Credit Card Fraud Detection

Presented by – GROUP 13

Branch/Section – AI-D

Group 13 members-

Shivam Kumar (202401100300232)

Shivam Agarwal (202401100300231)

Shikhar Maheshwari(202401100300230)

Shriyans Jain(202401100300242)

Shubham Mishra(202401100300245)

Sidh Agarwal(202401100300247)

01 Introduction

- Credit card fraud is a major global issue causing substantial financial losses.
- Fraudulent patterns constantly evolve, making detection challenging.
- Manual detection is inefficient and error-prone.
- Machine learning, especially unsupervised techniques, helps detect unknown fraud patterns without labeled data.

02 Problem Statement

- Fraudulent transactions are rare and hard to label.
- Supervised models need labeled data, which is not always available.
- Aim: Detect anomalies (fraudulent behavior) in transactional data using unsupervised learning.
- Focus: Minimize false negatives to avoid missed fraud cases.

Dataset Overview

- Dataset Source: Kaggle Credit Card Fraud Detection Dataset.
- Total Transactions: 284,807.
- Fraudulent Transactions: 492 (≈0.17%).
- Features: 30 anonymized columns due to privacy.
- Highly imbalanced dataset poses a challenge for detection.

Data Preprocessing

- caled numerical features using StandardScaler.
- Applied PCA (Principal Component Analysis) for dimensionality reduction.
- Handled missing values and outliers.
- Split dataset into training and testing sets.
- Ensured no data leakage between training and evaluation phases.

Methodology

Isolation Forest

Concept: Based on the idea that anomalies are data points that are few and different. It isolates anomalies instead of profiling normal data.

One-Class SVM

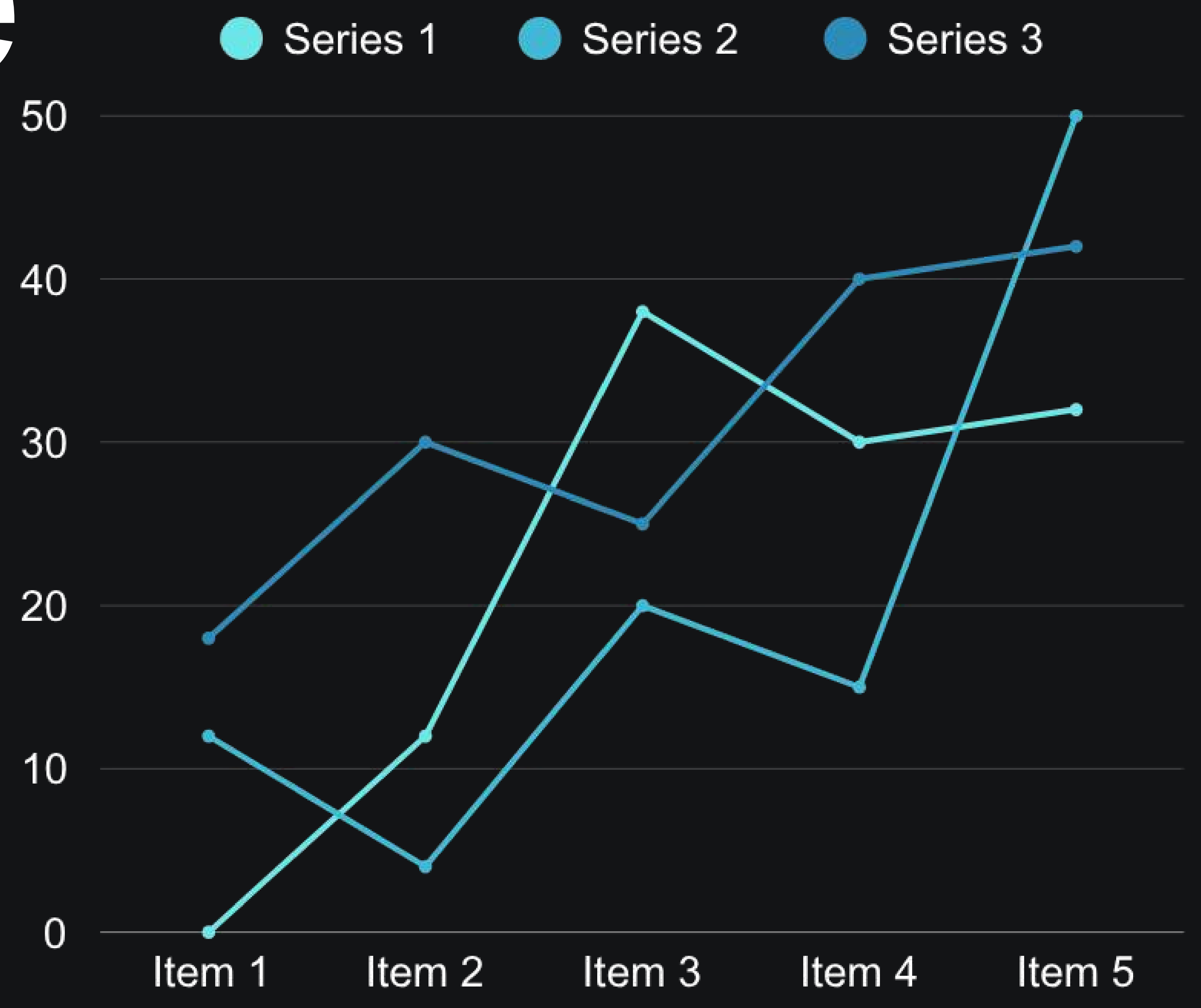
Concept: Learns the boundary that separates normal data from everything else (assumes most data is normal).

Autoencoder

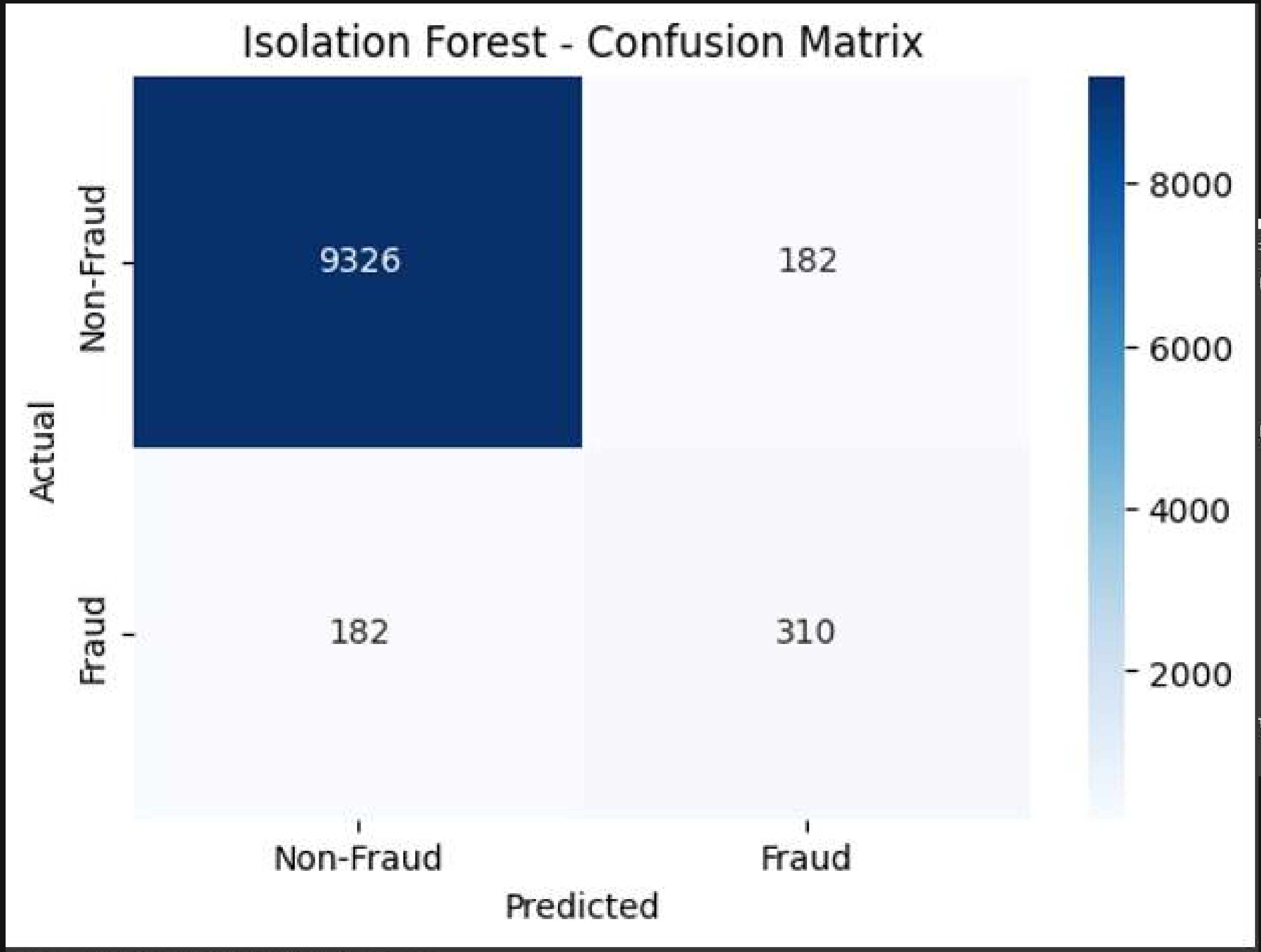
Concept: A type of neural network trained to reconstruct its input. Anomalies have high reconstruction error.

Performance Metrics

- Precision = TP / (TP + FP): Accuracy of positive predictions.
- Recall = TP / (TP + FN): Ability to find all frauds.
- F1-Score = 2 * (Precision * Recall) / (Precision + Recall): Harmonic mean
- AUC-ROC: Measures model's ability to distinguish between classes.



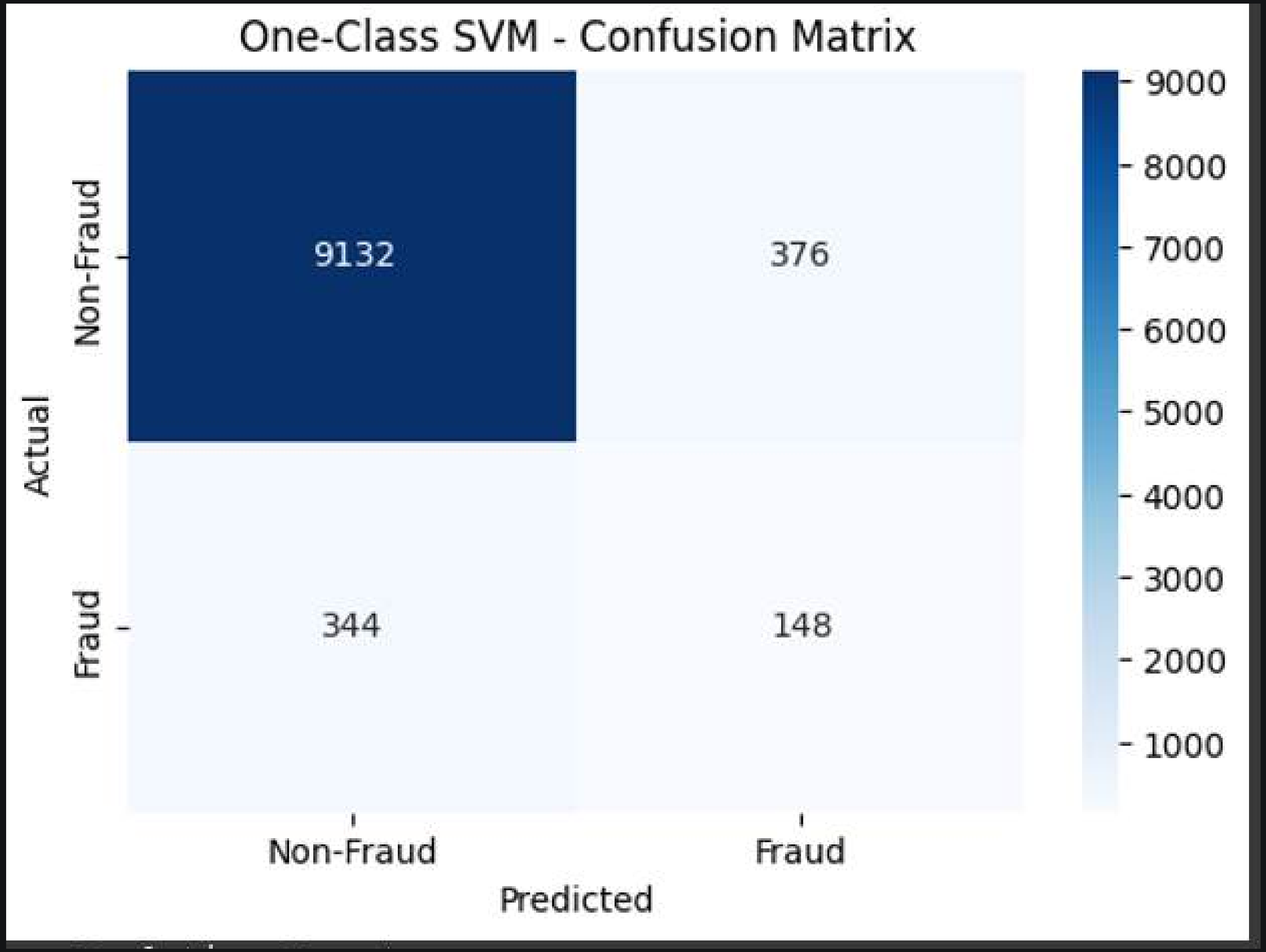
RESULT




```
=== One-Class SVM ===
Confusion Matrix:
 [[9132  376]
 [ 344  148]]

Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.96      0.96      9508
           1       0.28      0.30      0.29       492

    accuracy                           0.93     10000
   macro avg       0.62      0.63      0.63     10000
weighted avg       0.93      0.93      0.93     10000
```

RESULT

RESUL
T




```
=== Isolation Forest ===
Confusion Matrix:
 [[9326  182]
 [ 182  310]]

Classification Report:
              precision    recall  f1-score   support

           0       0.98      0.98      0.98      9508
           1       0.63      0.63      0.63       492

    accuracy                           0.96     10000
   macro avg       0.81      0.81      0.81     10000
weighted avg       0.96      0.96      0.96     10000
```

“Conclusion”

- Unsupervised learning is effective for fraud detection in imbalanced datasets.
- Models like Autoencoders and Isolation Forest can detect unknown fraud patterns.
- Results show strong potential for real-world use with proper tuning.
- Future improvements: hybrid models, real-time integration, more feature engineering.

References

- Kaggle: Credit Card Fraud Detection Dataset
- Scikit-learn documentation
- Keras & TensorFlow: Autoencoder examples
- Research papers on anomaly detection in finance

Thank You