

Detección de Fraude en Canastas de Compra mediante Árboles de Decisión

Franklin Manuel Manjarres

En este proyecto se implementó un modelo de detección de fraude en canastas de compra utilizando árboles de decisión y bosques aleatorios, priorizando la interpretabilidad y la identificación de patrones clave. El algoritmo `DecisionTreeClassifier` se configuró con parámetros como `criterion="entropy"`, `max_depth=None` y ajustes para clases desbalanceadas, logrando una exactitud del 88.41% en pruebas. Sin embargo, la matriz de confusión reveló un alto número de falsos negativos (197), indicando que el 80.7% de los fraudes no se detectaron. Características como `FULFILMENT CHARGE | RETAILER`, costos totales y productos específicos (Apple/Samsung) fueron determinantes en las predicciones.

Para optimizar el modelo, se ajustó el umbral de clasificación (0.3), incrementando los verdaderos positivos ($47 \rightarrow 118$) y reduciendo los falsos negativos ($197 \rightarrow 126$), pero con un aumento significativo de falsos positivos ($27 \rightarrow 157$), lo que implica mayores riesgos de bloquear transacciones legítimas. Adicionalmente, un `RandomForestClassifier` mejoró la exactitud a 89.97%, destacando características similares. El proceso incluyó limpieza de datos (reducción de 2,457 a 525 variables) y análisis de importancia de variables. En conclusión, el modelo prioriza la detección de fraude a costa de precisión en transacciones legítimas.