Credit Card Fraud Detection

Submitted by:

Khushi Somaiya

Agenda

- Problem Statement
- Objective
- Key Insights
- Cost Benefit Analysis
- Appendix

Problem Statement

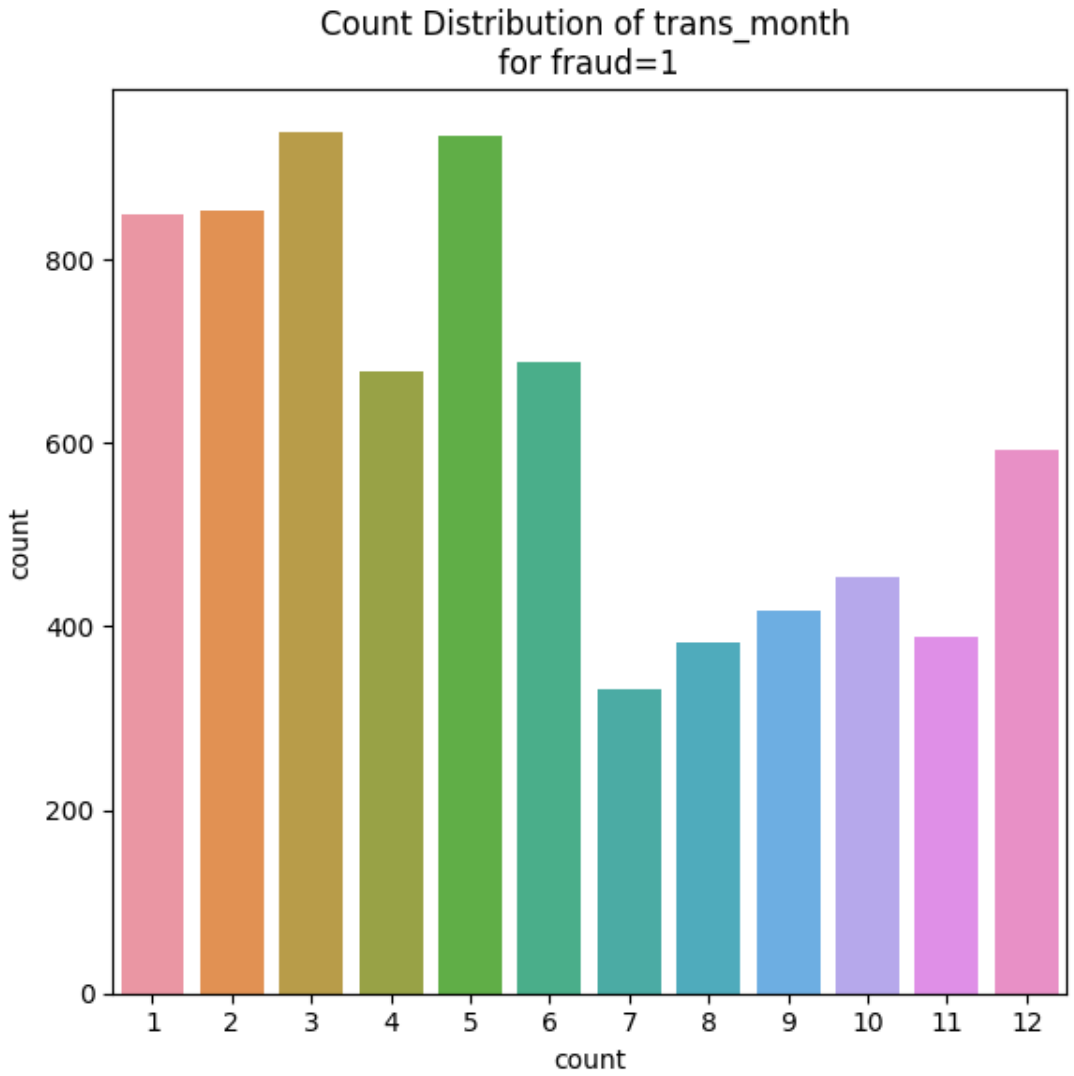
- Finex, a leading financial service provider, is experiencing a significant rise in unauthorized credit card transactions. Fraudsters exploit various methods, causing substantial financial losses to the bank.

- Finex lacks the necessary technology to detect and prevent fraudulent activities promptly, resulting in delayed customer complaints and financial losses.

- Customers often discover unauthorized transactions too late, leading to delayed complaint registration and the bank incurring substantial losses.

Objective

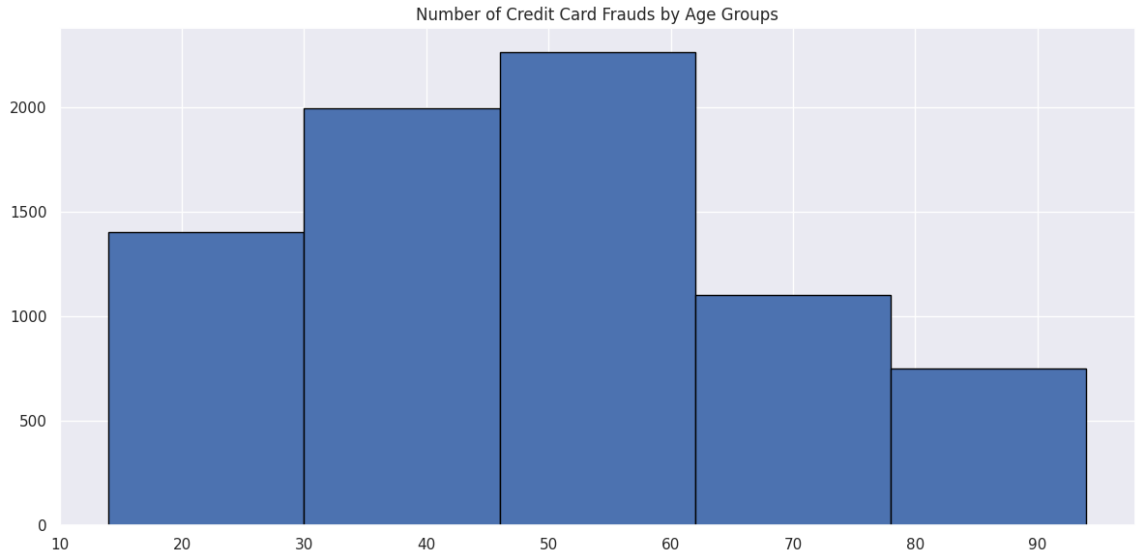
- Develop a machine learning model for credit card fraud detection, utilizing historical transaction data to identify and prevent unauthorized transactions proactively.
- Analyze the financial impact of fraudulent transactions, including chargebacks and customer reimbursements, to assess the potential cost savings of implementing the fraud detection system.



Key Insights - I

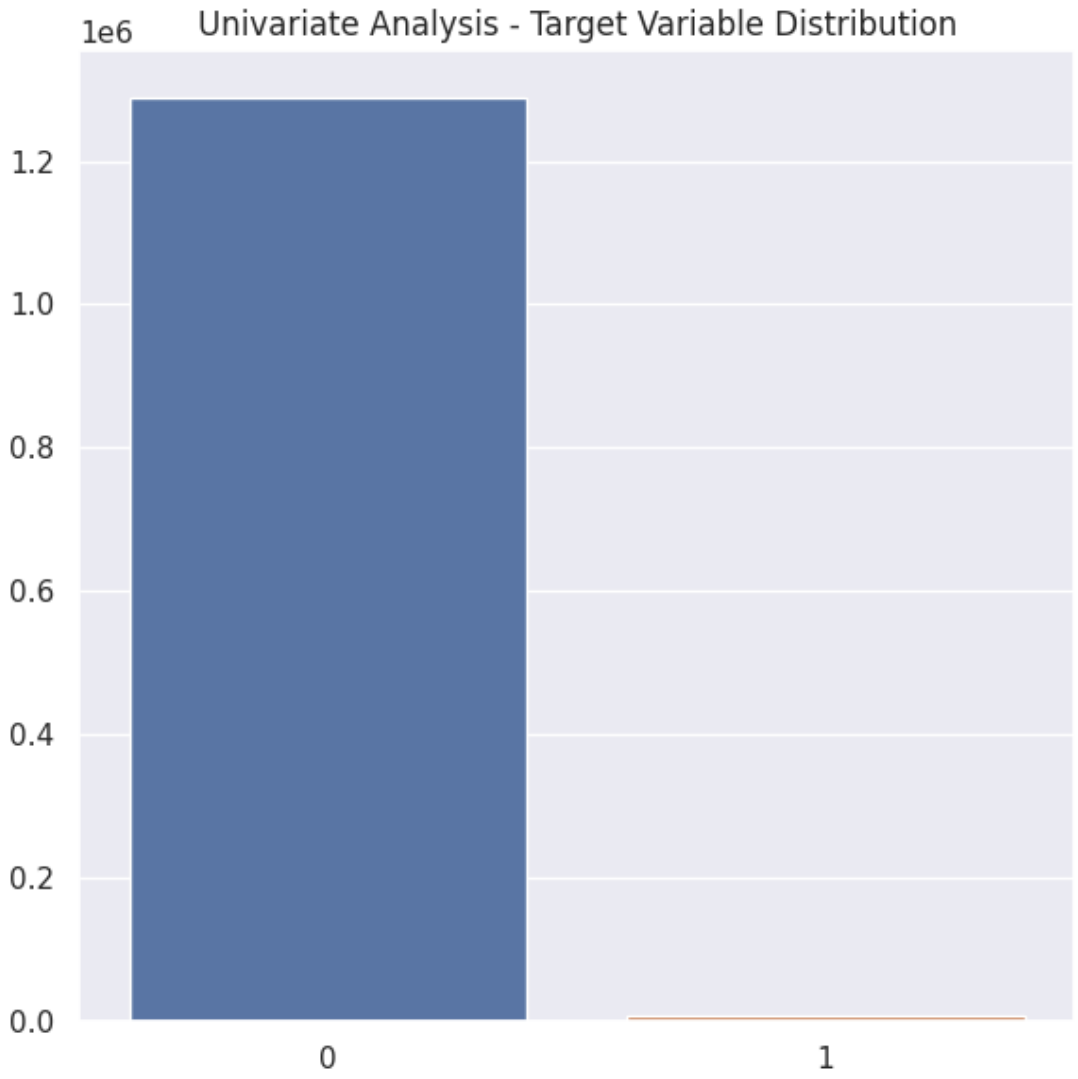
The count distribution of transaction months for fraudulent transactions reveals that these fraudulent activities are most prevalent in the months of May and March.

To effectively combat credit card fraud, Finex should focus on enhancing its security measures and fraud detection techniques during these peak months, in addition to implementing a year-round strategy.



Key Insights - II

The histogram chart representing fraudulent transactions across various age groups shows that the highest number of fraudulent transactions occur within the age range of 30 to 60. This age group seems to be particularly susceptible to credit card fraud.



Key Insights - III

The target variable "is_fraud" displays a significant imbalance, with fewer fraudulent transactions compared to legitimate ones.

Techniques like SMOTE and ADASYN can be used to address this imbalance by creating synthetic data points for the minority class.

Cost Benefit Analysis Part - 1

Finex processes an average of 77,183 transactions each month.

There are approximately 402 fraudulent transactions reported on average each month.

The average amount for a fraudulent transaction is approximately $530.

Cost Benefit Analysis Part - 2

Prior to deploying the fraud detection model, Finex incurred a cost of $213,060 per month.

After deploying the model, the cost of providing support for detected fraud transactions totals $358.5.

The cost associated with undetected transactions is $34,980 per month.

After deploying the model, the total monthly cost is $35,338.5.

The implementation of the fraud detection model results in cost savings of $177,721 per month, as the cost incurred after model deployment is significantly lower than the cost incurred before deployment.

Appendix: Project Pipeline

Understanding Data

Exploratory Data Analytics

Train/Test Data Splitting

Model Building or Hyperparameter tuning

Model Evaluation

Business Impact

Model	Recall on Train	Recall on Test	AUC Score
XGBoost (without sampling method)	0.70	0.49	0.98
Random Forest (without sampling method)	0.26	0.22	0.96
XGBoost - SMOTE	0.86	0.80	0.95
XGBoost - ADASYN	0.86	0.80	0.95
Decision Trees - SMOTE	0.86	0.81	0.93
Decision Trees - ADASYN	0.85	0.82	0.93
Random Forest - SMOTE	0.82	0.78	0.93
Random Forest - ADASYN	0.81	0.79	0.93
Decision Trees (without sampling method)	0.30	0.28	0.73
Logistic Regression - ADASYN	0.73	1.00	0.61
Logistic Regression - SMOTE	0.77	1.00	0.56
Logistic Regression (without sampling method)	0.00	0.99	0.52

Appendix: ML Model Summary

- The table provides an overview of various machine learning models with their respective performance metrics when applied to the task of fraud detection.
- Since it's more critical to accurately identify fraudulent transactions, recall is used as the primary evaluation metric.
- The Decision Tree model with ADASYN sampling not only demonstrates high recall (sensitivity) in detecting frauds but also achieves a strong AUC score, making it a robust choice for fraud detection in this context.

Appendix: Video Link

https://drive.google.com/file/d/1zchlNM84LTQJW0Unaz-ilslPu1DVuSHk/view?usp=share_link

Thank You