



Proyecto Final: Fraude en compras con tarjetas



Sharon Nicole Rojas, Maria Paulina Bolivar y Maria Paulina Alzate

Caso de estudio

Base original

Conjunto de datos de transacciones de tarjetas de crédito simuladas que contiene transacciones legítimas y fraudulentas desde el 1 de enero de 2019 hasta el 31 de diciembre de 2020. Cubre tarjetas de crédito de 1000 clientes que realizan transacciones con un grupo de 800 comerciantes.



Caso de estudio

Base modificada

Conjunto de datos de transacciones de tarjetas de crédito simuladas que contiene transacciones legítimas y fraudulentas desde el 1 de enero de 2019 hasta el 31 de diciembre de 2020. Cubre 226 transacciones realizadas por personas en el Estado de Florida.



Objetivos del negocio

Comprender los patrones de las transacciones fraudulentas y desarrollar modelos que puedan identificarlas con mayor precisión.

- Reducir las pérdidas por fraude
- Cumplir con las regulaciones
- Optimizar las estrategias de prevención del fraude
- Identificar nuevas tendencias de fraude



Técnica de análisis

Para la identificación de la probabilidad de fraude en tarjetas de créditos en diferentes comercios en Estados Unidos desde el 1 de enero de 2019 hasta el 31 de diciembre de 2020, se utilizará dos modelos de clasificación, el primero fue del tipo **KNN Means** para analizar su correlación significativa y el segundo es el modelo **Neuronal Madaline** para realizar el pronóstico



ARCHIVOS

BASE PRINCIPAL
(Kaggle)

BASE UTILIZADA

Descripción de variables

La estructura de la base de datos utilizada para la configuración de los modelos tienen la siguientes características:

- La BD tiene una muestra de 226 transacciones con tarjeta de crédito.
- Cada una de las transacciones está caracterizada por 7 variables (siendo una de ellas la de fraude) de tipo cuali-cuantitativo.
- La variable de referencia en el KNN es Is_fraud, y en el MADALINE es la amt.

Variables	Tipo de variable
Gender	Cuantitativa-Bivariada
States	Cuantitativa (única)
City	Cuantitativa
Category	Cuantitativa
Job	Cuantitativa
AMT (valor transacción)	Cuantitativa (continua)
Is_fraud	Cuantitativa-Bivariada

Descripción de variables

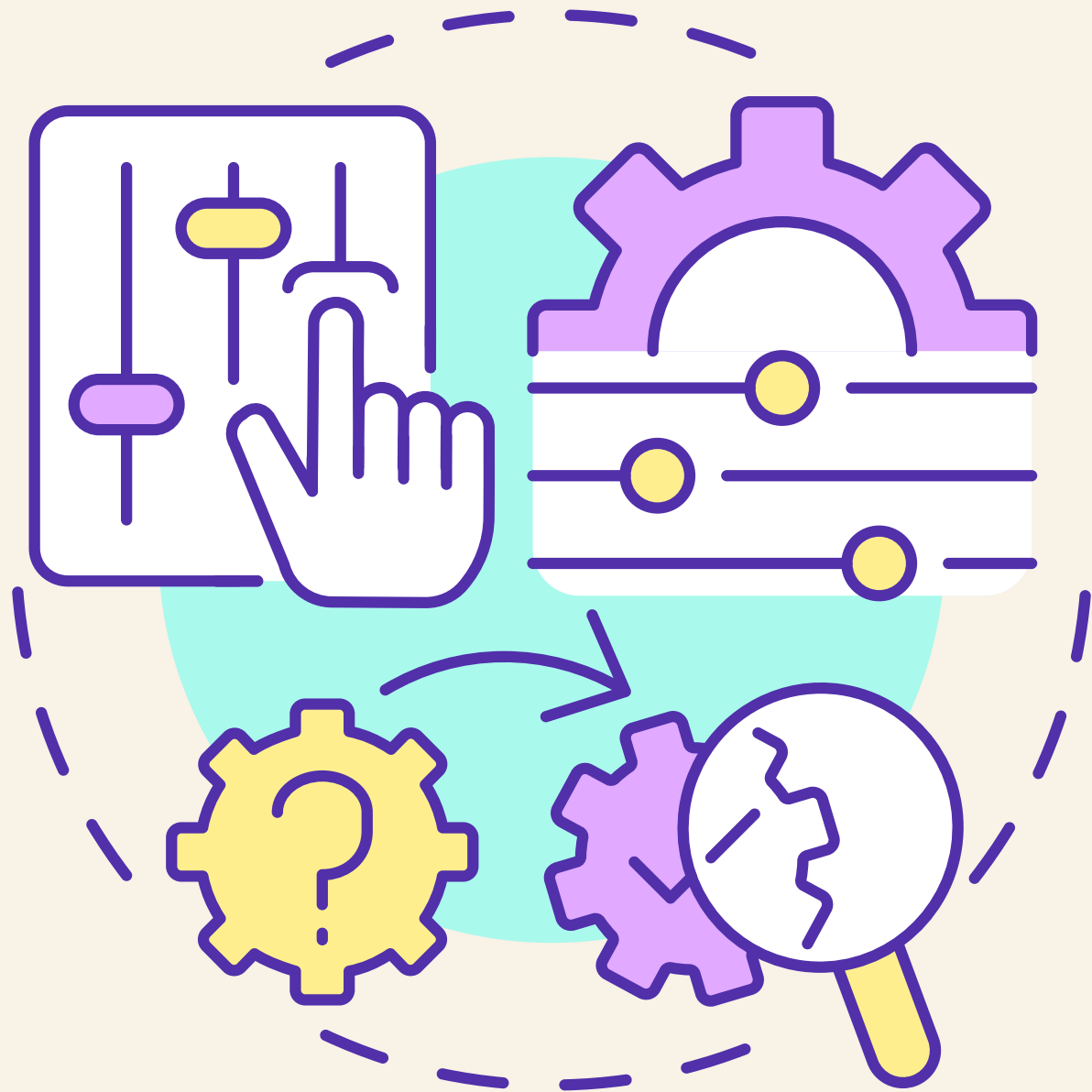
Gender	
Número	Significado
0	Masculino
1	Femenino

States	
Número	Significado
0	Florida

City	
Número	Significado
0	West Palm Beach
1	Orlando
2	Tampa
3	Jacksonville
4	Miami

Category	
Número	Significado
0	Entertainment
1	Food dining
2	Health fitness
3	Home
4	Kids pets
5	Misc pos
6	Personal care
7	Shopping pos
8	Travel

Descripción de variables



Jobs	
Número	Significado
0	Accountant
1	Advertising account planner
2	Airline pilot
3	Armed forces training and education office
4	Barista
5	Biomedical scientist
6	Clothing/textile technologist
7	Community education officer
8	Editor, magazine features
9	Engineer, land
10	Health physicist
11	Prison officer
12	Teacher
13	Water engineer
14	Web designer
15	Wellsite geologist

Preguntas del análisis

- ¿Qué tan fuerte es la correlación entre las variables predictoras y la variable objetivo ("amt")?
- ¿Qué tan efectivo fue el modelo KNN para la clasificación de los datos?
- ¿A qué cluster pertenece un usuario femenino (0) que compró en la categoría personal care (6), residente de la ciudad de West Palm Beach (0) en el estado de Florida (0), quien trabaja como Health physicist (10) y realizó una transacción por valor de 1000? ¿Cuál es la probabilidad de cometer fraude dentro de este cluster?
- ¿Qué tan efectivo fue el modelo Madaline para la predicción de la información?
- ¿Cómo se pueden interpretar los coeficientes de efectos independientes para comprender la influencia de cada variable predictora en el monto de la transacción?
- ¿Qué implica la discrepancia notable entre las medias de los grupos en el conjunto de datos? ¿Cómo podría afectar esto la precisión del modelo?





ANÁLISIS DE LOS MODELOS

Gracias