

Inteligencia Artificial

Act 13: Programando Random Forest en Python

Estudiante: José Luis Calderón Galarza - 2132939

Docente: Luis Ángel Gutiérrez Rodríguez

1 Introducción

Random Forest es un algoritmo de aprendizaje supervisado basado en la combinación de múltiples árboles de decisión. Su principal ventaja radica en su capacidad para reducir el sobreajuste y mejorar la precisión del modelo mediante el uso de un conjunto de árboles que trabajan en paralelo. Este método es ampliamente utilizado en tareas de clasificación y regresión, destacándose por su robustez y efectividad en conjuntos de datos desequilibrados. En este reporte, se explora la aplicación de Random Forest para la detección de transacciones fraudulentas en tarjetas de crédito.

2 Metodología

Se utilizó el conjunto de datos `creditcard.csv`, el cual contiene transacciones etiquetadas como normales o fraudulentas. Se llevaron a cabo los siguientes pasos:

1. Carga y exploración de datos:

```
df = pd.read_csv("creditcard.csv")
print(df.head(n=5))
print(df.shape)
```

2. División del conjunto de datos:

```
y = df['Class']
X = df.drop('Class', axis=1)
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.7)
```

3. Entrenamiento de regresión logística:

```
clf = LogisticRegression(C=1.0, penalty='l2', random_state=1, solver="newton-cg", class_weight='balanced')
clf.fit(X_train, y_train)
```

4. Evaluación del modelo:

```
pred_y = clf.predict(X_test)
mostrar_resultados(y_test, pred_y)
```

5. Entrenamiento de bosques aleatorios:

```
model = RandomForestClassifier(n_estimators=100, max_features='sqrt', class_weight="balanced")
model.fit(X_train, y_train)
```

3 Resultados

3.1 Matriz de Confusión 1

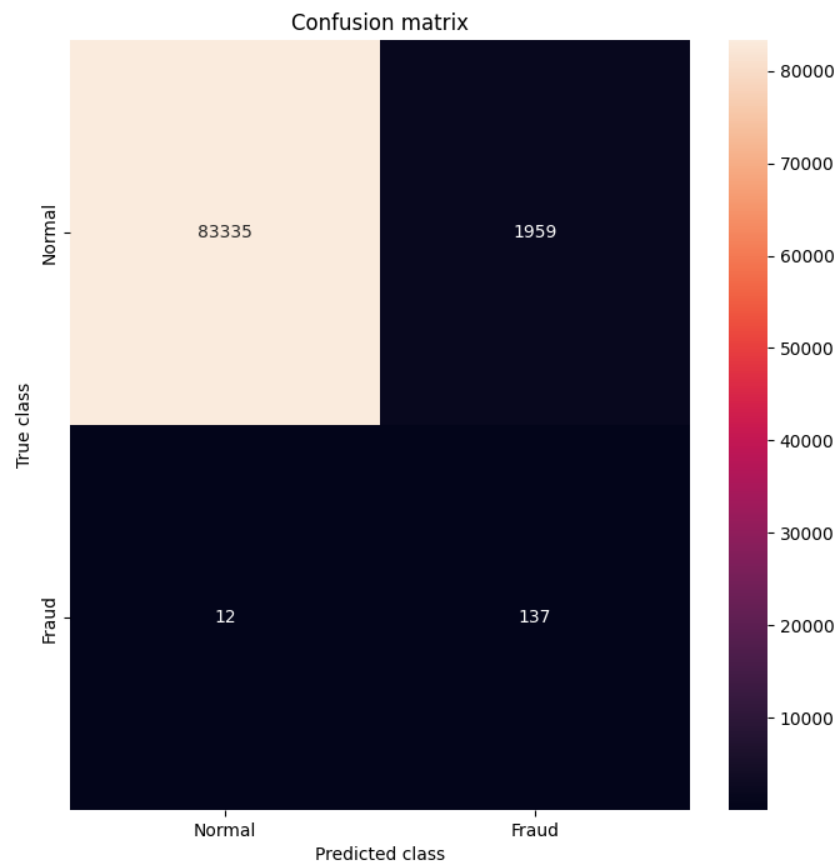


Figure 1: Matriz de confusión del modelo.

3.2 Matriz de Confusión 2

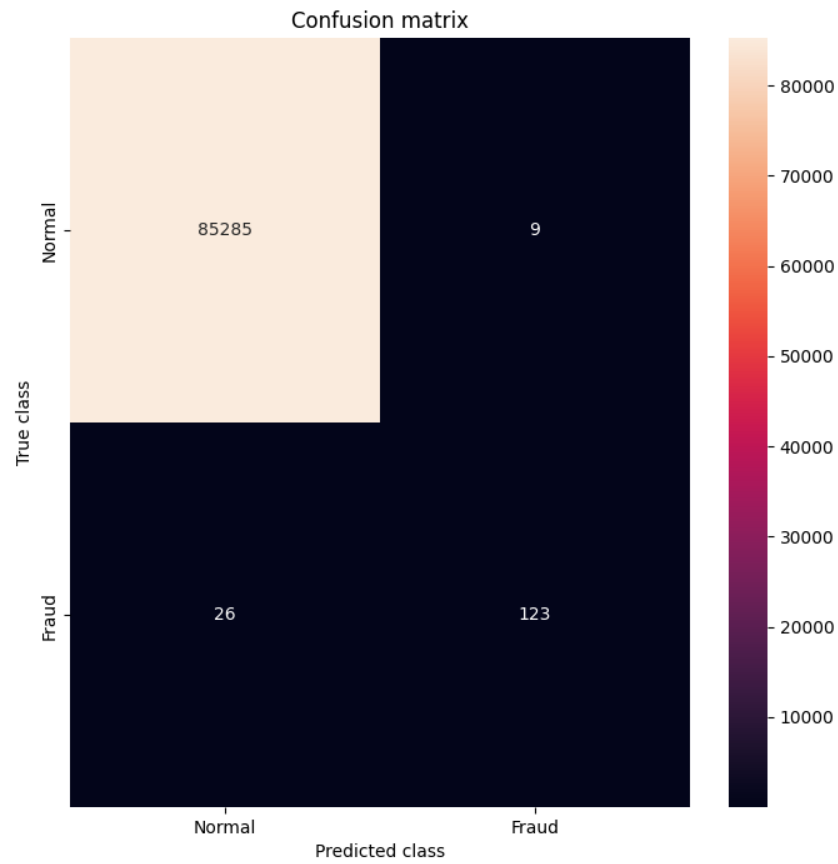


Figure 2: Matriz de confusión del modelo.

3.3 Matriz de Confusión 3

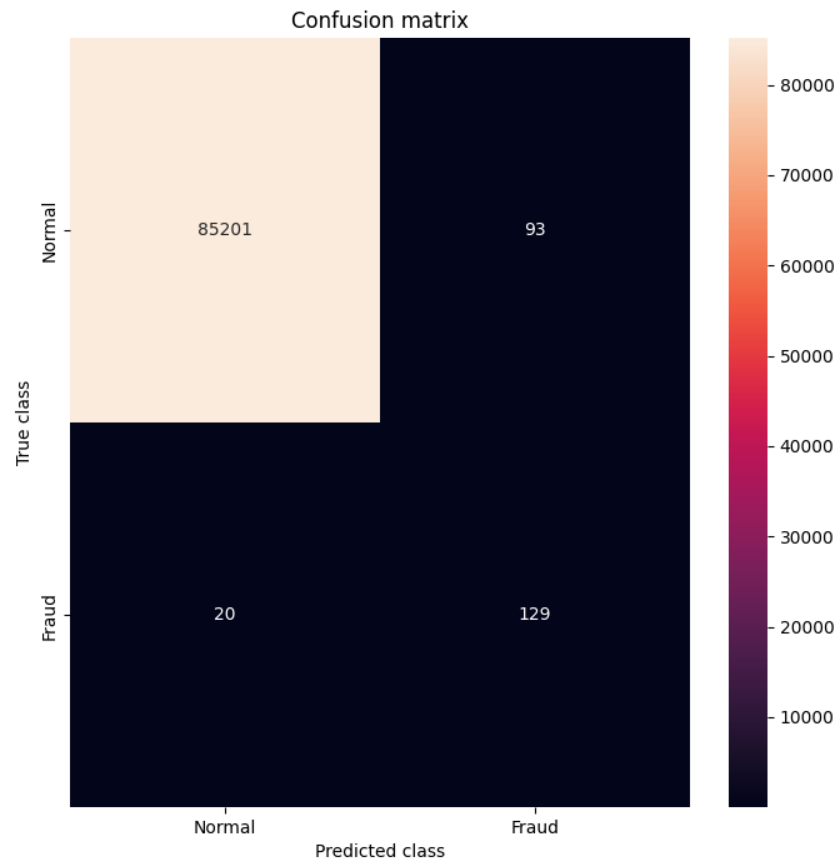


Figure 3: Matriz de confusión del modelo.

3.4 ROC AUC Score

El *ROC AUC score* obtenido es:

$$0.93$$

Este valor indica un buen desempeño del modelo en la clasificación de fraudes.

4 Conclusión

El uso de Random Forest en la detección de fraudes en transacciones con tarjeta de crédito ha demostrado ser efectivo, logrando un alto grado de precisión. La combinación de múltiples árboles de decisión ha permitido mejorar la capacidad de generalización del modelo, reduciendo el sobreajuste. Sin embargo, aún existen casos de falsos negativos que podrían minimizarse mediante un ajuste fino de hiperparámetros y técnicas avanzadas de balanceo de datos. Este experimento

confirma la importancia de Random Forest como una herramienta robusta para la clasificación de datos en problemas de detección de anomalías.