de resultados. Asimismo, buscamos fortalecer la seguridad y la privacidad de los datos manejados, alineándose con las mejores prácticas y cumplimiento normativo. Finalmente, aspiramos a que estas tecnologías nos permitan innovar y mantenernos a la vanguardia en nuestra industria, ofreciendo soluciones más inteligentes y adaptativas a las necesidades de prevenir el fraude en la industria de pagos con tarjetas débito y crédito en Colombia.

4 Diseño de integración y Automatización de Datos para IA (Diagrama)



5 integración de Datos *(Segunda entrega)*

6 Automatización de Datos *(Segunda entrega)*

7 IA *(Segunda entrega)*

8 Proximos pasos *(Tercera entrega)*

9 Lecciones aprendidas *(Tercera entrega)*

10 Bibliografía