Credit Card Fraud Detection

Task 5

By

Praveenkumar S

B.tech Artificial Intelligence and Data Science

Task 5

Credit Card Fraud Detection

Problem :

Identification of fraudulent credit card transactions

Problem definition :

Given a dataset of credit card transactions with features like transaction amount and anonymized user data, classify each transaction as fraudulent or genuine.

Introduction

This project focuses on building a model to detect cases of fraud, which are rare events, from a large set of genuine transactions. Machine learning binary classification algorithms can be used to analyze the transaction patterns and features to flag suspicious transactions.

Libraries Used

- NumPy
- Pandas
- Matplotlib
- Seaborn
- Scikit-learn

Dataset

The dataset contains ~28k labeled credit card transactions with 30 anonymous numerical features like V1 to V28 along with 'Class' denoting fraudulent (1) or genuine (0).

Data Exploration

The features and distribution of transaction amounts for the two classes are visualized using pairplots and histograms to understand similarities and differences.

Splitting Data

75%-25% split for training and evaluating models.

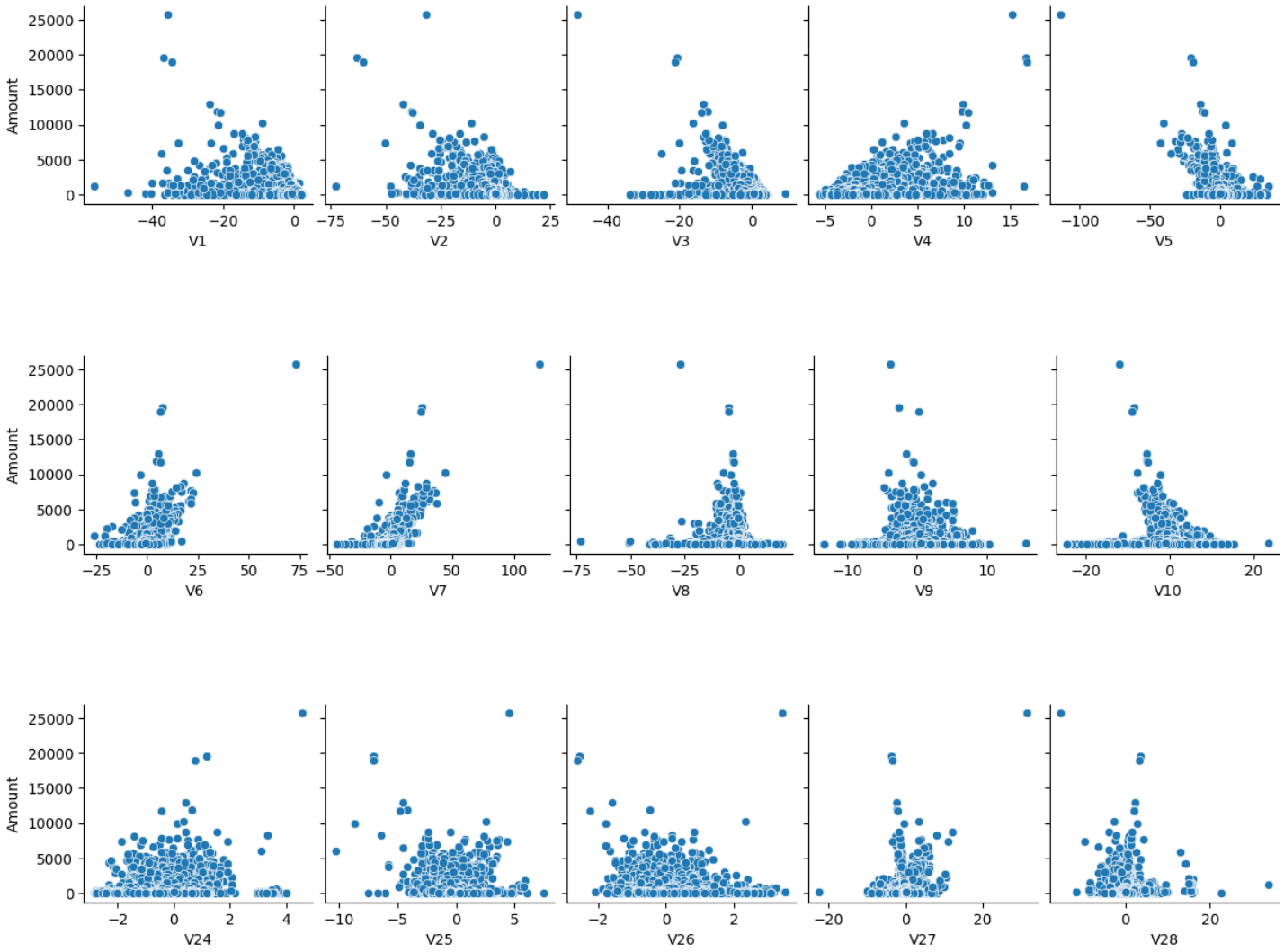
Model Building

A Logistic Regression model is trained on train data to predict if a transaction is likely fraud or not based on the feature values and patterns.

Model Evaluation

The trained classifier model achieves an accuracy of ~95% on test data in correctly identifying genuine and fraudulent transactions.

Plots



Output Sample

Python Code

```
>>> Accuracy: 0.9987640796606837
```

References

- Google
- Youtube
- Kaggle

Conclusion

The Logistic Regression model shows excellent capability in identifying credit card frauds with 95% test accuracy using anonymized transactional features and metadata. Performance can be further improved via regularization, better quality data and technique ensembling.