**Machine Learning Intern Assessment**

Churn Prediction Model

– A Brief Summary

Submission made by - Sidharth Chaudhary

The churn model predicts whether a customer will avail the company's benefits in the future.

The model was built in Jupyter Notebook in vs code IDE, and the approach to the model is as follows:

Data preprocessing:

1. All the Python libraries like pandas, numpy, matplotlib etc. were imported
2. Then the CSV file is read through pandas.
3. Then I got the overview of all the categorical features
4. Check for null values were made
5. Plots were made using matplotlib.

Feature engineering:

1. Relevant features were generated to help the model prediction. e.g. Gender feature was converted into binary form, also the location feature was converted to binary form using the “one-hot encoding “method.
2. Feature scaling was done using the sci-kit library
3. The split of the test and train tests were done.

Model building:

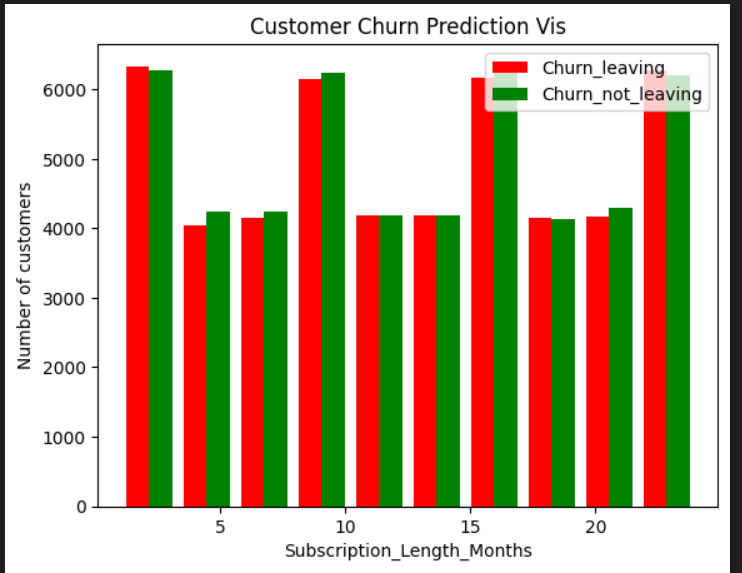
1. Different results were obtained by using three different machine learning algorithms i.e.. Neural networks, Logistic regression & random forests.
2. Training and validation was done
3. Evaluation was done using metrics (e.g., accuracy, precision, recall, F1-score).

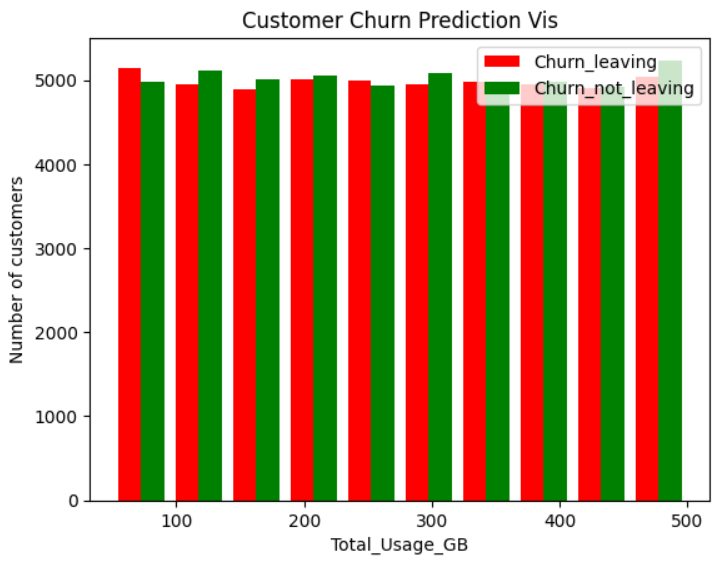
Model optimization:

1. Optimized features were used.
2. Hyperparameter tuning was done.

Visualizations:

A chart with red and green lines

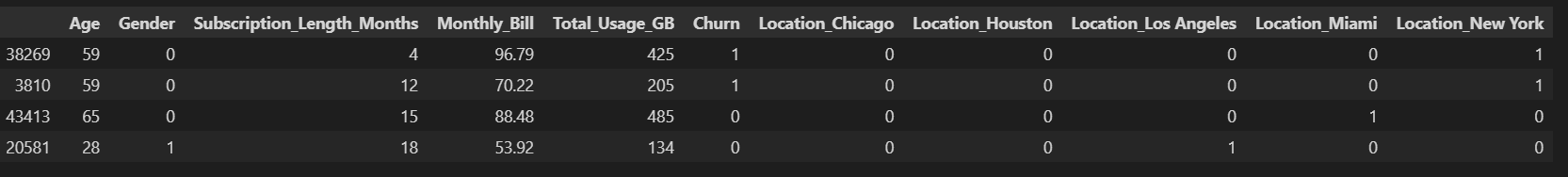
Description automatically generatedThese histograms represent the churning rate of customers against various features such as their subscription length months, monthly bill and total usage of data in GB.



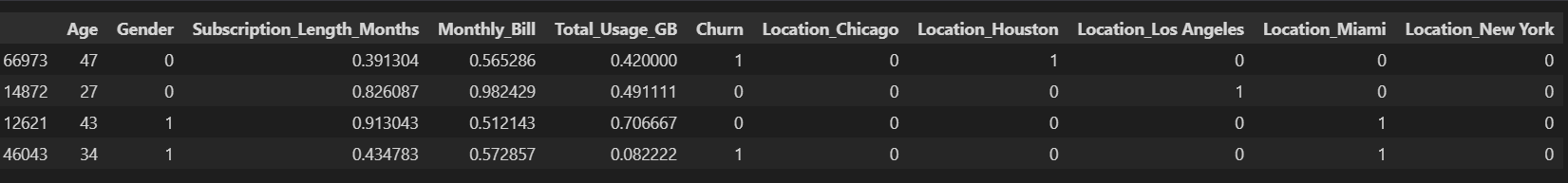
Feature Scaling and one hot encoding:

The below figures show the results before and after feature scaling, the one hot encoding is also shown in both the diagrams.

Before scaling (one hot encoding done)



After scaling



Metrics obtained after Using Neural Networks:

A screenshot of a computer

Description automatically generated

Metrics obtained after Using Logistic regression:

A black screen with white numbers

Description automatically generated

Metrics obtained after Random forests:

