CAPSTONE PROJECT

Project Name : FindDefault (Prediction of Credit Card fraud) Project Report

Name : Pranav Lokhande
Email : PranavLokhande504@gmail.com

1.Problem Statement :

Credit card fraud involves unauthorized use of card information for purchases or cash withdrawals. It's crucial for companies to detect fraudulent transactions to protect customers. The provided dataset includes credit card transactions from September 2013, involving 284,807 transactions, of which only 492 are fraudulent (0.172%). The goal is to develop a classification model to predict whether a transaction is fraudulent.

2.Dataset Overview :

Total Transactions: 284,807
Fraudulent Transactions: 492

3. Exploratory Data Analysis (EDA)

A count plot showing the distribution of fraudulent vs non-fraudulent transactions.
A correlation heatmap to visualize the relationships between different features in the dataset.

4. Data Cleaning and Preprocessing

Standardization: The Amount feature was standardized.
Missing Values: No missing values were found in the dataset.
Time was initially considered but later dropped due to low correlation with the target variable

5.Dealing with Imbalanced Data

Over-sampling: Increasing the number of fraudulent transactions using techniques like SMOTE (Synthetic Minority Over-sampling Technique) imblearn.over_sampling ADASYN

6. Feature Engineering

Interaction Terms: Created interaction terms to capture relationships between features.
Feature Scaling: Applied Min-Max scaling to ensure all features are on the same scale

7.Model Building and Evaluation

We make
Logistic Regression
Decision Tree
Random Forest
XGBoost

Model Performance Metrics:

Base on the all model performance we select the best model
Recall (Sensitivity): The ratio of correctly predicted fraudulent transactions to all actual fraudulent transactions.

Model Results:

Model	AUC	Avg Precision
Logistic Regression	0.9302	0.5059
Decision Tree	0.9045	0.6559
Random Forest	0.9495	0.8274
XGBoost	0.9676	0.8685

Model Deployment:

The best-performing model is serialize using the pickle module for deployment We make the model deploy use to pickle file best_xgb model First to Save the model to a file

```
import pickle

# Save the best model and scaler
model_filename = 'best_fraud_Prediction_model.pkl'  # Model saved as .pkl
scaler_filename = 'scaler.pkl'  # Scaler saved as .pkl
```

```
# Save model and scaler using pickle
with open(model_filename, 'wb') as model_file:
    pickle.dump(best_model_instance, model_file)

with open(scaler_filename, 'wb') as scaler_file:
    pickle.dump(scaler, scaler_file)

print(f"Best model saved as {model_filename}")
print(f"Scaler saved as {scaler_filename}")
```

```
predictions = loaded_model.predict(X_test)

# Output the predictions
print("Predictions on test data:", predictions)
```

Conclusion:

This project effectively applied machine learning techniques to detect fraudulent transactions in a highly imbalanced dataset. By addressing class imbalance with methods like SMOTE and experimenting with various classifiers, we achieved strong performance metrics, particularly in precision and recall, both crucial in fraud detection.

The model's high AUC score demonstrates its ability to differentiate between fraud and legitimate transactions.