Maharshi Karve Stree Shikshan Samstha's
Cummins College of Engineering for Women, Pune
(An Autonomous Institute affiliated to Savitribai Phule Pune University)

23PCCE501L – Artificial Intelligence and Machine Learning Laboratory

TY BTECH COMP DIV A

Group 4

Transaction Fraud Detection Using Machine Learning Techniques

UCE2023453 : Janhavi Patki
UCE2023471 : Shriya Tijare
UCE2023473 : Esha Vasagadekar

01

Introduction

- Financial transactions have increased massively through online banking, UPI, e-commerce, wallets, etc.
- Along with this growth, fraudulent transactions have also increased.
- Manual rule-based systems often fail to detect evolving fraud patterns.
- Machine Learning helps identify fraud by learning patterns from historical data.
- This project uses Logistic Regression & Random Forest models to classify transactions as fraud or normal
- Our project adds an extra layer of intelligence: Location-based behavior analysis.

02

Objective

- Build a reliable ML system to detect fraudulent transactions.
- Train and compare Logistic Regression & Random Forest on a real transactions dataset.
- Preprocess data: scaling, handling imbalance, stratified split.
- Evaluate using Accuracy, Precision, Recall, F1-score, ROC-AUC.
- Build a Gradio-based user interface to test fraud risk in real-time.
- Analyze model behavior, outcomes, and challenges.
- Integrate user location + transaction location similarity

03

Problem Statement

To develop an efficient and scalable machine learning model capable of detecting fraudulent financial transactions using supervised learning techniques and deploy it with an interactive interface.

Challenges Addressed:

- Real-time decision requirements
- Evolving fraud patterns
- Need for accurate + interpretable predictions
- Existing datasets do not contain location information, limiting fraud analysis.
- Need a system that combines transaction patterns + location behavior.

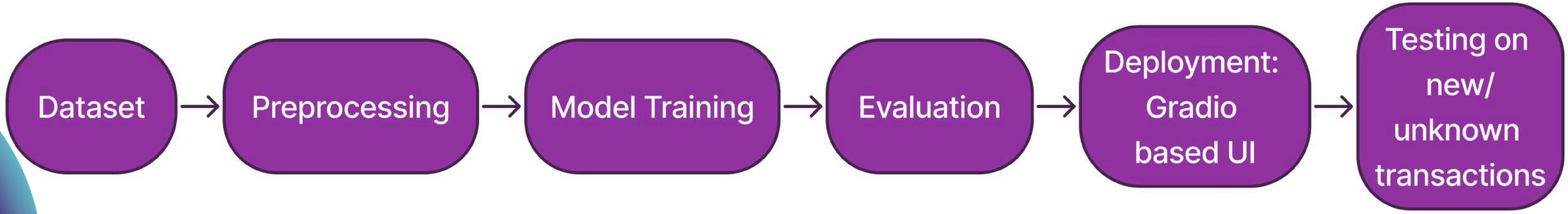
04

Market Analysis

- Digital payments in India grew over 50% yearly due to UPI and online banking.
- Fraud cases such as phishing, unauthorized transactions, OTP theft, and merchant scams are rising.
- RBI reports growing financial cybercrime cases every year.
- Businesses require:
    a. Real-time fraud detection
    b. Explainability
    c. Scalability
    d. High fraud recall (avoid missed frauds)

05

Methodology



Outcomes

- Models trained successfully on anonymized transaction dataset
- Random Forest performed best in terms of recall & AUC
- Generated Confusion Matrix, ROC Curve, Accuracy Table
- Created a working Gradio app for predicting fraud probability
- Created a donut chart showing contribution of the features

07

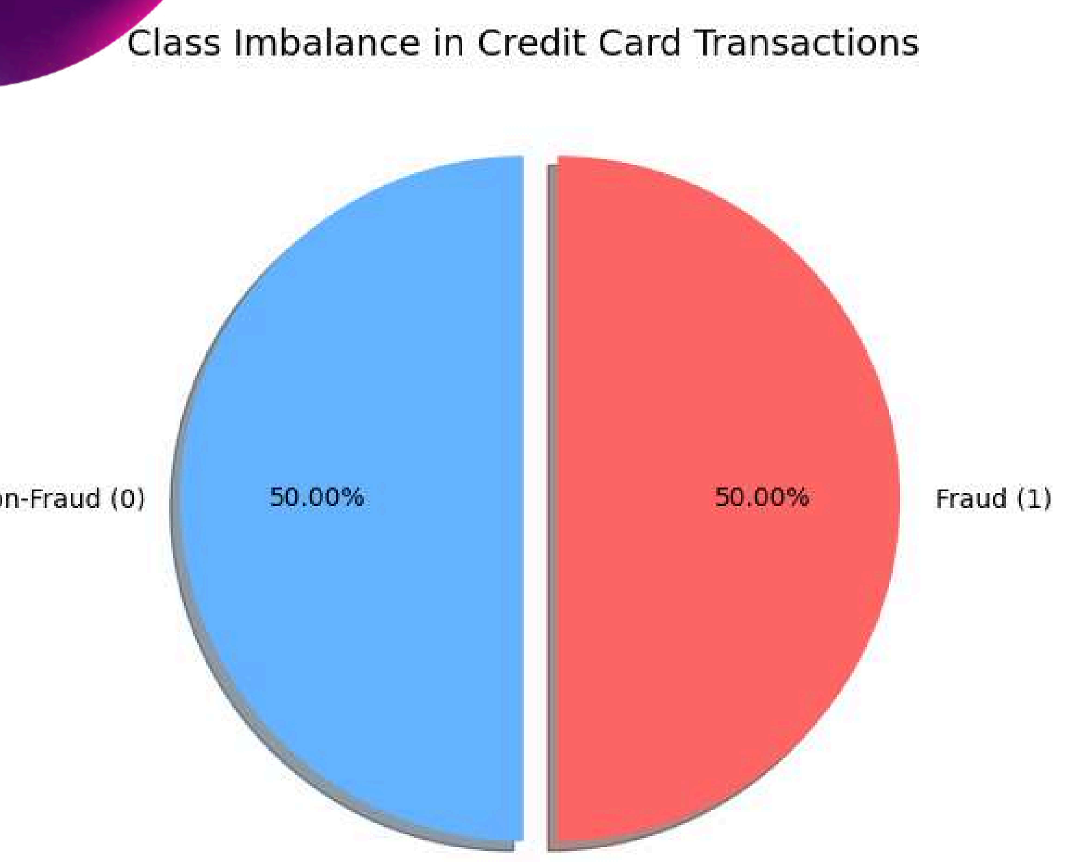
Results

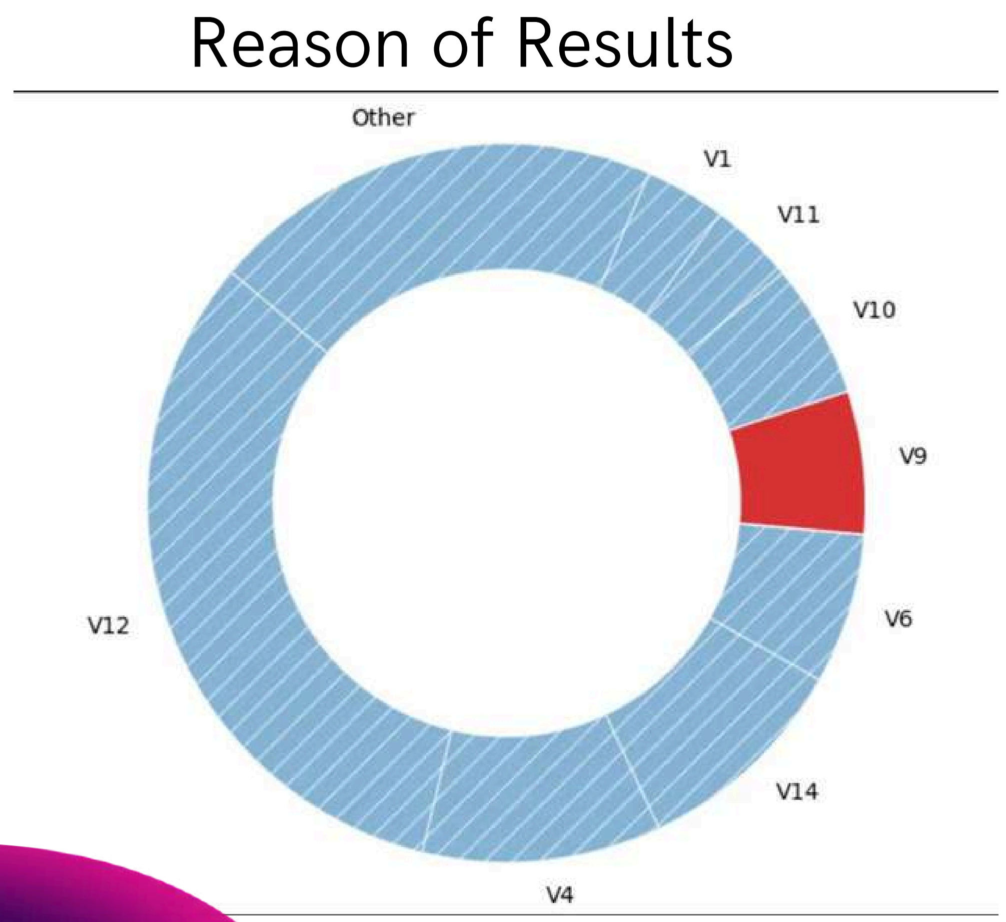
```
==============================================================
Model: Logistic Regression
==============================================================
              precision    recall  f1-score   support

           0     0.9976    0.9990    0.9983     56863
           1     0.9990    0.9976    0.9983     56863

    accuracy                         0.9983    113726
   macro avg     0.9983    0.9983    0.9983    113726
weighted avg     0.9983    0.9983    0.9983    113726

Confusion matrix:
  TN: 56,805   FP: 58
  FN: 137   TP: 56,726
```

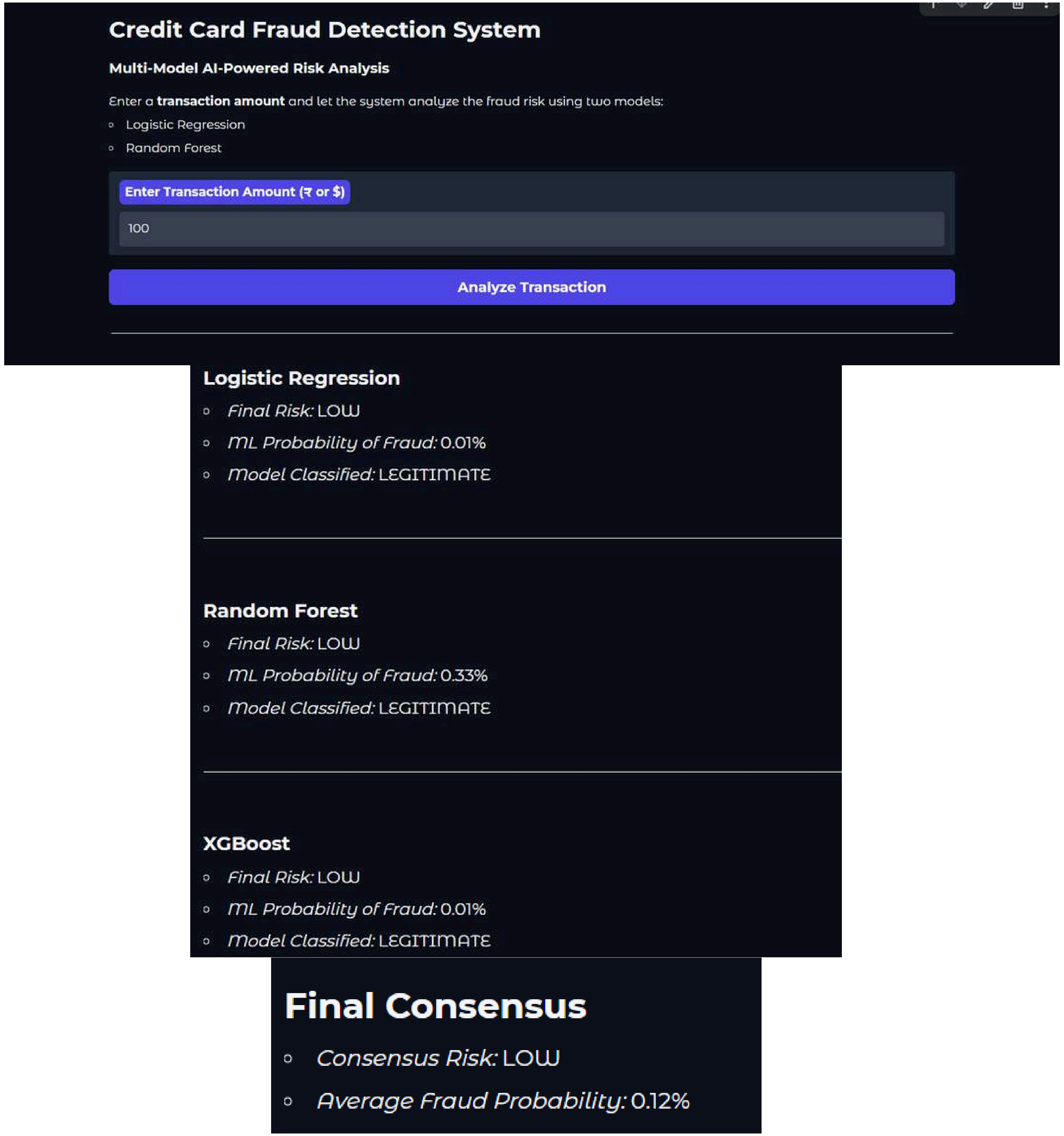
```
==============================================================
Model: Random Forest
==============================================================
              precision    recall  f1-score   support

           0     0.9997    0.9998    0.9998     56863
           1     0.9998    0.9997    0.9998     56863

    accuracy                         0.9998    113726
   macro avg     0.9998    0.9998    0.9998    113726
weighted avg     0.9998    0.9998    0.9998    113726

Confusion matrix:
  TN: 56,852   FP: 11
  FN: 16   TP: 56,847
```

```
==============================================================
Model: XGBoost
==============================================================
              precision    recall  f1-score   support

           0     0.9996    0.9999    0.9997     56863
           1     0.9999    0.9996    0.9997     56863

    accuracy                         0.9997    113726
   macro avg     0.9997    0.9997    0.9997    113726
weighted avg     0.9997    0.9997    0.9997    113726

Confusion matrix:
  TN: 56,855   FP: 8
  FN: 23   TP: 56,840

==============================================================
```

```
Overall comparison:
              Model  Accuracy  AUC-ROC  True Positives  False Negatives  False Positives
Logistic Regression  0.998285 0.999807           56726              137               58
      Random Forest  0.999763 0.999996           56847               16               11
            XGBoost  0.999727 0.999977           56840               23                8

Best model by AUC-ROC: Random Forest (AUC: 1.0000)
```


Class Imbalance in Credit Card Transactions
n-Fraud (0)
50.00%
50.00%
Fraud (1)


Reason of Results
Other
V1
V11
V10
V9
V6
V14
V4
V12


Credit Card Fraud Detection System
Multi-Model AI-Powered Risk Analysis
Enter a transaction amount and let the system analyze the fraud risk using two models:
Logistic Regression
Random Forest
Enter Transaction Amount (₹ or $)
100
Analyze Transaction
Logistic Regression
Final Risk: LOW
ML Probability of Fraud: 0.01%
Model Classified: LEGITIMATE
Random Forest
Final Risk: LOW
ML Probability of Fraud: 0.33%
Model Classified: LEGITIMATE
XGBoost
Final Risk: LOW
ML Probability of Fraud: 0.01%
Model Classified: LEGITIMATE
Final Consensus
Consensus Risk: LOW
Average Fraud Probability: 0.12%

08

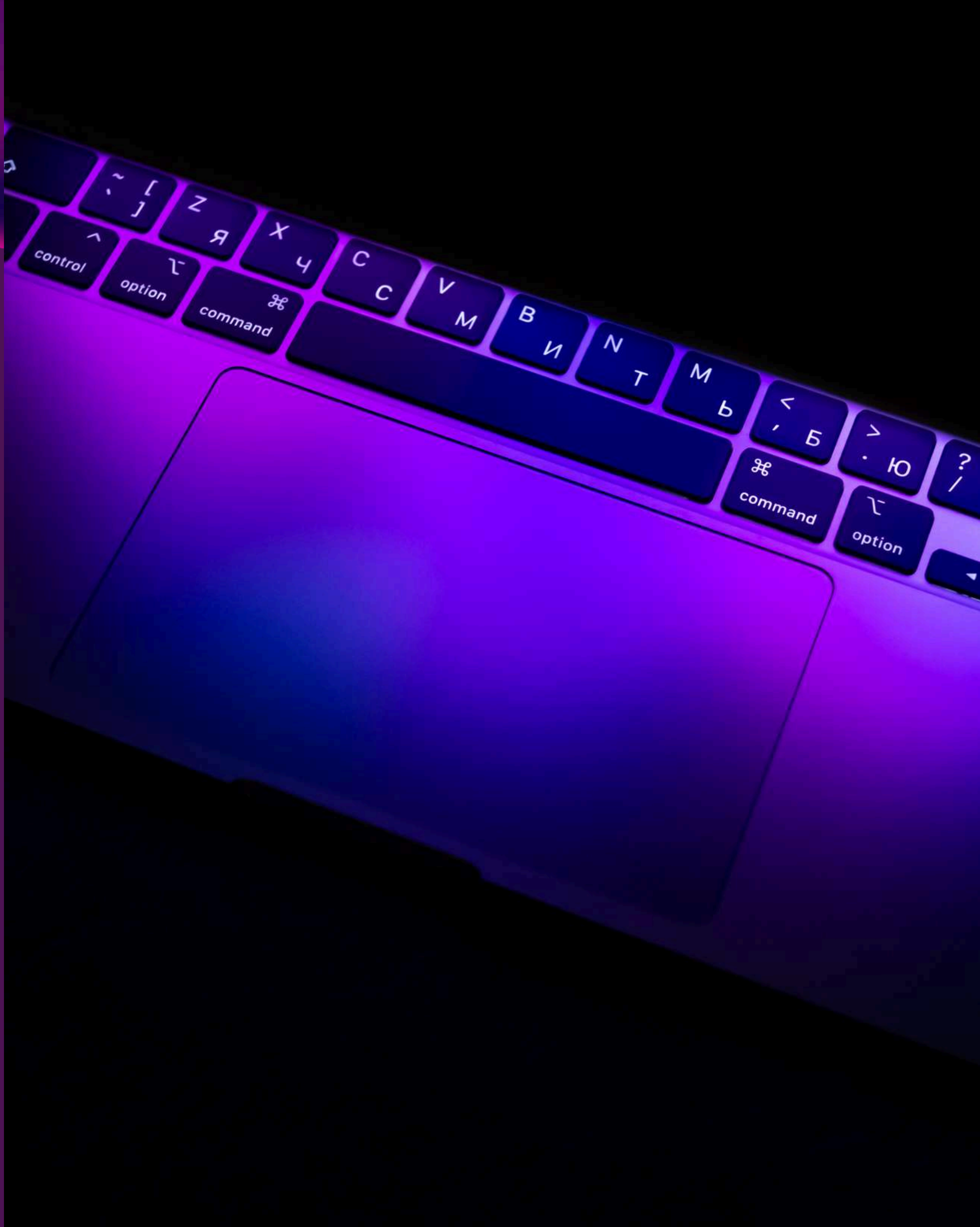
Conclusion

- Developed an end-to-end Transaction Fraud Detection system.
- Compared LR and RF and XGBoost .
- Achieved extremely high accuracy due to clean anonymized dataset.
- Demonstrated real-time prediction using Gradio.
- System can be extended for banks, wallets, e-commerce, and merchant fraud protection.

09

Future Scope

- Add SHAP explain-ability
- Deploy model as API or web service
- Integrate deep learning or anomaly detection
- Add real-time streaming data
- User behavior profiling
- Adaptive retraining pipeline



Acknowledgement

We sincerely thank:

- Our project guide :
  - Varsha Pimprale
  - Dr. Sandhya Arora
- Open-source contributors of datasets/tools
- Kaggle for providing dataset access

References

- Gradio documentation
- Kaggle: Fraud Transaction Dataset
- West, Bhattacharya. “Intelligent financial fraud detection”

THANK YOU