UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA

DESARROLLO DE SOFTWARE

EXTRACCIÓN DE CONOCIMIENTO EN BASES DE DATOS

REPORTE DE CASO DE ESTUDIO APLICANDO IA/ML/DM/BIG DATA

PRESENTA:

MILDRED VILLASEÑOR RUIZ

ÁNGEL RICARDO CHÁVEZ ZARAGOZA

DARON TARÍN GONZÁLEZ

RICARDO ALONSO RÍOS MONRREAL

DOCENTE:

ING. LUIS ENRIQUE MASCOTE CANO

Chihuahua, Chih., 23 de septiembre de 2025

Contenidos

Introducción 3
Introducción al caso 3
Justificación del dominio 3
Beneficios y KPIs 4
Herramientas y lenguajes 4
Diagrama de alto nivel 5
Conclusión 5
Referencias 6

Introducción

El crecimiento del comercio electrónico y la banca digital ha incrementado de manera exponencial el volumen de transacciones financieras realizadas a través de canales como aplicaciones móviles, cajeros automáticos y plataformas web. Este panorama ha generado nuevos retos en materia de seguridad, ya que los fraudes mediante transferencias no autorizadas y clonación de tarjetas se han vuelto cada vez más frecuentes. Ante esta situación, las instituciones financieras requieren soluciones avanzadas que permitan integrar grandes volúmenes de datos históricos con la capacidad de procesar información en tiempo real para detectar patrones anómalos. El uso de un Data Warehouse combinado con modelos de Machine Learning y tecnologías de Big Data representa una estrategia clave para identificar comportamientos fraudulentos, reaccionar de forma inmediata y proteger tanto los activos financieros como la confianza de los clientes en la institución.

Introducción al caso

SecureBank procesa millones de transacciones diarias a través de múltiples canales como web, app móvil y cajeros automáticos. Con el incremento del comercio electrónico y la banca digital, se ha registrado un aumento significativo en fraudes mediante transferencias no autorizadas y pagos con tarjetas clonadas. El reto es implementar un sistema que combine un Data Warehouse para consolidar datos históricos y capacidades analíticas en tiempo real para detectar patrones anómalos en las transacciones. Esto permitirá a los equipos de fraude actuar rápidamente para mitigar pérdidas financieras, proteger a los clientes y mantener la confianza en la institución. rando equipos de fraude, clientes y tecnología, con impacto directo en la seguridad y confianza.

Justificación del dominio

El uso de Machine Learning es fundamental para abordar este problema porque permite ir más allá de las reglas estáticas tradicionales, aprendiendo comportamientos típicos y detectando anomalías nuevas y no conocidas previamente. Además, al combinar ML con técnicas de Big Data se puede

procesar y analizar grandes volúmenes de transacciones en tiempo real, lo cual es indispensable para responder con rapidez a incidentes críticos. Así, ML habilita un sistema de detección flexible, escalable y con capacidad de mejora continua, esencial para la lucha contra fraudes en entornos bancarios modernos.

Beneficios y KPIs

La implementación de un sistema de Data Warehouse con analítica en tiempo real y modelos de Machine Learning aportará beneficios estratégicos y operativos a SecureBank:

Reducción de pérdidas financieras por fraude

- **Métrica**: Disminuir al menos un 40% el monto asociado a transacciones fraudulentas en el primer año de implementación.

Mejora en la precisión de detección

- **Métrica**: Alcanzar una tasa de precisión superior al 90% en la clasificación de transacciones sospechosas, reduciendo los falsos positivos que generan fricción con clientes legítimos.

Optimización del tiempo de respuesta

- **Métrica**: Reducir el tiempo promedio de detección y reacción de incidentes de fraude a menos de 1 minuto desde la ocurrencia de la transacción, permitiendo bloquear operaciones antes de concretarse.

Herramientas y lenguajes

Java + Spring Boot

Para la capa de servicios críticos (ETL, API de inferencia y comunicación con el Data Warehouse) se utiliza Java con Spring Boot, ya que es un lenguaje fuertemente tipado que asegura integridad, transaccionalidad y alta concurrencia. Permite construir microservicios que orquestan la limpieza, transformación y carga de datos (ETL), así como la exposición de la API de inferencia para evaluar transacciones en tiempo real.

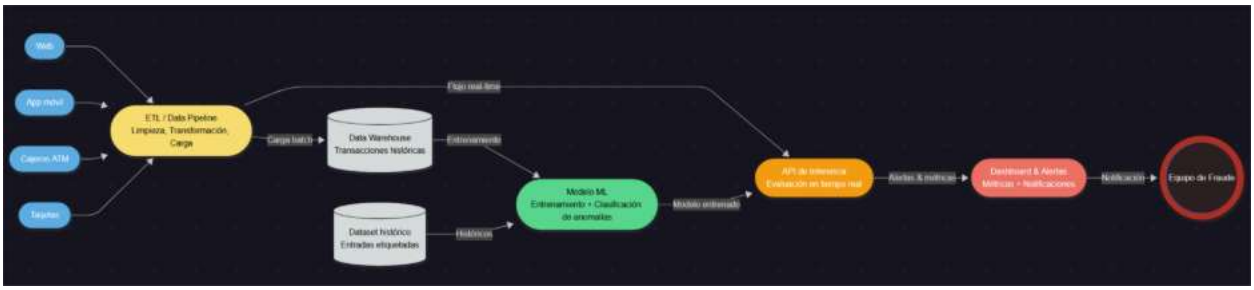
Apache Kafka

En el flujo de datos en tiempo real, Kafka funciona como backbone de mensajería distribuida, conectando múltiples fuentes de transacciones (web, app móvil, cajeros, tarjetas) con los módulos analíticos. Esto garantiza la ingestión continua y escalable de millones de eventos, alimentando tanto el Data Warehouse (para históricos) como el modelo de ML (para clasificación inmediata de anomalías).

Python + TensorFlow/PyTorch

El entrenamiento y despliegue de modelos de detección de fraude se realiza en Python, usando frameworks como TensorFlow o PyTorch. Estos modelos procesan datasets históricos y, una vez entrenados, se integran con la API Java para inferencia en producción. De este modo, se combina la flexibilidad y potencia de Python en Machine Learning con la estabilidad de un lenguaje fuertemente tipado en la operación bancaria.

Diagrama de alto nivel



Conclusión

La detección de fraudes bancarios en tiempo real es un desafío estratégico que exige un enfoque integral, combinando infraestructura tecnológica robusta, técnicas analíticas avanzadas y metodologías de Machine Learning. La solución propuesta para SecureBank integra herramientas de almacenamiento, procesamiento y modelado que permiten no solo reducir pérdidas financieras, sino también mejorar la experiencia del cliente al disminuir falsos positivos y garantizar respuestas ágiles. Este tipo de sistemas se convierten en un pilar fundamental para la continuidad del negocio

en un entorno digital cada vez más dinámico y vulnerable a ciberamenazas. Los próximos pasos deben enfocarse en la implementación gradual, pruebas piloto controladas y la mejora continua de los modelos, asegurando la escalabilidad y adaptación a nuevas modalidades de fraude.

Referencias

A Guide to Preventing Fraud Detection in Real-Time with Apache Flink. (n.d.). Alibaba Cloud Community. https://www.alibabacloud.com/blog/a-guide-to-preventing-fraud-detection-in-real-time-with-apache-flink_602024

IBM. (2024). Detecting and preventing fraud in banking with AI and machine learning. IBM. https://www.ibm.com/topics/fraud-detection

Use streaming ingestion to make ML-backed decisions in near-real time | Amazon Web Services. (2023, November 9). Amazon Web Services. https://aws.amazon.com/blogs/machine-learning/use-streaming-ingestion-with-amazon-sagemaker-feature-store-and-amazon-msk-to-make-ml-backed-decisions-in-near-real-time/