



Detección de Fraude Financiero

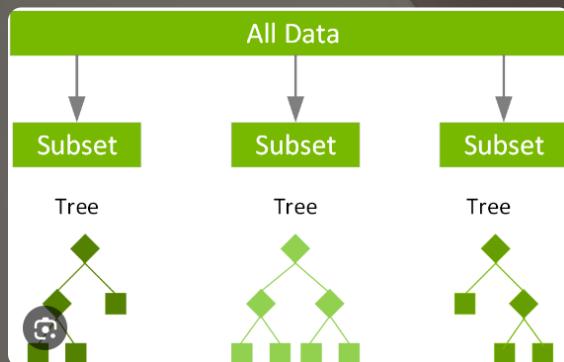
Modelo de detección de fraudes, diseñado para enfrentar desafíos cruciales en el ámbito financiero. Desde la gestión del desbalanceo de datos hasta la optimización del rendimiento, nuestro enfoque aborda obstáculos significativos en la identificación temprana de transacciones fraudulentas.

- JORGE ALZATE
- JACKELINE GÓMEZ

Modelo de Detección de Fraude

Empleamos el algoritmo XGBoost, una potente técnica de aprendizaje automático basada en árboles de decisión.

1. **Regularización:** Mecanismos incorporados, como `gamma`, `lambda`, y `alpha`, para controlar la complejidad del modelo y prevenir el sobreajuste.



step	type	amount	nameOrig	oldbalanceOrg	newbalanceOrig	nameDest	oldbalanceDest	newbalanceDest	isFraud	isFlaggedFraud	
0	1	9839.64	C1231006815	170136.00	160296.36	M1979787155	0.00	0.00	0	0	
1	1	1864.28	C1666544295	21249.00	19384.72	M2044282225	0.00	0.00	0	0	
2	1	TRANSFER	181.00	C1305486145	181.00	0.00	C553264065	0.00	0.00	1	0
3	1	CASH_OUT	181.00	C840083671	181.00	0.00	C38997010	21182.00	0.00	1	0
4	1	PAYMENT	11668.14	C2048537720	41554.00	29885.86	M1230701703	0.00	0.00	0	0
...	
101608	10	PAYMENT	7477.02	C513257306	785.00	0.00	M524833426	0.00	0.00	0	0
101609	10	CASH_OUT	282252.35	C210473293	220339.29	0.00	C1172042998	95156.01	91286.97	0	0
101610	10	PAYMENT	17289.01	C807582280	0.00	0.00	M334249577	0.00	0.00	0	0
101611	10	TRANSFER	1347146.45	C1315779140	1605.00	0.00	C1631408038	541639.43	2186953.43	0	0
101612	10	CASH_OUT	469539.21	C515691325	31133.00	0.00	C1383702768	0.00	469539.21	0	0



Transformaciones realizadas

- **Codificación Categórica:** One-Hot Encoding para "Type" mediante `pd.get_dummies`.
- **Escalado Numérico:** Min-Max Scaling en "Step", "Amount", "OldBalanceOrg", "NewBalanceOrig", "OldBalanceDest", y "NewBalanceDest".
- **Eliminación de Columnas:** "NameOrig" y "NameDest" removidas por irrelevancia.
- **Manejo de Datos Faltantes:** Identificación y eliminación de valores nulos en el conjunto de datos.

Balanceo de Clases con SMOTE:

- **SMOTE:** Aplicamos la técnica Synthetic Minority Over-sampling Technique (SMOTE) para balancear las clases de "isFraud".
- **Nuevo DataFrame:** Creamos un nuevo DataFrame (df_resampled) con variables escaladas y clases balanceadas, utilizando las muestras generadas por SMOTE.

	amount	oldbalanceOrg	newbalanceOrig	oldbalanceDest	newbalanceDest	isFlaggedFraud	type_CASH_IN	type_CASH_OUT	type_DEBIT	type_PAYMENT	type_TRANSFER	isFraud
0	0.000984	0.004369	0.004116	0.000000	0.000000	0	0	0	0	1	0	0
1	0.000186	0.000546	0.000498	0.000000	0.000000	0	0	0	0	1	0	0
2	0.000018	0.000005	0.000000	0.000000	0.000000	0	0	0	0	0	1	1
3	0.000018	0.000005	0.000000	0.000623	0.000000	0	0	1	0	0	0	1
4	0.001167	0.001067	0.000767	0.000000	0.000000	0	0	0	0	1	0	0
...
202989	0.001259	0.000323	0.000000	0.000000	0.000000	0	0	0	0	0	1	1
202990	0.000249	0.000064	0.000000	0.000000	0.000000	0	0	0	0	0	1	1
202991	0.003321	0.000853	0.000000	0.000710	0.000127	0	0	1	0	0	0	1
202992	0.009791	0.002514	0.000000	0.000000	0.000000	0	0	0	0	0	1	1
202993	0.014923	0.001993	0.000000	0.000000	0.003832	0	0	1	0	0	0	1
202994	0.000000	0.000000	0.000000	0.000000	0.000000	0	0	0	0	0	0	0

Definir parámetros para XGBoost

Definir parámetros para XGBoost

```
# Definir parámetros para XGBoost
params = {
    'eta': 0.01,
    'max_depth': 5,
    'min_child_weight': 1,
    'subsample': 0.8,
    'colsample_bytree': 0.8,
    'gamma': 0.1,
    'lambda': 1,
    'alpha': 0.1,
    'objective': 'binary:logistic',
    'n_estimators': 1000
}
# Crear y entrenar el modelo XGBoost con ajustes adicionales
xgb_model = xgb.XGBClassifier(**params)
xgb_model.fit(
    X_train, y_train,
    eval_metric="auc",
    eval_set=[(X_train, y_train), (X_test, y_test)],
    early_stopping_rounds=50,
    verbose=10
)
```

Criterios de Detección de Fraude

1. **Tipos de Transacción (type)**: Analiza patrones distintivos en los diferentes tipos de transacciones.
2. **Montos de Transacción (amount)**: Evalúa la magnitud de las transacciones.
3. **Historial del Cliente (oldbalanceOrg, newbalanceOrig)**: Examina el historial de transacciones de los clientes.
4. **Patrones de Destinatario (oldbalanceDest, newbalanceDest)**: Evalúa las transacciones con destinatarios específicos.
5. **Marcador de Fraude (isFlaggedFraud)**: Considera transacciones marcadas como intentos ilegales de transferir grandes cantidades de dinero en una sola operación.

Resultados del Modelo y su Exactitud

El modelo logra una precisión sólida, con un valor de precisión general del 0.9852 y una métrica ROC AUC de 0.9904. Estos resultados respaldan la eficacia del modelo en comparación con métodos convencionales.

