Bank Account Fraud Detection

Problem Statement 1

-TuringTitans

Overview

- Objective
- Proposed Approach
- Flow of the solution
- Result
- Tech Stack
- Future Scope

Objectives

Develop an effective fraud detection system for banking transactions

Objective 01

Performing EDA to identify the potential features, and understand the characteristics of fraudulent and legitimate transactions.

Objective 02

Data preprocessing,performing feature engineering to optimize the dataset for model training.

Objective 03

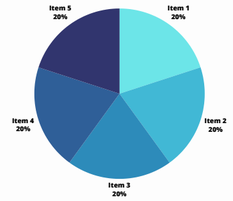
Develop a ML model for classification of transactions,considering the imbalanced class distribution.

Objectives

Develop an effective fraud detection system for banking transactions

Objective 04

Evaluate the performance of the model using appropriate metrics.



Objective 05

Provide insights and recommendations for integrating the it to the bank’s existing processes.

Proposed Approach

- Preprocessing of data
- Exploratory Data Analysis
- Selecting a resampling techinque: SMOTE(Synthetic Minority Over-sampling Technique) for oversampling
- Feature Engineering using Anova test(F test and P-value)
- 18 features are selected

- Model Selection:
    a)Decision tree
    b) Random Forest
    c)LightGBM
    d)XGBoost
    e)LSTM
- We got best results for XGBoost model with F1-Score and Accuracy of 97%
- Explainability of Fraud Detection

Flow of the solution

01

Cleaning and Prepossing of data

02

Performing EDA , feature Engineering and handling missing values

03

Selecting models for fraud Detection:

1. Decision Tree
2. Random Forest
3. XGBoost
4. LightGBM
5. LSTM

05

Performance evaluation metrics
Accuracy
precision
Recall
F1 score
AUC curve
ROC curve

06

Selecting the best model.
We have achieved accuracy of 97%

07

Explainability of Fraud Detection

Results

01

XGBoost

Precision: 0.9802789655874568

Recall: 0.9759318855636926

F1 Score: 0.9781005955547993

Accuracy: 0.9781942965808313

AUC-ROC Score: 0.9781896320395 13

Results

Comparison of Models

Model	Precision	Recall	F1 Score
XGBoost	0.876002	0.787387	0.829334
LightGBM	0.978374	0.970522	0.974432
Random Forest	0.905961	0.930646	0.918138
Decision Tree	0.876002	0.787387	0.829334

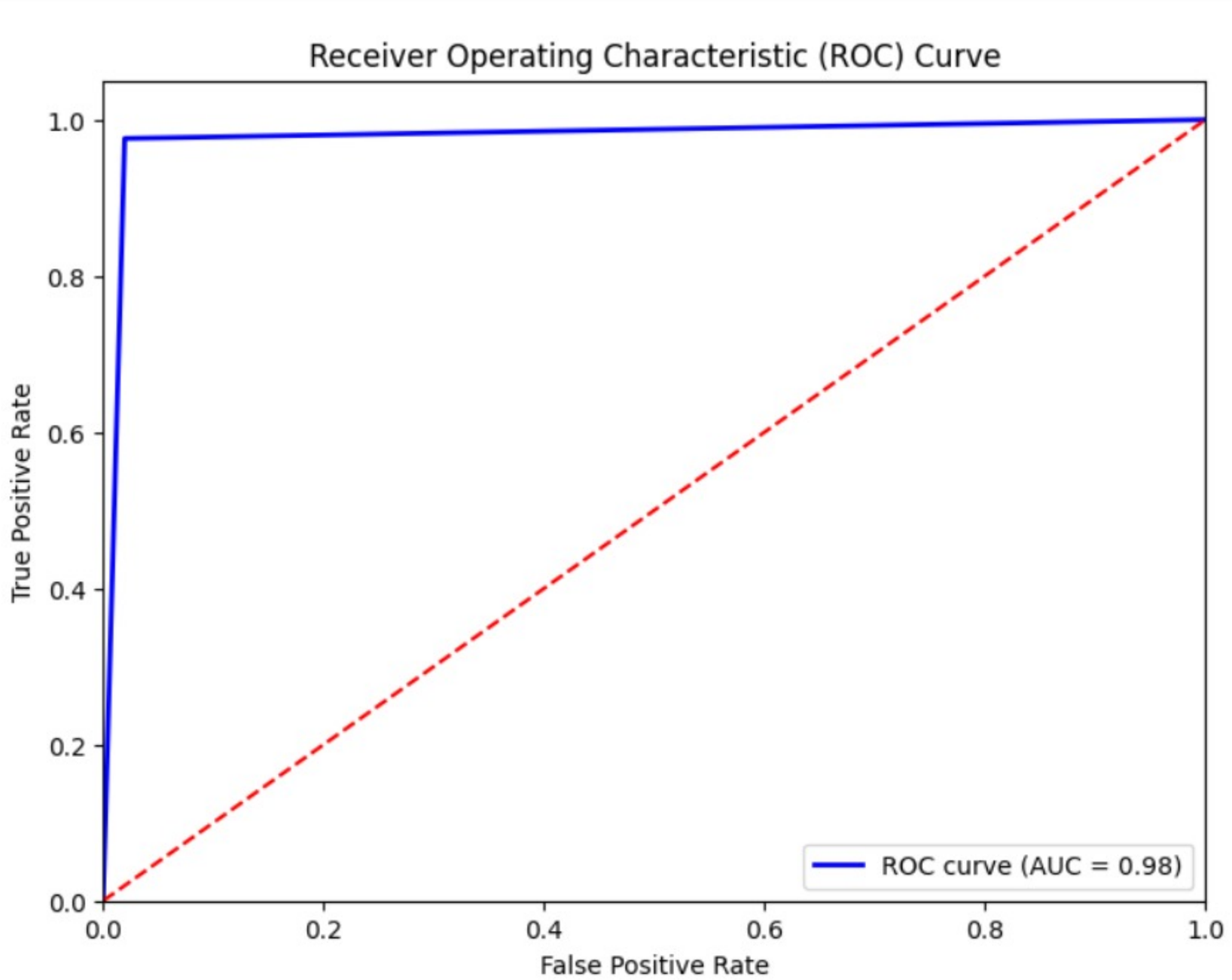
Results

Accuracies of the models with changing number of features

No of features	Decision tree	Random forest	LightBGM	XGBoost
10	80	81	93	93.9
15	82	89	96	96.7
18	86	91	97	97.7
20	86	92	97	97

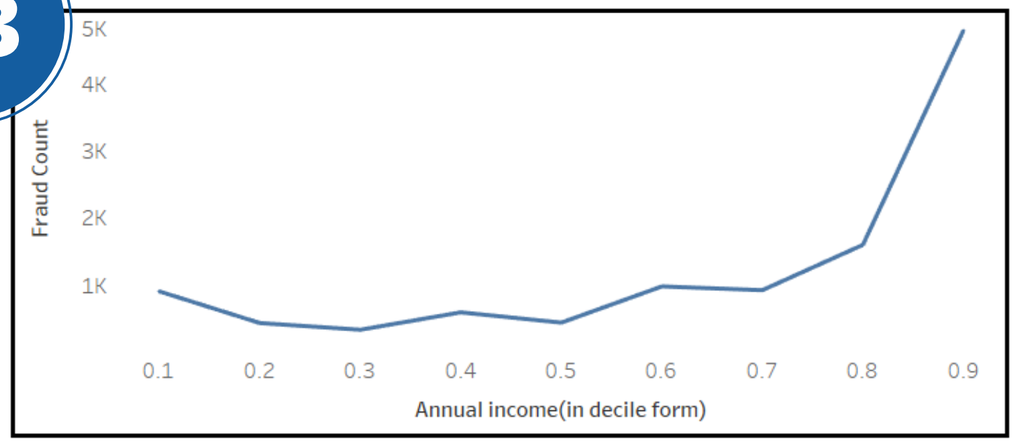
02

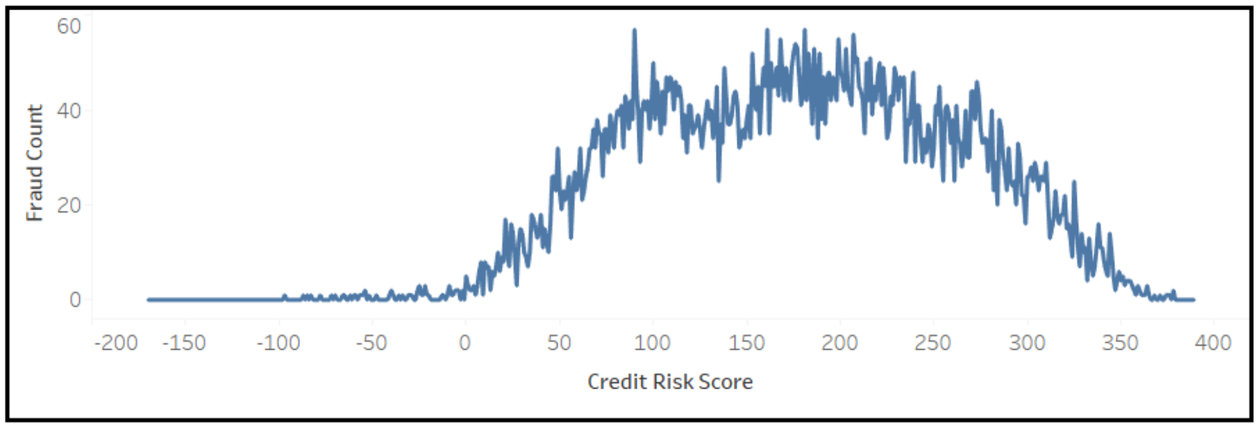
Results

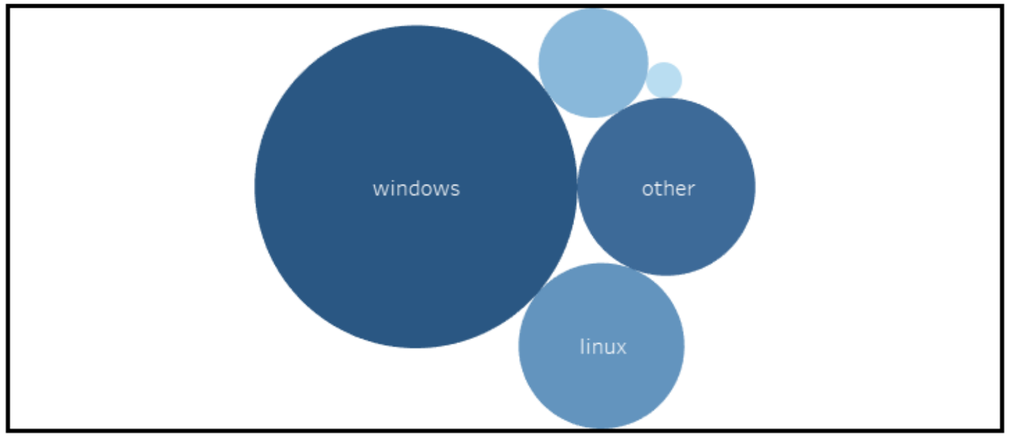

Receiver Operating Characteristic (ROC) Curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (AUC = 0.98)

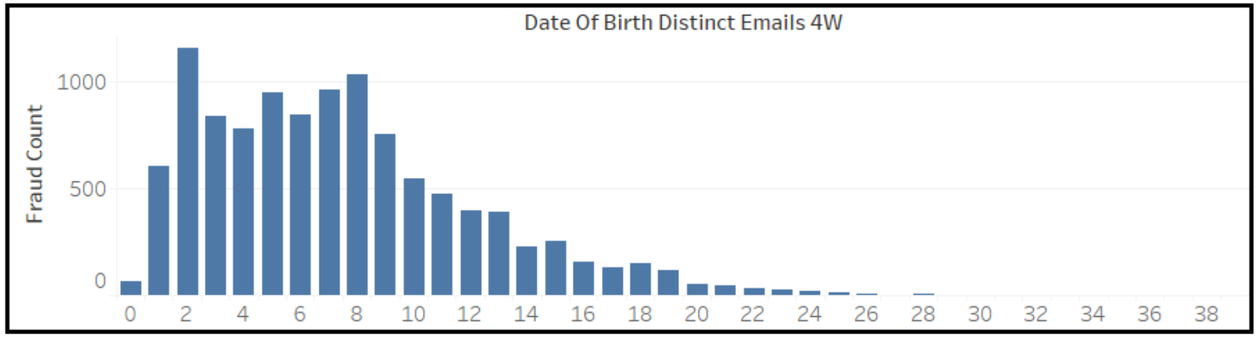
03

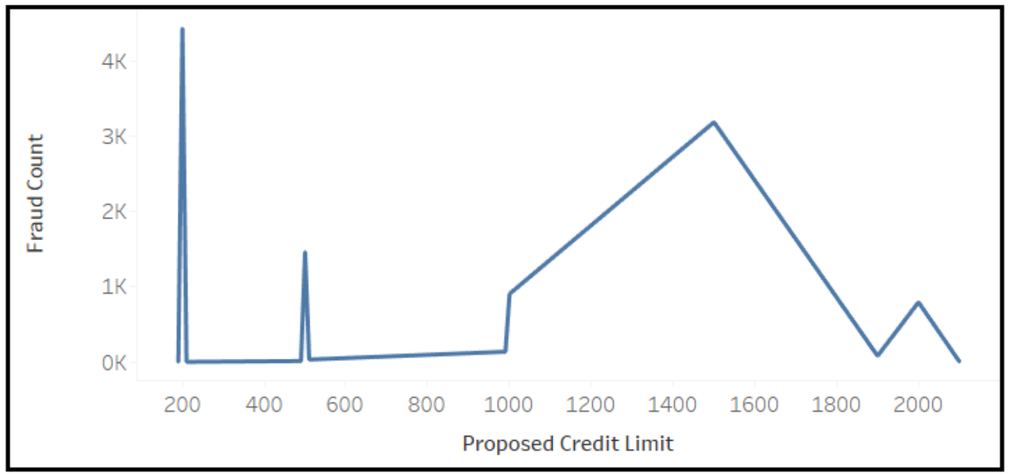
Feature Analysis

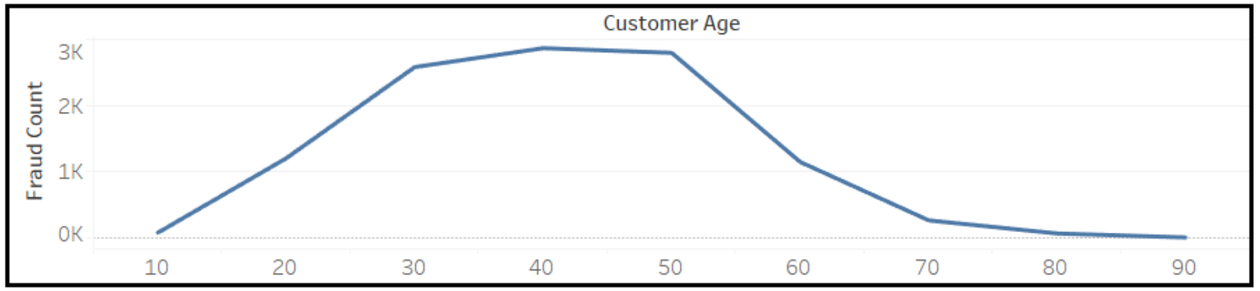


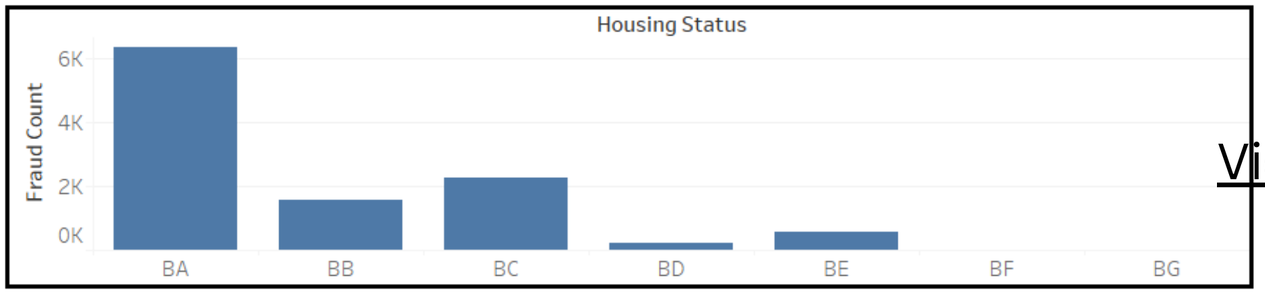












Viewlink

Tech Stack

- Python
- Jupyter Notebook
- Visual Studio Code
- Tableau
- libraries used :

pandas, numpy, matplotlib ,seaborn, sklearn , imblearn, lightgbm , xgboost .

Future Scope

- Online Learning: Implement online learning algorithms that can continuously update the model's parameters as new data becomes available. This allows the model to adapt to changes in fraud patterns over time without requiring retraining from scratch.
- Temporal Analysis: Identify time-based patterns in fraudulent activity. Detect spikes or recurring trends in fraud at specific times.
- Privacy-Preserving Techniques: Protect sensitive customer data while still detecting fraud. Ensure compliance with privacy regulations.
- Behavioral Analysis: Analyze customer behavior to spot unusual patterns. Looks for changes in spending habits or transaction times.

THANK YOU!

Tanmayee Kulkarni
Aarushi Chopkar
Manasi Kole
Aakanksha Dorage