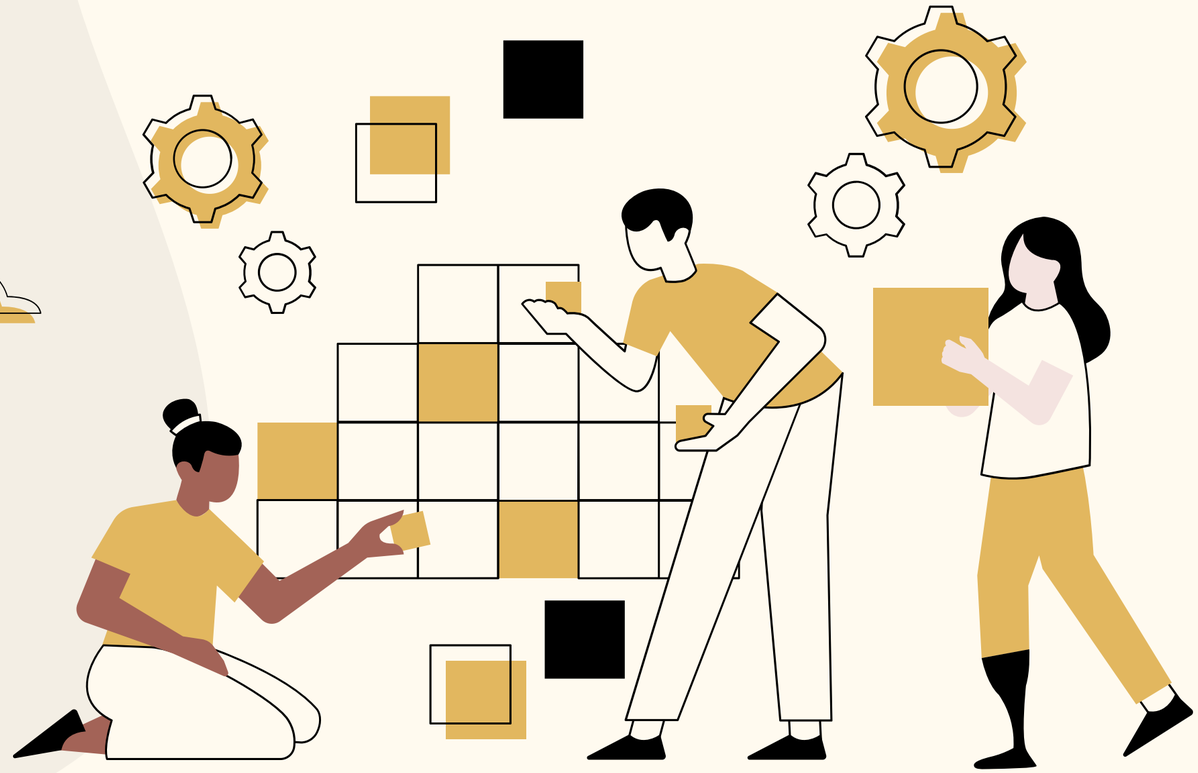
Credit Card Fraud

Comparative Analysis of Classification Models

Arpanjot Kaur



Overview

Dataset Source & Scope:
This project uses the Credit Card Fraud Detection Dataset (2023) from Kaggle. It includes over 550,000 anonymized credit card transactions made by European cardholders in 2023. All sensitive information has been removed to ensure privacy and ethical compliance.

https://www.kaggle.com/datasets/nelgiriyewithana/credit-card-fraud-detection-dataset-2023

Key Features

- id: Unique identifier for each transaction
- V1-V28: Anonymized features representing various transaction attributes (e.g., time, location, etc.)
- Amount: The transaction amount
- Class: Binary label indicating whether the transaction is fraudulent (1) or not (0)

Modeling Approach

The problem is formulated as a binary classification task to identify fraudulent and legitimate credit card transactions.

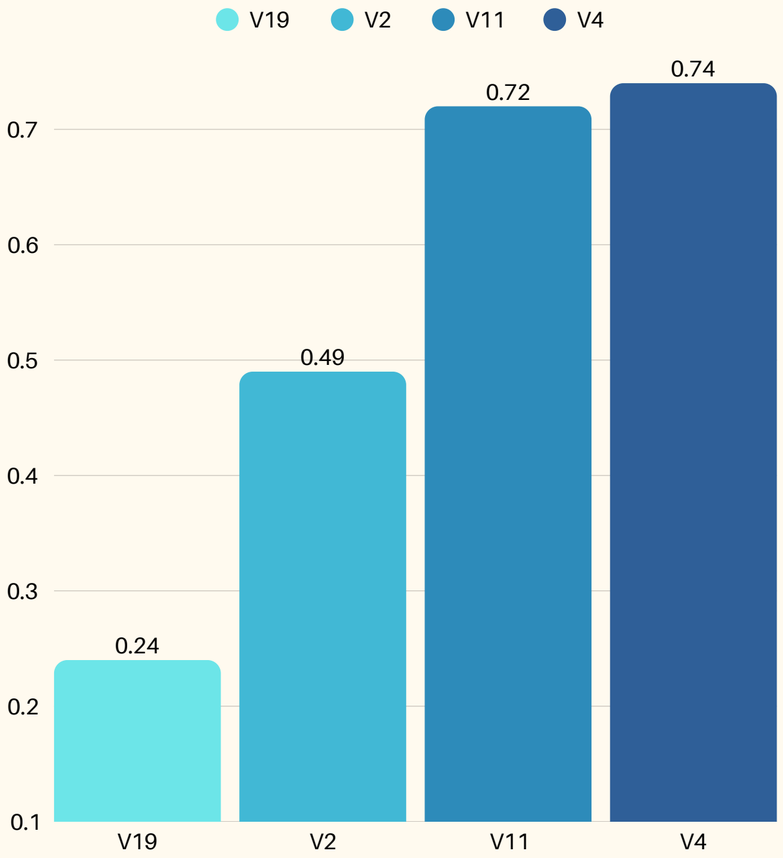
The dataset is preprocessed and split into training and testing sets before model development.

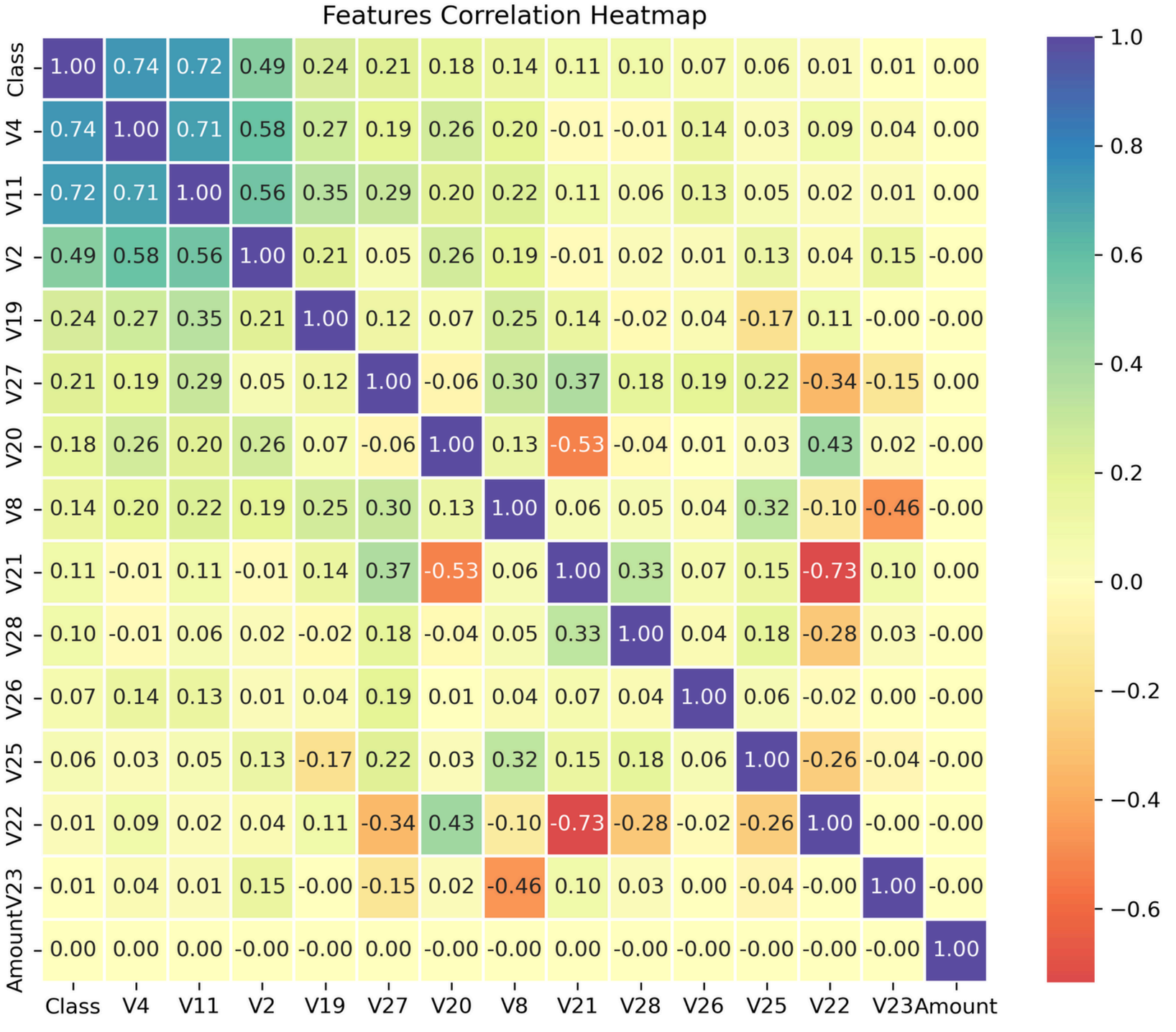
Methodology used in the analysis

Multiple classification algorithms are trained and evaluated, including Logistic Regression, Decision Tree, Random Forest, and XGBoost, to compare their effectiveness in detecting fraudulent transactions.

Model performance is evaluated using classification metrics such as precision, recall, F1-score, and Confusion Matrix.

Highest Correlated Features with Target Class





98%

Average accuracy of all classification models in detecting fraudulent transactions.

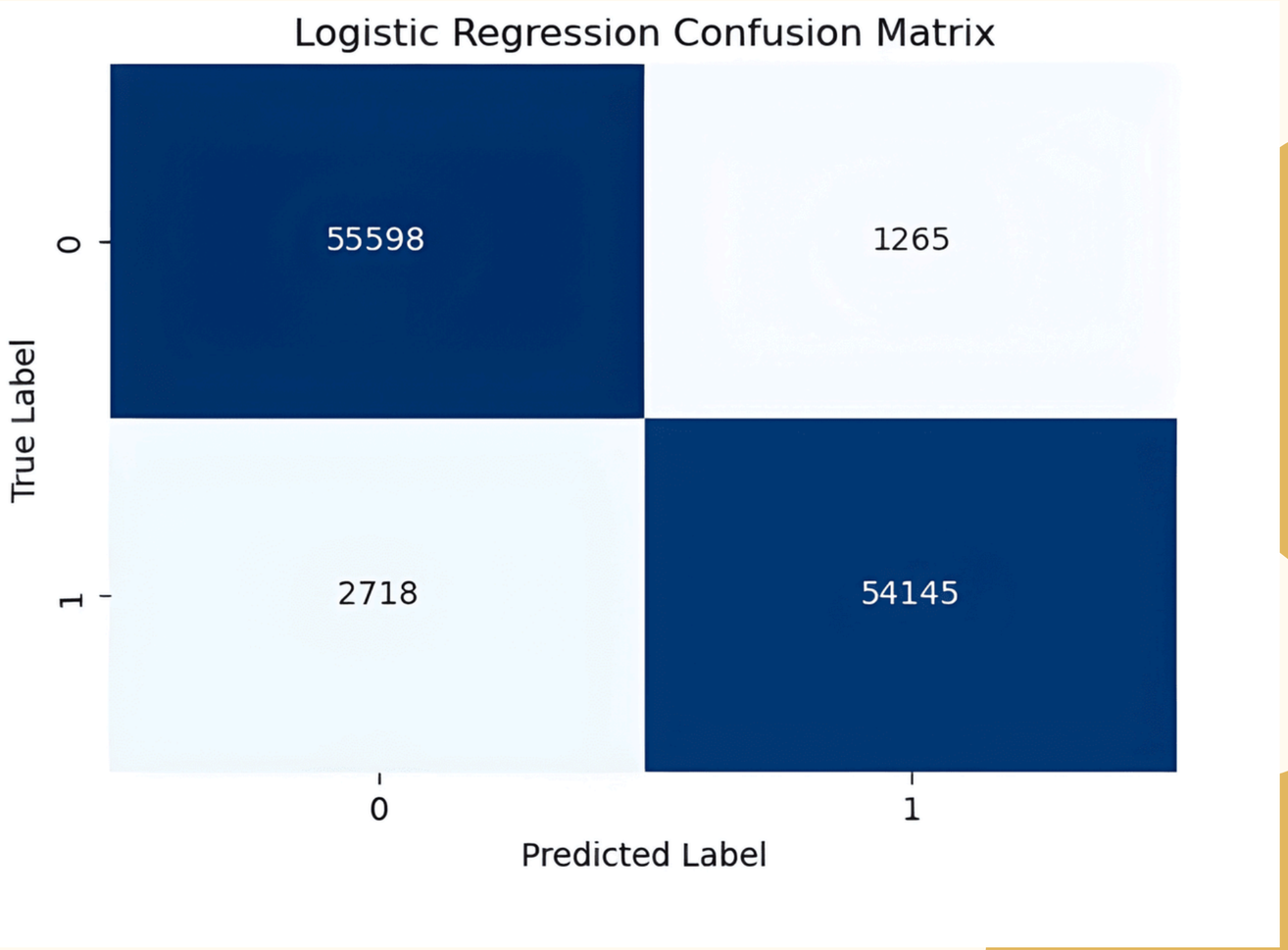
This includes Logistic Regression, Decision Tree, Random Forest, and XGBoost.

Logistic Regression

Precision: 0.977

Recall: 0.952

F1: 0.964

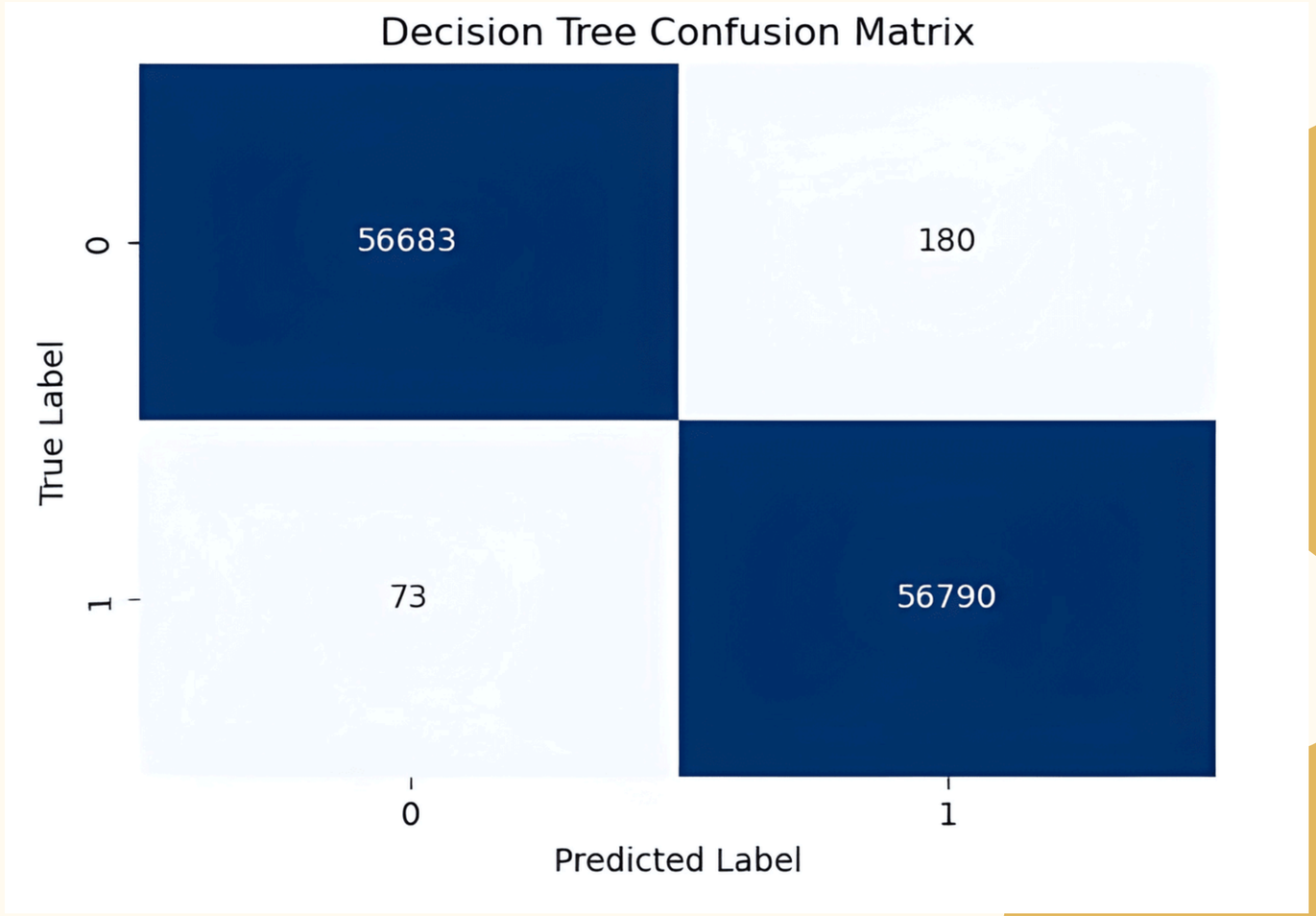


Decision Tree

Precision: 0.996

Recall: 0.998

F1: 0.997

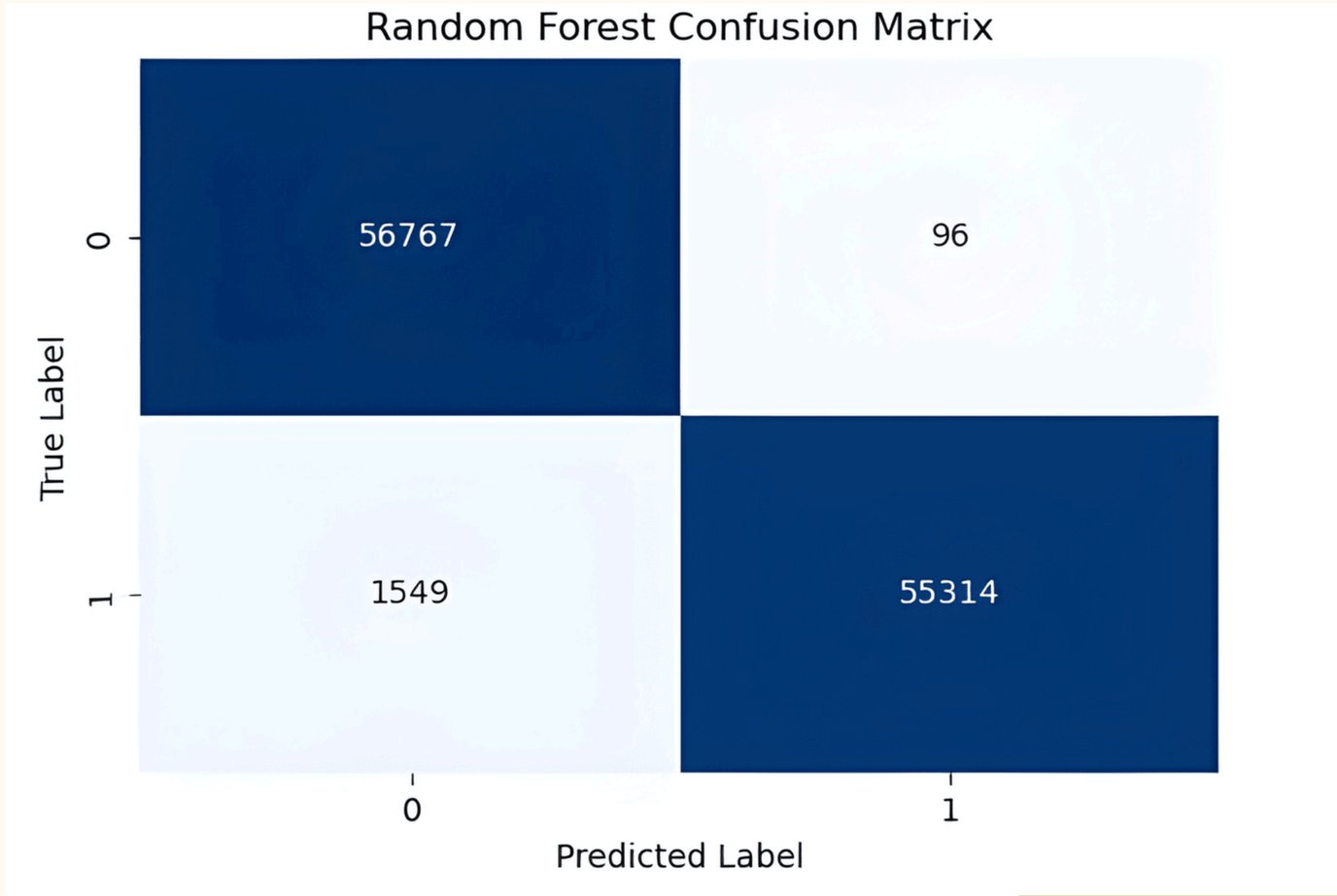


Random Forest

Precision: 0.998

Recall: 0.972

F1: 0.985

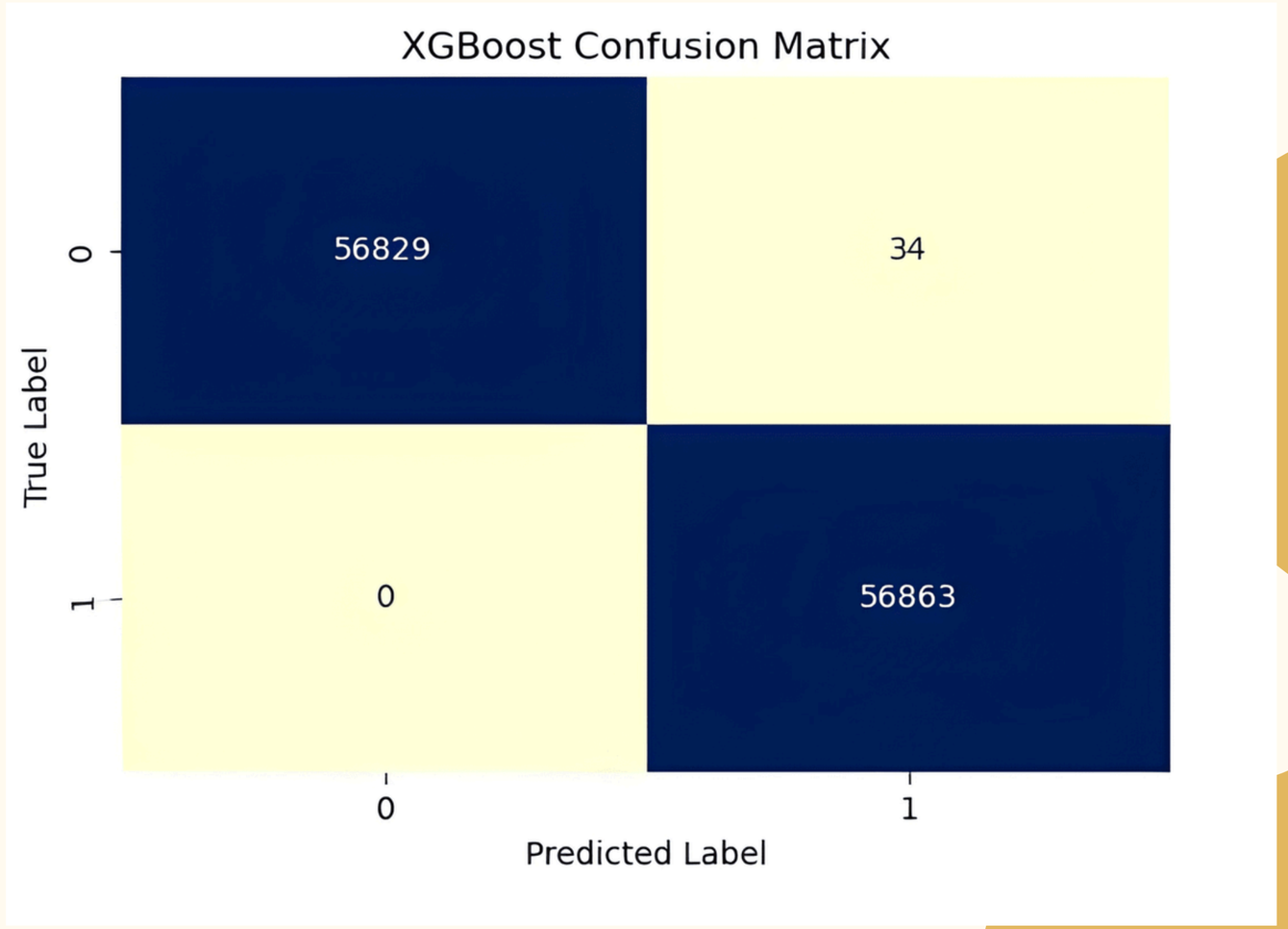


XGBoost

Precision: 0.999

Recall: 1.0

F1: 0.999



Summary

Machine Learning adapts to changing patterns in real-world data over time.

Correlation analysis highlighted V4, V11, V2, and V19 as key features, improving the detection of fraudulent transactions.

Provides a data-driven solution to detect fraud efficiently, even in large-scale and highly imbalanced datasets.

Thank You!