Fraud detection in financial transactions

In this Presentation

Here's what we'll cover:

Inroduction

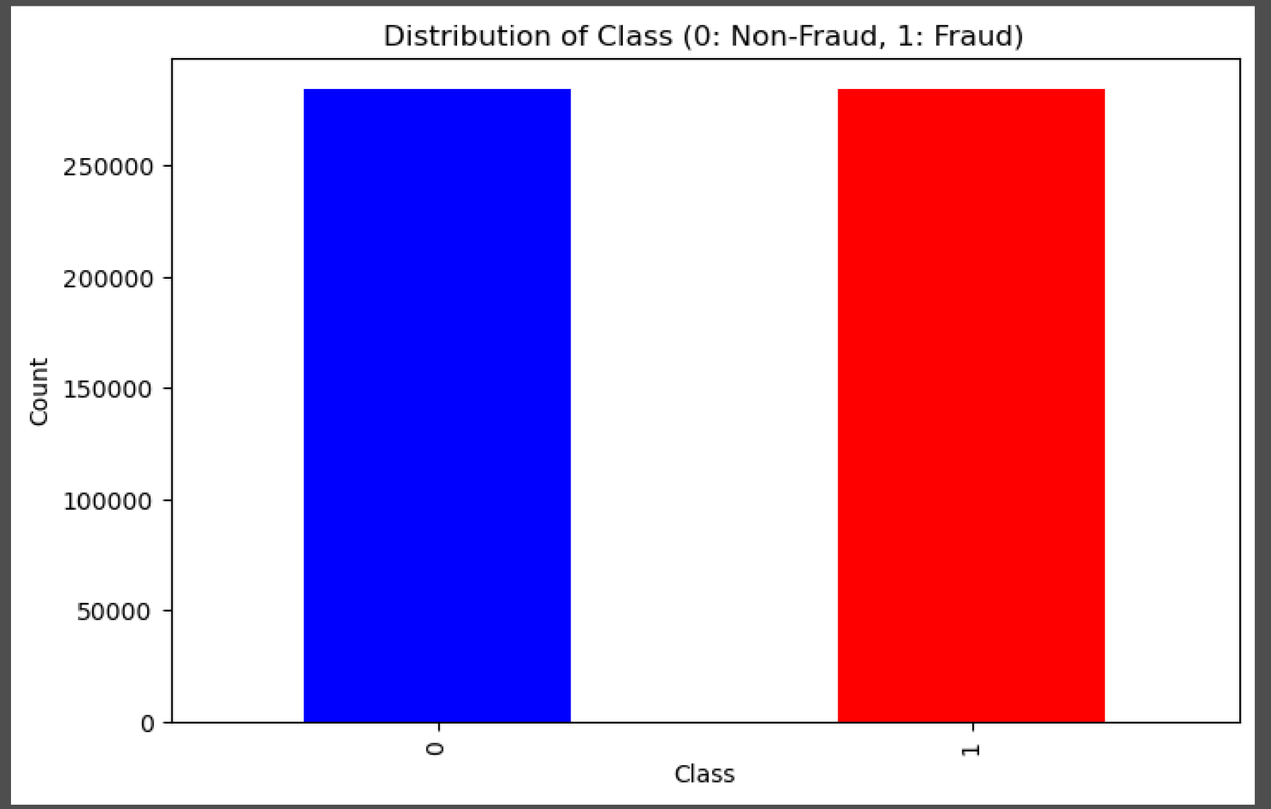
Credit card fraud detection is a set of methods and techniques designed to block fraudulent purchases, both online and in-store. This is done by ensuring that you are dealing with the right cardholder and that the purchase is legitimate.

Use Case and Data Set

Dataset was taken from Kaggle

This dataset contains credit card transactions made by European cardholders in the year 2023. It comprises over 550,000 records, and the data has been anonymized to protect the cardholders' identities. The primary objective of this dataset is to facilitate the development of fraud detection algorithms and models to identify potentially fraudulent transactions.

In fraud detection, achieving a balanced dataset, where the number of fraudulent and non-fraudulent transactions is similar, is crucial. This balance is fundamental for accurate model evaluation. It ensures that the model isn't biased toward the majority class, providing a more realistic representation of its performance.



Data Quality Assessment and Exploration

missing values

There were no missing values, ensuring a robust foundation for analysis.

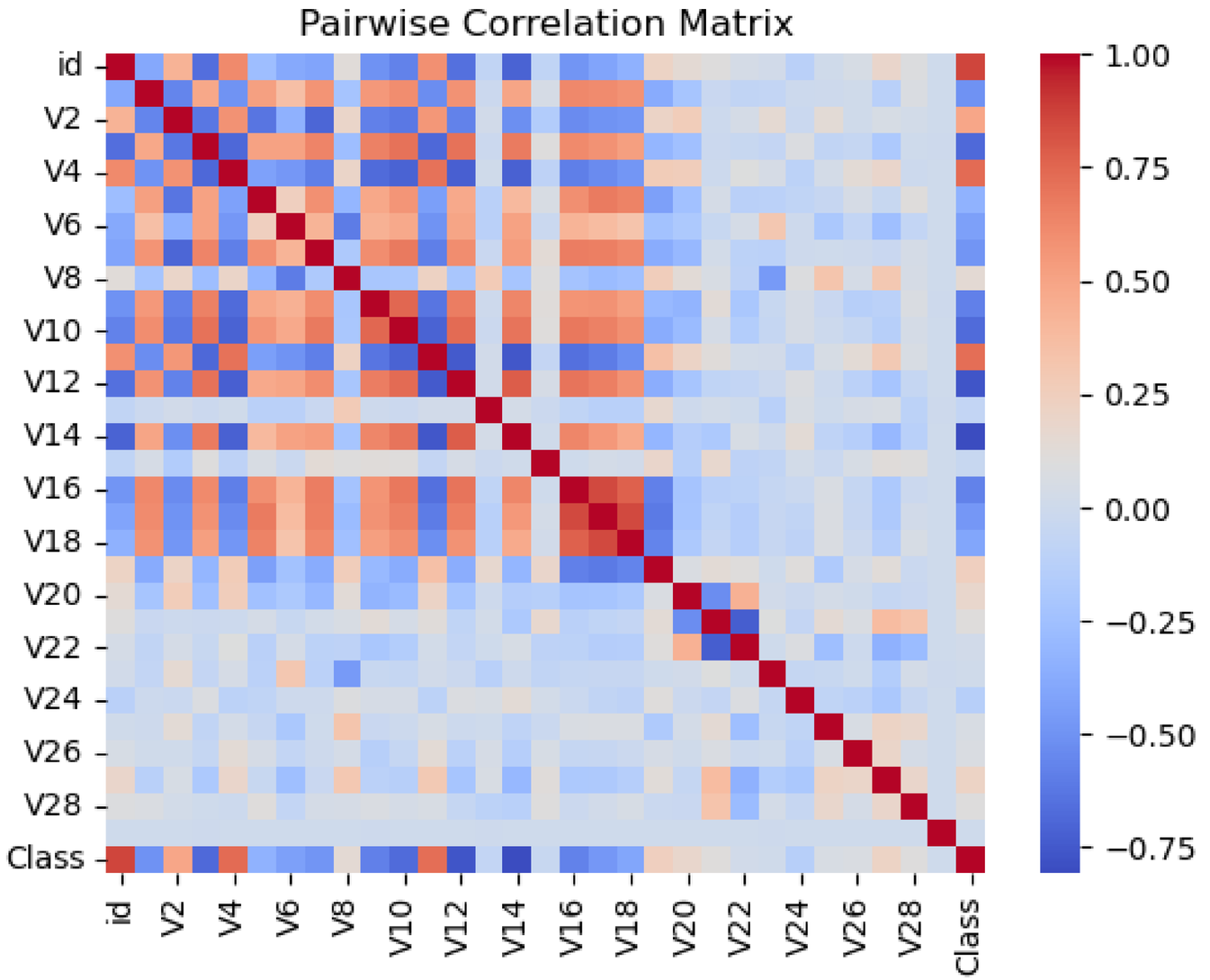
Outliers

There were no outliers, ensuring a robust foundation for analysis.

Multicollinearity

Variance Inflation Factor (VIF). Here, I present a snippet of the VIF values for each variable. The good news is that none of the VIF values exceeded the commonly accepted threshold of 10, indicating no issues of multicollinearity among the features.

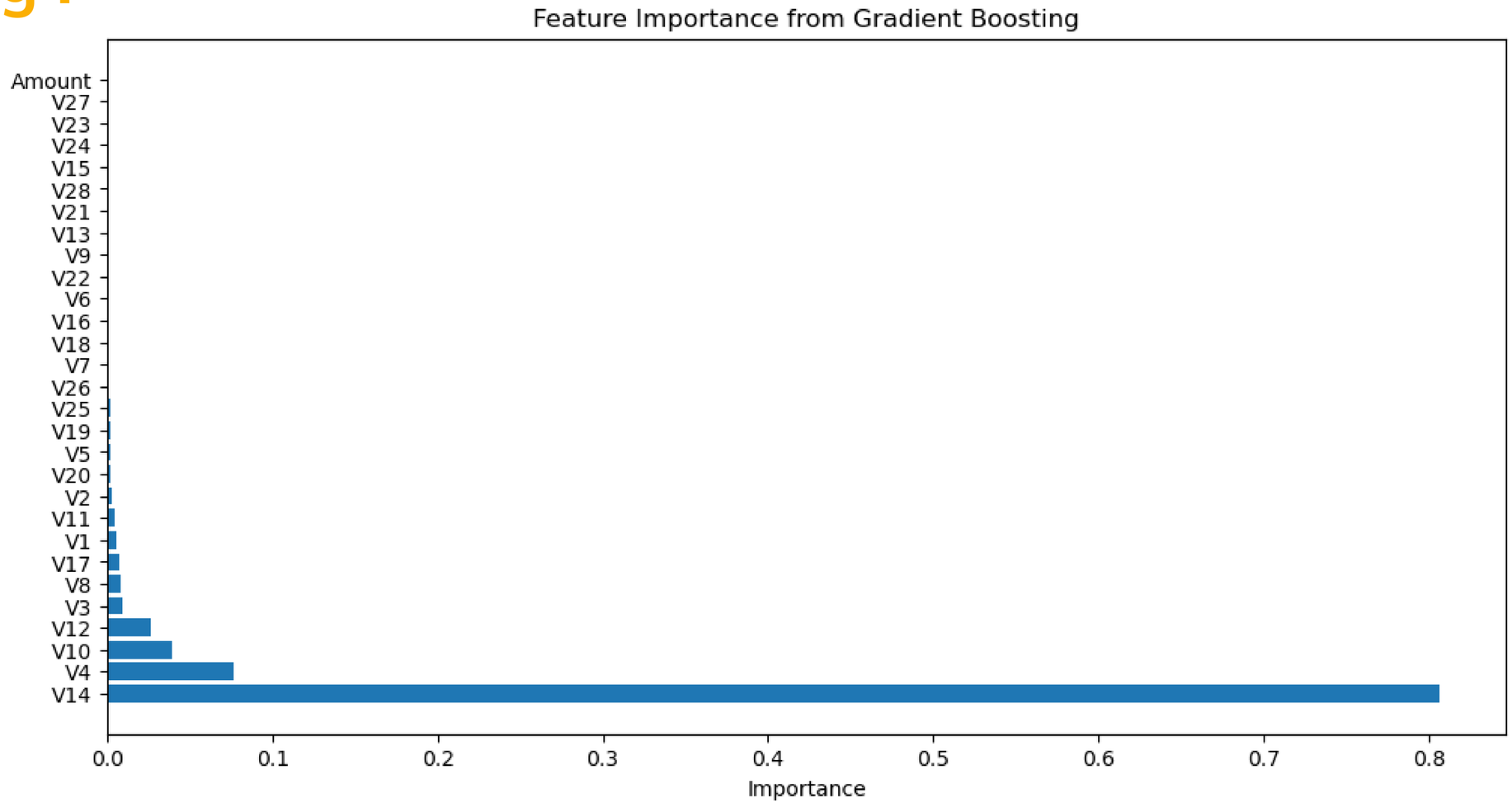
Correlation matrix



Feature Engineering and Model Performance Indicators

Feature Engineering :

We’ve used the gradient boosting algorithm technique due to its iterative in order to do the feature selection.



Model Performance Indicators

Accuracy

It provides an overall measure of correct predictions, suitable when classes are balanced.

F1 score

Balancing precision and recall, F1 score is crucial for imbalanced datasets. It considers both false positives and false negatives, making it a robust metric.

Confusion Matrix

Breaks down the model's performance, helping to understand true positives, true negatives, false positives, and false negatives. This is essential, especially in fraud detection, where the cost of false positives and false negatives can be significantly different.

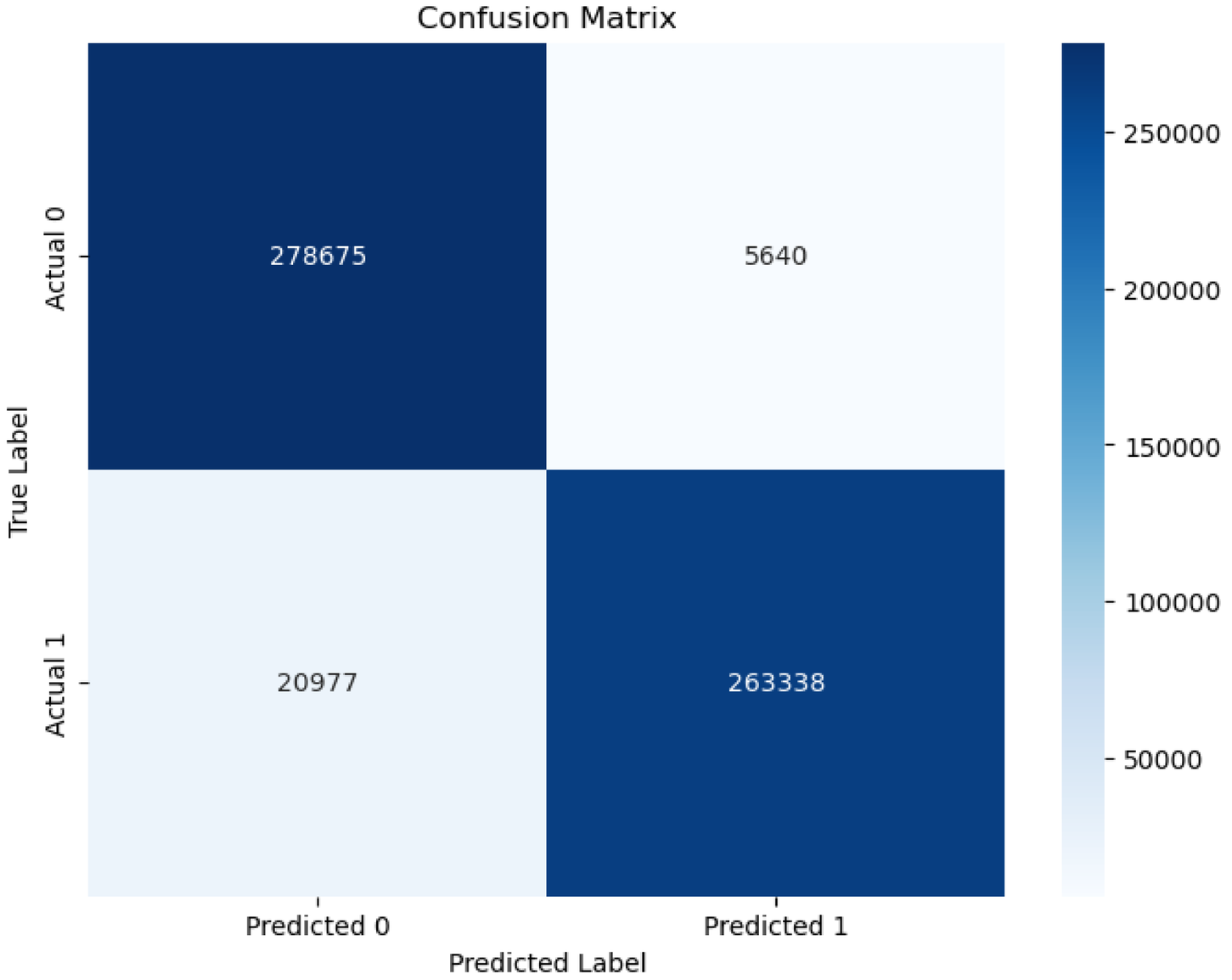
Algorithms and Model Performance Comparison

Machine Learning Models :

- Logistic Regression
- Random Forest
- Support Vector Machine

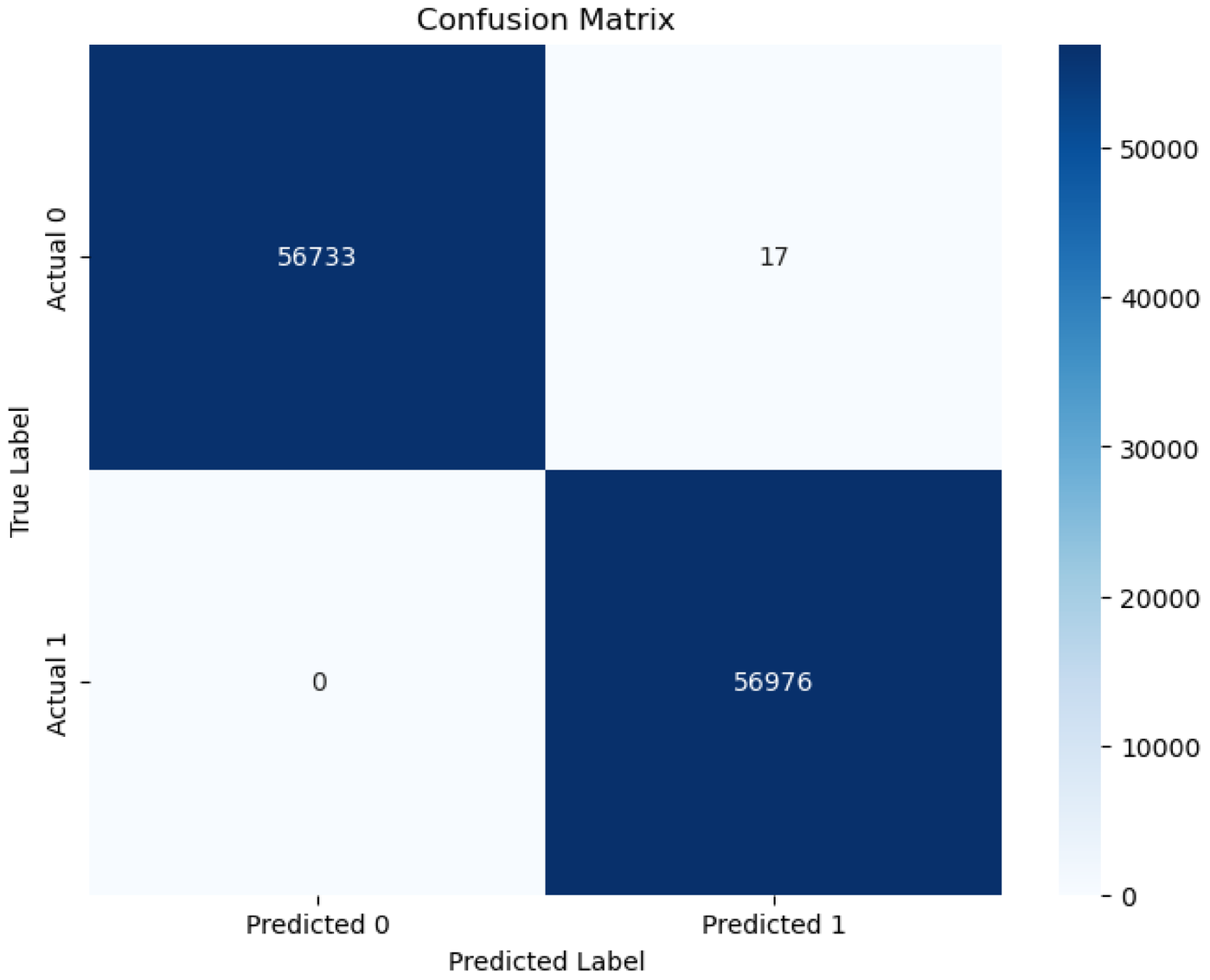
Deep Learning Model :

- Multiple-Layer Perceptron



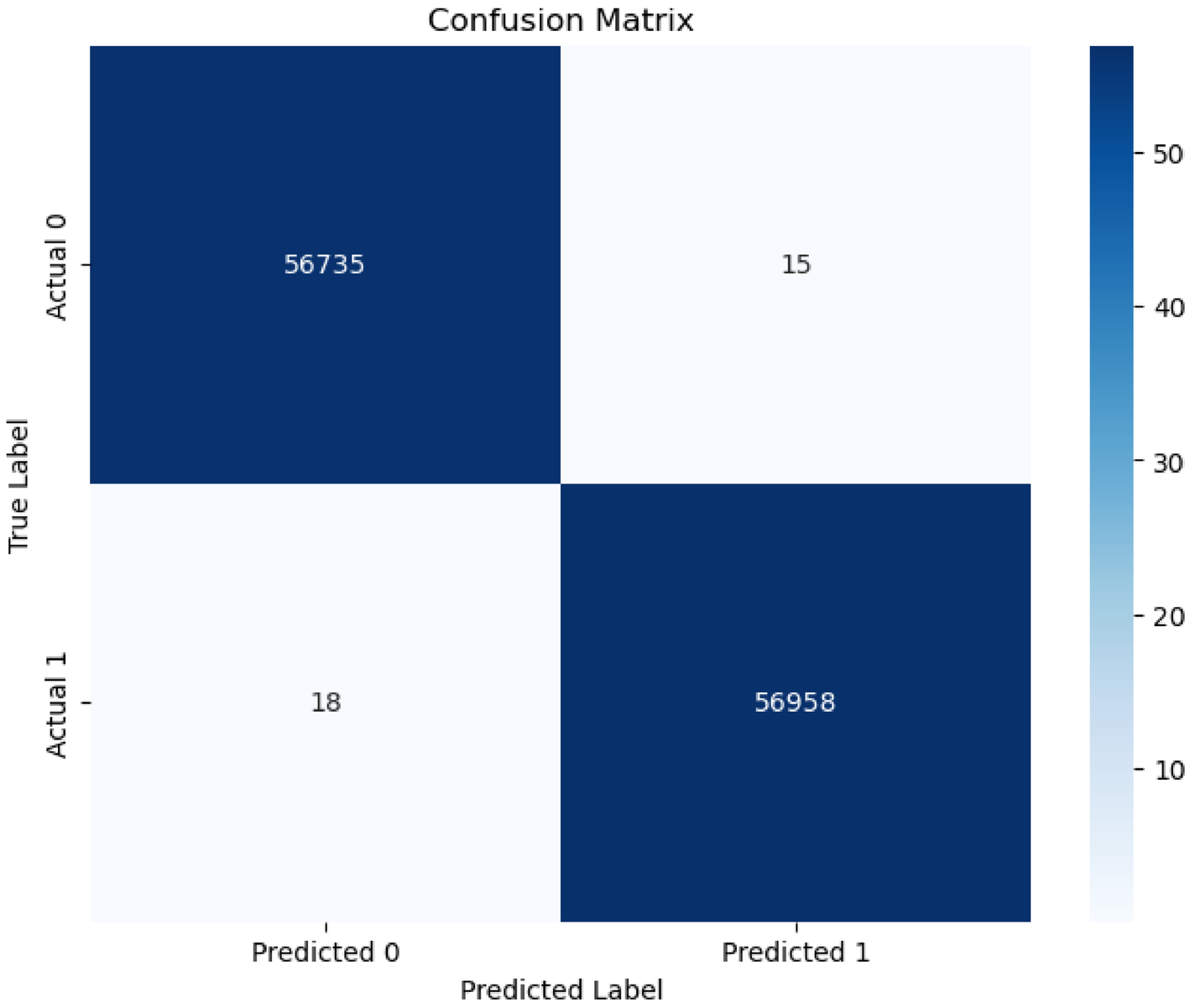
Logistic Regression:

Random Forest:



Deep Learning Models :

TensorFlow Keras-based MLP neural network



Why Specific Methods Were Chosen

Feature Selection :

Gradient Boosting Algorithm

Implementing models:

- scikit-learn
- Keras & Tensorflow

ML Algorithm

- Random Forest
- Support Vector Machine
- Logistic Regression

DL Algorithm

- Multiple-Layer Perceptron

Conclusion

- **Project Focus**: My project centers on the vital task of credit card fraud detection, addressing the significance of securing financial transactions.
- **Key Achievements**:
  - Robust Models: Developed robust machine learning and deep learning models to effectively identify fraudulent transactions.
  - Insightful Exploration: Conducted thorough data exploration, revealing patterns and insights crucial for model development.
- **Challenges and Learnings**:
  - Optimization: Learned valuable lessons in optimizing model performance, emphasizing the importance of thoughtful algorithm selection.
- **Continuous Learning**

Thank you for your watching