Credit Card Fraud detection

Submitted by-
Sawal Malhotra

Introduction

- Finex is a leading financial service provider in Florida that offers range of products and financial services.
- Banks are facing huge revenue and profit loss due to unauthorised fraudulent transactions in past years.
- Banks have received complaints from Customers for unauthorised transaction on their debit/credit card.
- Fraudsters use Stolen and lost card and hack private system to access personal and confidential information.
- Unethically, they are also involved in ATM Skimming at POS terminals such as gas stations, shopping malls and ATMs that do not send alerts.
- Most fraudulent activities are reported to happen during odd hours of the day leaving no suspicion of the malicious action.

Problem Statement

- Finex being not equipped with latest technologies, wants to identify the possible root cause of such fraudulent activity.

- Due to increase in Digital payment channels and other factors, its becoming difficult for financial firms to identify and predict fraudulent transaction. By the time, they are aware huge losses have already been made from Customers account which Banks are obliged to pay.

- They want a long term solution that would help bank generate revenue in future and have better customer experience.

Credit card customer life cycle

- **Onboarding the customers:** Customers are identified by marketing teams and applications are approved or declined as part of the process.
- **Issuing account details:** Once application is approved, physical cards or credentials are shared with accountholder.
- **Customer management:** It involves grouping and segmenting the customers based on their transaction type, etc.
- **Account maintenance:** It involves updating customer personal details based on their request like address , contact details, etc.
- **Closure/recovery:** Collecting or recovering due payments in case customer is opting out of bank services or refuses to make a payment.

Types of credit card fraud

- **Application fraud during onboarding:** When identity of genuine customers is used to apply for new account by criminals.

- **Mail theft while issuing account details:** Criminals tie up with Delivery companies to stole the confidential details.

- **Transaction fraud during Customer management:** It occurs due to physical or digital card details being stolen. It is one of the most common frauds that happens.

- **Identity takeover during account maintenance:** Customer information is gathered by criminals and later they contact Banks posing as a genuine customer.

How to identify a fraud

There can be different patterns of Fraud transaction :

- **Anomalous amount**: Making a higher amount of transaction all of a sudden i.e. around 4 to 5 times more than the average spent amount.

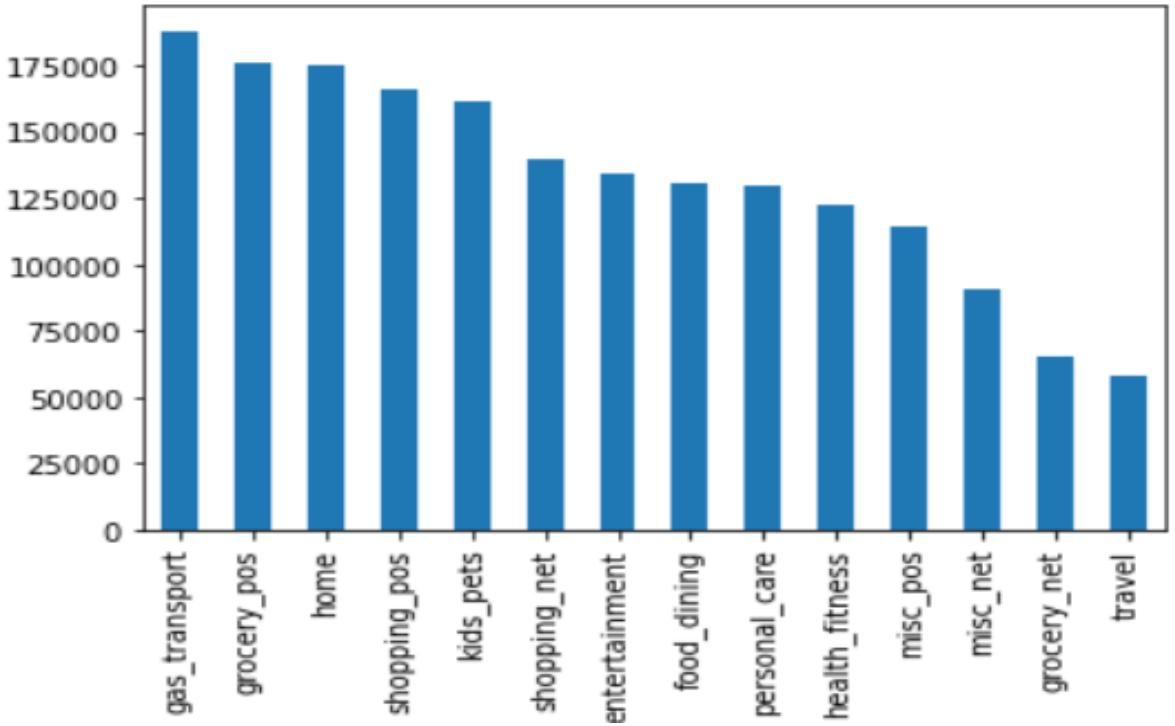
- **Anomalous time:** Making transactions during the Odd hours of the day, can be a pattern for transaction to be fraud.

- **Anomalous location:** Based on demography location, if card transaction is made all of sudden from other country or in some other currency. It can be a sign for online fraud transaction.

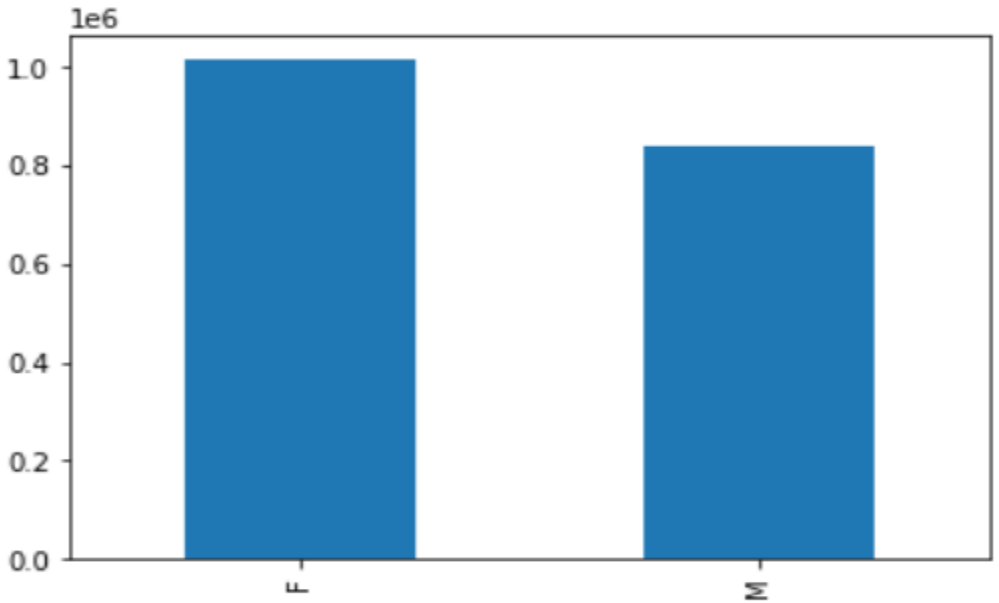
- **Transaction frequency:** Fraudsters can make transactions until he’s been caught by the Bank or identified(Repeater Fraud)

Steps performed for Fraud detection

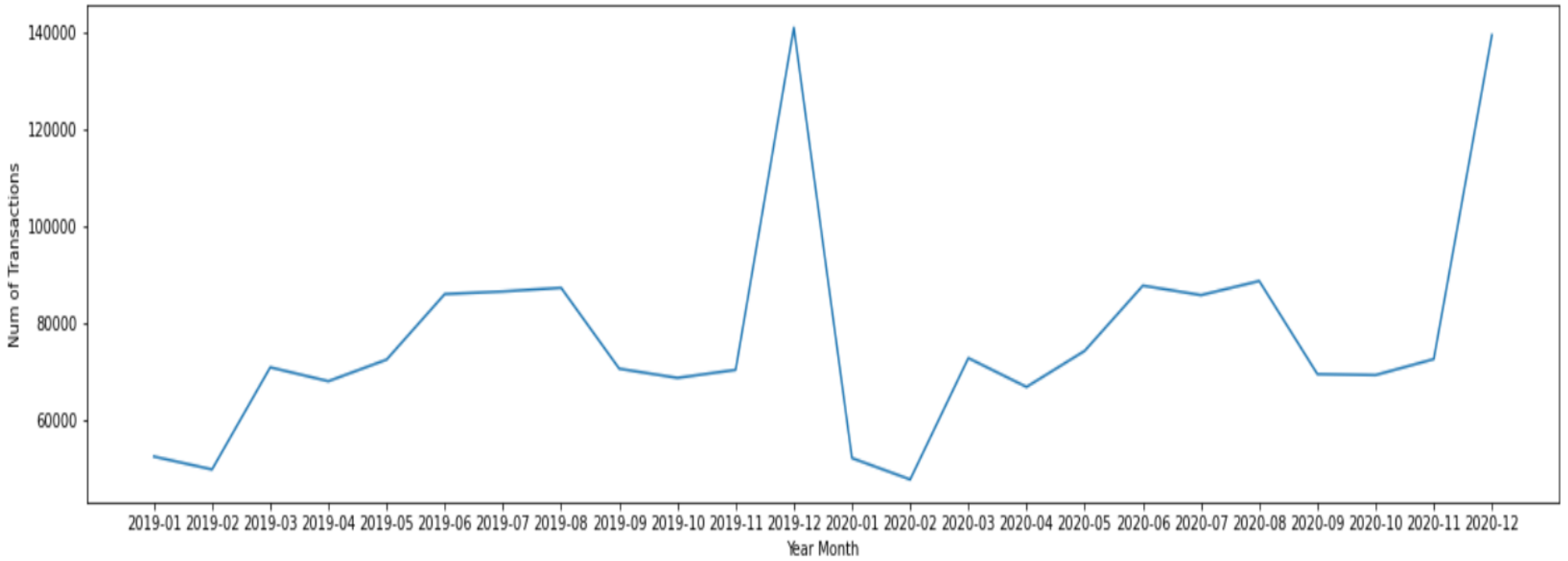
- Data understanding and data cleaning including conversion of columns data to appropriate format.
- Exploratory Data Analysis(EDA) with Univariate and Bivariate analysis to identify distribution and relation of different variables.
- Train and Test data splitting for training and testing the model to check its performance for our problem statement.
- Model building and model evaluation using Confusion matrix and Classification report metrics.



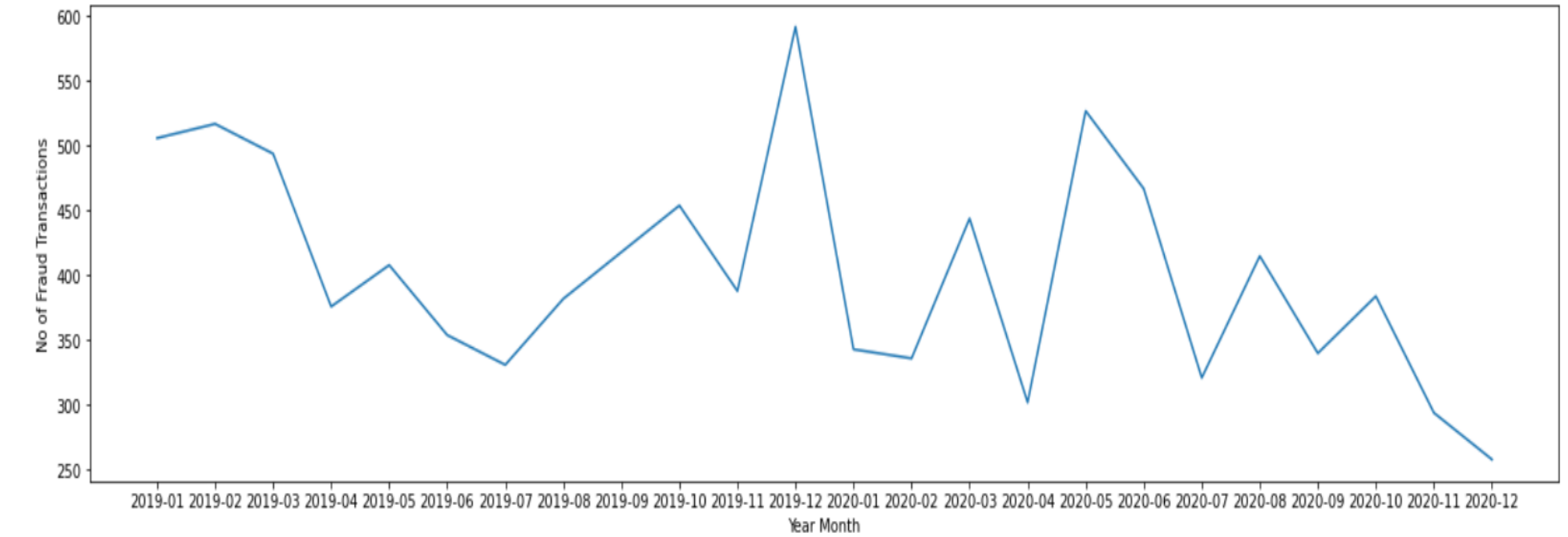
- Highest number of transactions were made at Gas_transport terminal
- Least number of transactions were made for travel purpose.



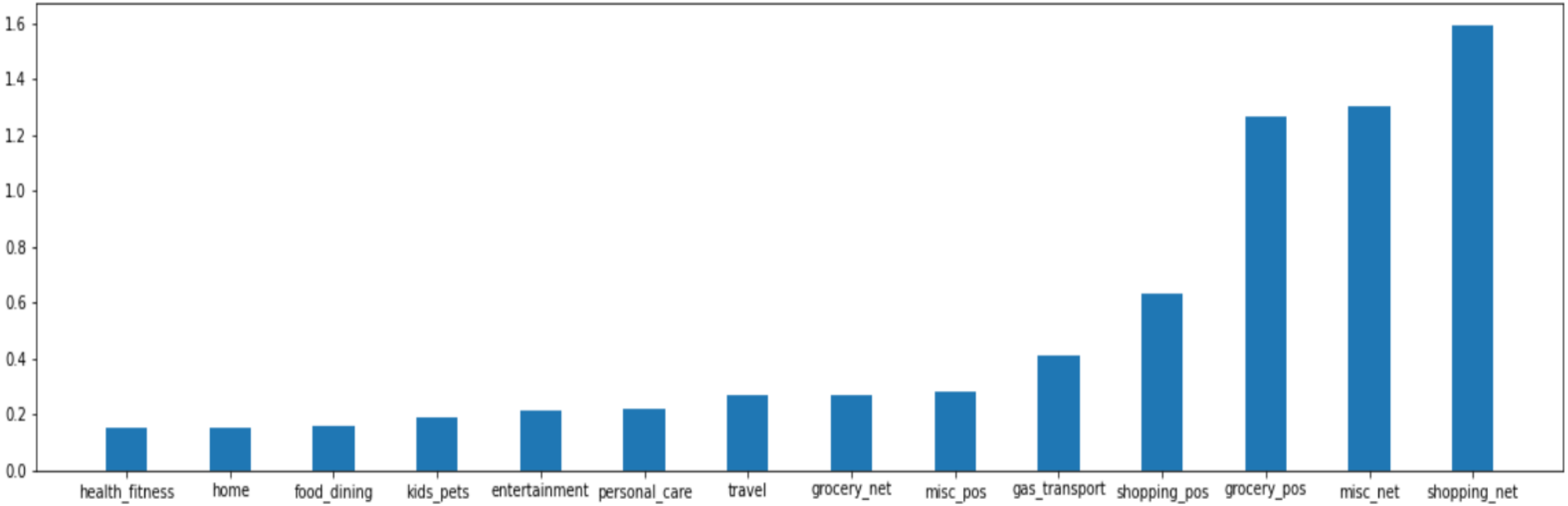
- Females have shown to make more Online card transactions as compare to Males.
- It can be said that Females tend to shop more using Debit/credit cards.



- Maximum number of overall transactions were recorded in December month, may be due to the Christmas Celebration in US.
- Least number of transaction were observed during start of new year in February month.



- Highest number of **Fraud** transactions were recorded in December, 2019.
- Least occurrence of fraud transactions were in December, 2020. Indicating, reduction in fraudulent transaction.



- 'Shopping_net' category is having the maximum count of Fraud transaction, while the 'health_fitness' category have recorded least no of fraud transactions.

- Top 3 categories whihc company should focus to readuced fraudlent transactions are 'shopping_net', 'misc_net' & 'grocery_pos'.

Building the real time model

- There was high Class imbalance in data set , wit less than 1% of fraud transaction sample data.

is_fraud	count	percentage
0	1842743	99.478999
1	9651	0.521001

- To mitigate, oversampling technique was used to balance Class dataset. On the other hand , these sampling techniques were not used as they were inefficient like- Under sampling (to reduce no of data points from majority class) and Smote(to increase data point in minority class with new feature) which

- Data was split into Train and Test data sets in ratio of 70:30.Model would first learn on training dataset and later performance of model would be checked on the Test model.

- Being a classification business problem, various models haven implemented such as Logistic regression, Decision Tree, Random forest and Gradient Boosting model.

- Models have been tested on both Training and test dataset to check for their accuracy and recall.

- Out of all , Decision tree was found to be the Best model for credit card fraud detection solution, as it had the highest recall metric value.

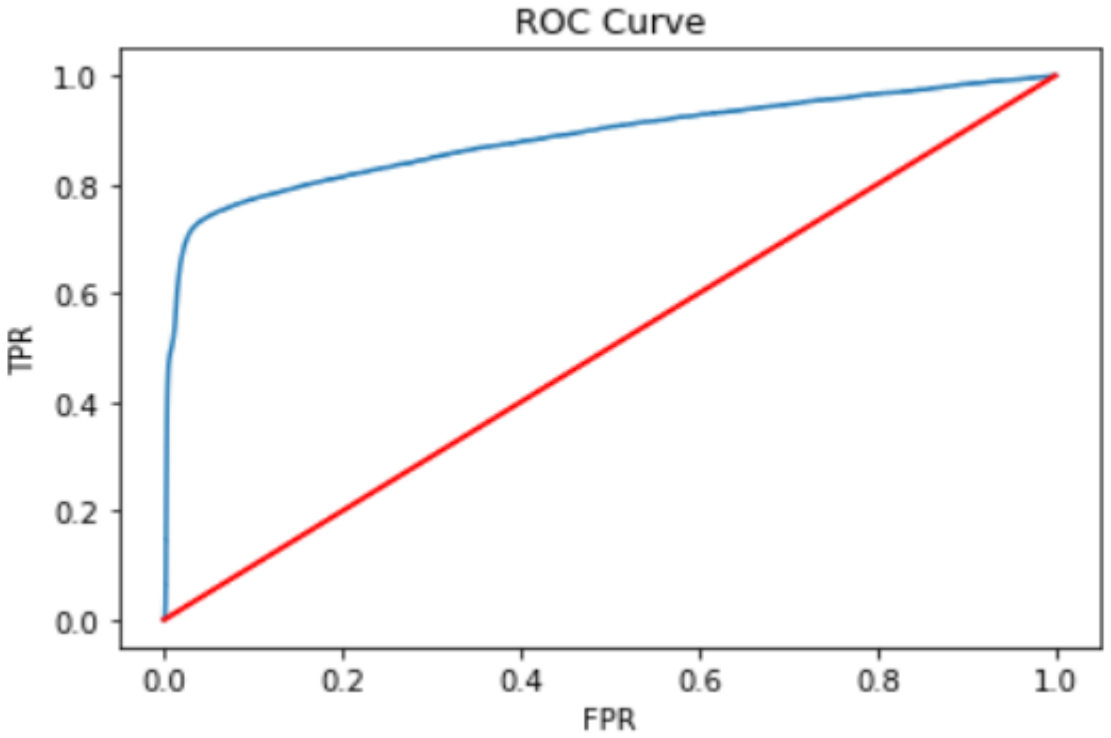
Logistic Regression model result

	precision	recall	f1-score	support
0	0.79	0.92	0.85	1290008
1	0.91	0.76	0.83	1289955
accuracy			0.84	2579963
macro avg	0.85	0.84	0.84	2579963
weighted avg	0.85	0.84	0.84	2579963

- For Train Data, model is having Accuracy around 84-85% and Recall of 76%.

	precision	recall	f1-score	support
0	0.79	0.92	0.85	552823
1	0.91	0.76	0.83	552876
accuracy			0.84	1105699
macro avg	0.85	0.84	0.84	1105699
weighted avg	0.85	0.84	0.84	1105699

- For Test Data, model is having Accuracy around 84-85% and Recall of 76%.



- ROC (receiver operating characteristic) curve tell us the performance of classification model at all classification thresholds.
- TPR(True Positive Rate) is around 80% which indicates the number of Fraud Transactions correctly predicted by the model.
- TPR needs to be as high as possible for a good model.

- Overall, LR model is good with high accuracy and recall. But, Recall value can be further improved and hyper tuned to increase the performance.

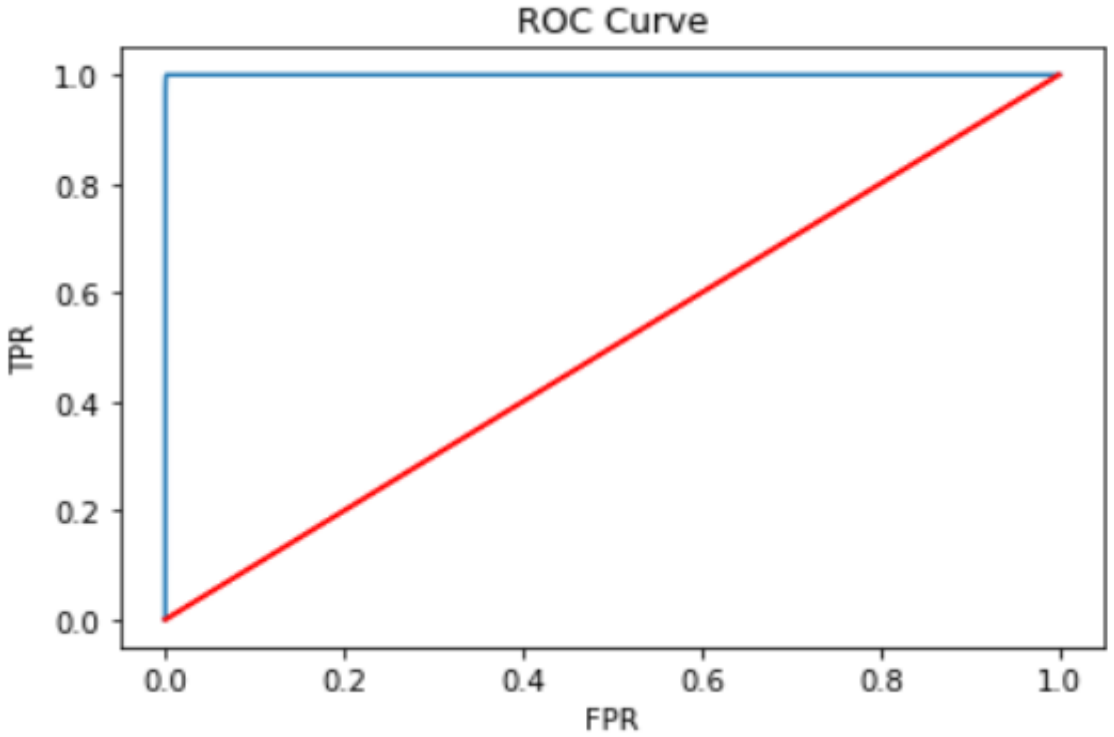
Decision tree model result

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1290008
1	1.00	1.00	1.00	1289955
accuracy			1.00	2579963
macro avg	1.00	1.00	1.00	2579963
weighted avg	1.00	1.00	1.00	2579963

- For Train Data, model is having Accuracy of 100% and Recall of 100%.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	552823
1	1.00	1.00	1.00	552876
accuracy			1.00	1105699
macro avg	1.00	1.00	1.00	1105699
weighted avg	1.00	1.00	1.00	1105699

- For Test Data, model is having Accuracy of 100% and Recall of 100%.



- ROC (receiver operating characteristic) curve tell us the performance of classification model at all classification thresholds.
- TPR(True Positive Rate) is exactly 100 % which indicates all Fraud Transactions were correctly predicted by the model.
- TPR needs to be as high as possible for a good model.

- Overall, Decision tree proves to be one of the **best model** with highest Recall and accuracy value.
- Random Forest model (i.e. a group of independent decision trees) have been similarly tested and it have high False Negative(i.e. Number of Actual fraud transactions, incorrectly identified as non-fraud) value.
- On other hand, Decision tree model have False Negative value of 0 indicating , all fraud transactions were correctly identified on both train and test dataset. Thereby, making it a **preferred model**.

Gradient Boosting model result

	precision	recall	f1-score	support
0	0.98	0.99	0.99	1290008
1	0.99	0.98	0.99	1289955
accuracy			0.99	2579963
macro avg	0.99	0.99	0.99	2579963
weighted avg	0.99	0.99	0.99	2579963

- Xtreme Gradient Boosting(XGBoost) is a robust modelling algorithm that provides parallel processing, better optimization and regularization.
- For Train Data, model is having Accuracy of 98-99% and Recall of 98%.

	precision	recall	f1-score	support
0	0.98	0.99	0.99	552823
1	0.99	0.98	0.99	552876
accuracy			0.99	1105699
macro avg	0.99	0.99	0.99	1105699
weighted avg	0.99	0.99	0.99	1105699

- For Test Data, model is having Accuracy of 98-99% and Recall of 98%.
- Overall, it is performing slightly less better than decision tree.

Financial Impact of model on Business

Before building the model:

- Company was facing high revenue loss and unable to predict the patterns of fraud transaction.
- Fraud transactions happening during Odd hours of day were not identified.
- Poor customer experience with lot of complaints.

After building the model:

- Real time model can predict fraud transaction patterns anytime, with one time built up cost associated with it.
- There would be less revenue loss due to fraudulent transaction to the Banks. If any transaction suspected as Fraud, OTP to be sent to Customer and second level authentication be done.
- Better customer experience in terms of less complaints and retaining existing customers. It would reduce the cost of customer care for each authentication call.

Recommendations

- Customer awareness to be increased and should be briefed about do's and don'ts while using Credit/Debit card.

- OTP based notification to be sent for Point of Sale(POS) which are lacking the feature. It would enable the Customer to report for any details theft immediately to the bank.

- Categories for which there were large no of fraud transactions should be closely monitored and OTP system to be built for security reasons.

Thank you