Project 2
Subject: Fraud Detection in Financial Transactions

1. Problem Overview:

The task is to detect fraudulent credit card transactions using a binary classification model. The dataset comprises 154,621 samples and 31 features, including a Class column, where 0 represents non-fraudulent transactions and 1 represents fraudulent transactions.

2. Dataset Exploration:

Shape of Dataset: 154,621 rows and 31 columns.
Missing Values: The dataset contained 1 missing value in 29 columns. Rows with missing values were dropped.
Imbalance in Data: Fraudulent transactions (Class 1) are a minority, indicating a highly imbalanced dataset.

3. Preprocessing:

Feature Scaling: StandardScaler was used to normalize the features for better performance of the neural network.
Data Split:
Training Data: 70% of the dataset (stratified split to maintain class proportions).

Testing Data: 30% of the dataset.

4. Model Architecture:

Model: A sequential neural network implemented using TensorFlow and Keras.

Layers:

Dense layers with ReLU activation for non-linearity.

Dropout layers (30%) to reduce overfitting.

Output layer with a sigmoid activation function for binary classification.

Optimizer: Adam.

Loss Function: Binary cross-entropy.

Epochs: 20.

Batch Size: 64.

5. Training:

The model achieved an accuracy of 99.92% on the validation data by the 20th epoch.

Both training and validation loss decreased steadily, demonstrating a stable training process.

6. Results on Test Data:

Accuracy: 99.92%

Confusion Matrix:

True Negatives: 46,275

False Positives: 10

False Negatives: 26
True Positives: 75
Classification Metrics:
Precision (Fraudulent): 88%
Recall (Fraudulent): 74%
F1-Score (Fraudulent): 81%
ROC-AUC Score: 0.9741, indicating excellent performance.

7. Observations:

Class Imbalance Impact:
Precision and recall for Class 1 (fraudulent transactions) are lower compared to Class 0. This is expected due to the imbalance in the dataset.
False negatives (missed fraudulent transactions) can have serious implications in real-world applications.
Model Strengths:
High accuracy, precision, and recall for non-fraudulent transactions (Class 0).
Excellent ROC-AUC score, indicating strong separability.
Model Weaknesses:
Slightly reduced recall for Class 1. This indicates the model misses some fraudulent transactions, which could be mitigated by further techniques like SMOTE for data balancing.

8. Visualization:

Loss Curve:

Training and validation loss decreased over epochs, demonstrating effective learning without significant overfitting.
Validation loss stabilized after around 15 epochs.
ROC Curve:

AUC = 0.9741, showing the model's ability to differentiate between fraudulent and non-fraudulent transactions effectively.

9. Future Improvements:
Addressing Imbalance:
Techniques like oversampling (SMOTE) or undersampling could help improve recall for fraudulent transactions.
Class weighting during model training can also be explored.
Alternative Models: Investigate ensemble techniques like Random Forests or Gradient Boosting for comparison.
Hyperparameter Tuning: Use Grid Search or Bayesian Optimization to refine model parameters.
Feature Engineering: Explore additional derived features to enhance predictive performance.

10. Conclusion:
The neural network achieved exceptional performance with high accuracy and ROC-AUC score. However, additional

efforts are needed to improve recall for fraudulent transactions, ensuring fewer false negatives in production scenarios.