Riskova 2024

Creado por:

Jose Ibarra

Gabriela Lopez

Silvana Jaramillo

"La seguridad no es solo un ideal, es una necesidad." – Eleanor Roosevelt

Riskova

Autores:

Jose Ibarra – Business Analyst

Focus: Utiliza herramientas de Business Intelligence para visualizar datos y crear informes y paneles interactivos.

Gabriela Lopez – Data Analyst

Focus: Responsable de recopilar, limpiar y analizar datos para extraer insights y generar valor empresarial.

Silvana Jaramillo – Machine Learning Engineer

Focus: Desarrolla algoritmos para entrenar modelos predictivos o clasificadores automatizados.

Introducción

El proyecto consistió en el desarrollo de un **MVP (Producto Mínimo Viable)** para una plataforma de detección de fraudes en pagos electrónicos. Este MVP integró aprendizaje automático y análisis de datos para proporcionar una solución práctica y funcional.

El MVP incluye:

1. **Procesamiento de Datos:**
   Se realizó la carga y limpieza de datos transaccionales, abordando problemas como valores nulos, duplicados y escalado de características para garantizar que la información estuviera lista para su análisis.
2. **Modelo Predictivo:**
   Se entrenó e implementó un modelo de machine learning capaz de identificar transacciones sospechosas con precisión, evaluando su desempeño con métricas como precisión, recall y F1-score.
3. **Monitoreo en Tiempo Real:**
   Se desarrolló un sistema interactivo que permite visualizar transacciones en tiempo real, junto con alertas automáticas en caso de detectar posibles fraudes.

Descripción:

Desarrollaremos un modelo predictivo para detectar transacciones fraudulentas en una plataforma de pagos electrónicos, utilizando técnicas de aprendizaje automático y análisis de comportamiento.

El sistema tiene el siguiente flujo:

1. Inicio (Home)

Objetivo: Presentar una introducción clara y concisa al proyecto.
Implementación:
Se diseñó una página inicial con una breve descripción del problema de fraude en tarjetas de crédito, destacando su impacto global y la importancia de una solución innovadora. También se incluyó un resumen de las áreas involucradas, como análisis de datos, machine learning y visualización, detallando los roles de cada equipo. La página ofrece botones intuitivos para navegar a las distintas secciones del sistema.

---

2. Carga y Exploración de Datos (Data Ingestion & Exploration)

Ingeniería de Datos:
Se habilitó la carga de datasets en formato CSV o conexión a bases de datos existentes. Se desarrolló una funcionalidad para mostrar estadísticas descriptivas de los datos, como el número de filas y columnas, tipos de datos y valores nulos, junto con gráficos de distribución de valores faltantes utilizando herramientas como Pandas y Matplotlib.

Análisis Exploratorio:
Se realizaron análisis descriptivos con gráficos interactivos que muestran distribuciones y correlaciones entre variables, ayudando a identificar patrones de fraude. Por ejemplo, se crearon histogramas que comparan transacciones legítimas y fraudulentas.

---

3. Transformación y Limpieza de Datos (Data Preprocessing)

Ingeniería de Datos:
Se gestionaron valores nulos, datos duplicados y valores atípicos mediante técnicas específicas, y se realizaron transformaciones como normalización, codificación y escalado. Cada transformación fue registrada en logs para su trazabilidad.

Interactividad:
Se implementaron opciones para que los usuarios seleccionaran y aplicaran técnicas de limpieza o transformación, brindando flexibilidad y control sobre el preprocesamiento de los datos.

---

4. Análisis de Negocio (BI Dashboard)

Analista BI:
Se desarrollaron tableros interactivos que presentan KPIs clave como el total de transacciones procesadas, porcentaje de fraudes detectados y tendencias temporales. Gráficos dinámicos, como mapas geográficos y líneas de tiempo, facilitaron el análisis de patrones de fraude por región y tiempo.

---

5. Modelado Predictivo (Machine Learning)

Data Scientist / Machine Learning Engineer:
Se entrenaron modelos predictivos utilizando técnicas de machine learning, como redes neuronales. Se presentó una comparación clara entre datos de entrenamiento y prueba, mostrando las métricas de evaluación (precisión, recall y F1-score).

Interactividad:
Los usuarios pudieron probar diferentes configuraciones de hiperparámetros y visualizar resultados como la curva ROC y la matriz de confusión.

---

6. Análisis Avanzado (Data Science Insights)

Data Scientist:
Se incluyó una interpretación detallada del modelo a través de gráficos de importancia de características y valores SHAP, ayudando a comprender qué variables fueron más relevantes para la detección de fraudes.

---

7. Pruebas en Tiempo Real (Fraud Detection Simulation)

Machine Learning / BI:
Se simuló la detección en tiempo real, permitiendo a los usuarios cargar transacciones para predecir su legitimidad. Los resultados se presentaron de forma gráfica e incluyeron explicaciones detalladas del modelo.

---

8. Reporte y Exportación

Analista BI:
Se generaron reportes descargables en formatos PDF y Excel, que incluyen gráficos y métricas clave. Un resumen ejecutivo detalla las contribuciones de cada área involucrada en el proyecto.

---

9. Documentación y Acerca del Proyecto

Equipo Completo:
Se redactó documentación técnica que describe el proyecto, las herramientas y frameworks utilizados, roles y contribuciones, y una guía de uso del sistema.

Datos:

Dataset Sintético de Simulación (Fraud Detection de Kaggle):
https://www.kaggle.com/datasets/kartik2112/fraud-detection

Consideraciones Éticas

En nuestra plataforma de detección de fraudes:

1. Privacidad: Utilizamos datasets anonimizados (PCA) y sintéticos, evitando información sensible o identificable.
2. Equidad: Evaluamos sesgos para garantizar decisiones imparciales, sin discriminación por género, ubicación u otras características.
3. Transparencia: Las alertas generadas son explicables y supervisadas por humanos, asegurando decisiones responsables.

4. Cumplimiento Legal: Nos alineamos con GDPR y principios éticos

internacionales. Estos principios garantizan una solución ética, segura y confiable.

Plataforma: Interfaz en **Streamlit** con backend en **Python** para detectar fraudes en tiempo real. Utiliza **Pandas** para procesar datos y visualizaciones con **Matplotlib y Seaborn**. Incluye un **modelo predictivo preentrenado** con datos anonimizados para garantizar precisión y ética

Cronograma:
https://trello.com/invite/b/67376c8ba867c3eacc420259/ATTIeea13f40c0e85b3803938599c 9b94947C33A1431/fintech

Requisitos del Proyecto

- **Versión de Python:** Asegúrate de tener instalada la versión 3.x de Python.
- **Dependencias:** Las librerías de Python necesarias deben estar instaladas (listadas en el archivo `requirements.txt`).
- **Streamlit:** Debe estar instalado para ejecutar la interfaz de la aplicación web.
- **Archivos de Datos:** Archivos CSV que serán procesados.

Cómo Iniciarlo

Paso	Comando	Descripción
Clonar el proyecto	`https://github.com/No-Country-simulation/c22-31-m-data-bi.git`	Clona el repositorio del proyecto en tu máquina local.
Instalar dependencias	`pip install -r requirements.txt`	Instala todas las dependencias necesarias para el proyecto.
Ejecutar el Detector Predictivo de Fraudes	`streamlit run main.py`	Ejecuta la aplicación con Streamlit.