Telecom Churn Case Study

Sri Nath S, Rudresh Sishodia, Sagnik Saha

Problem Statement

- In the telecom industry, customers are able to choose from multiple service providers and actively switch from one operator to another. In this highly competitive market, the telecommunications industry experiences an average of 15-25% annual churn rate. Given the fact that it costs 5-10 times more to acquire a new customer than to retain an existing one, customer retention has now become even more important than customer acquisition.
- To reduce customer churn, telecom companies need to predict which customers are at high risk of churn. In this project, you will analyse customer-level data of a leading telecom firm, build predictive models to identify customers at high risk of churn and identify the main indicators of churn.

Data Cleaning and Preparation

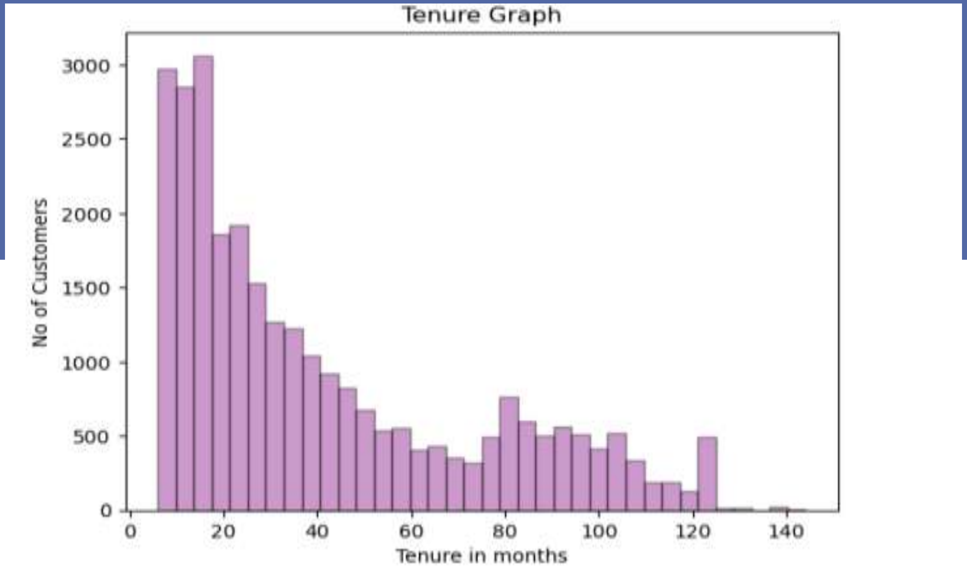
Data Preparation

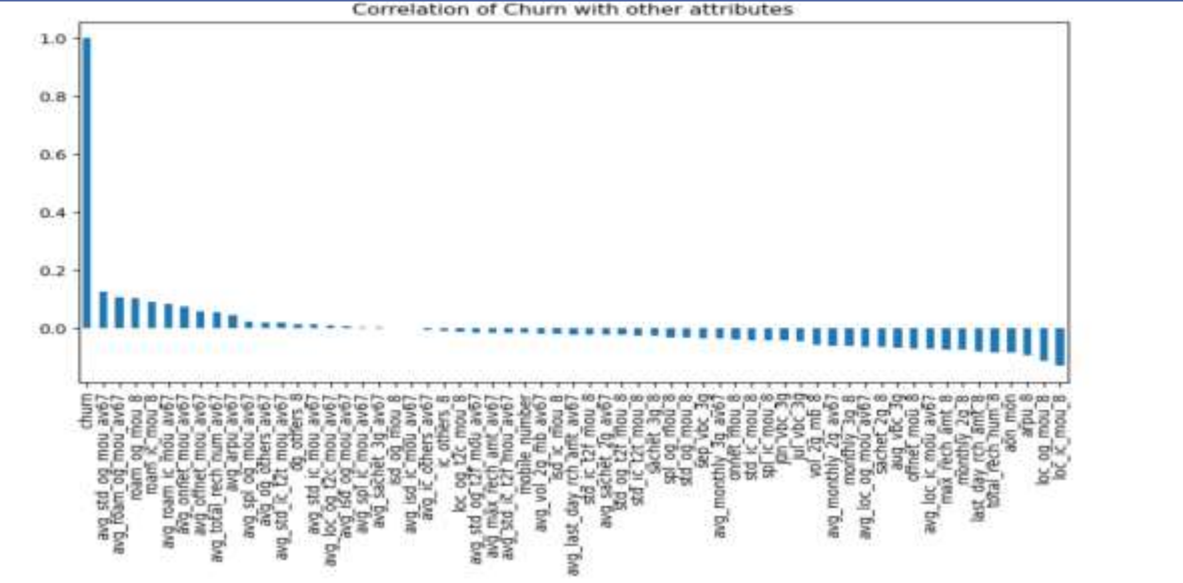
- For this case study, we would need to focus on the high value customers. To filter out the high value customers, calculated the total amount spent in the first two months(the good phase) and then filter the data based on the 70th percentile.
- Creating the Target/Churn column based on the usage for the last month. The customers who have not made any calls (either incoming or outgoing) AND have not used mobile internet even once will be tagged as churned customers.

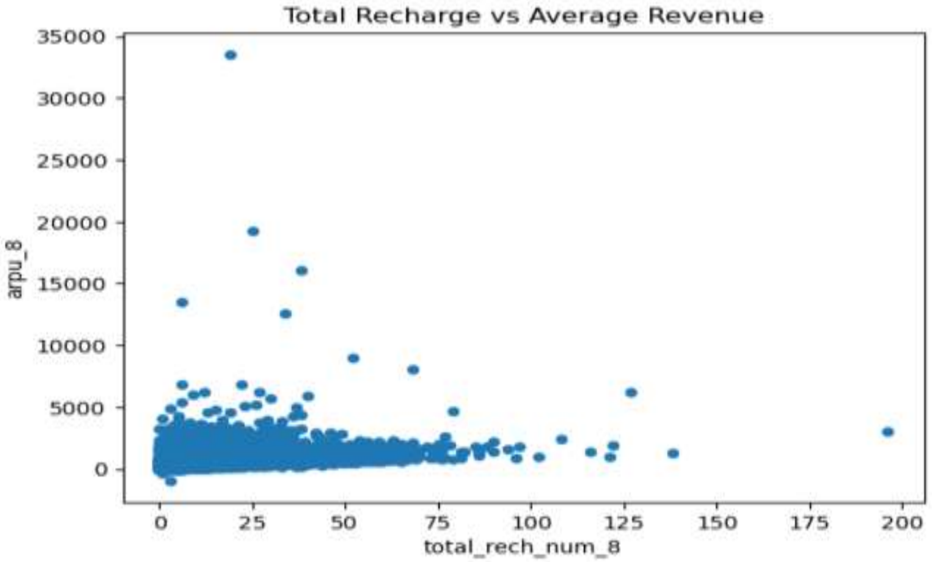
- Then dealing with columns which have only one unique value and dropping those columns as they wont be adding much insight in the model which will be created.
- Checking for missing values, columns with more than 40 percent missing values were dropped.
- Dropping columns which are highly correlated to avoid multicollinearity issue.
- Finally dropping the rows which have missing values to ensure there are no more missing values in the dataset.

Exploratory Data Analysis(EDA)

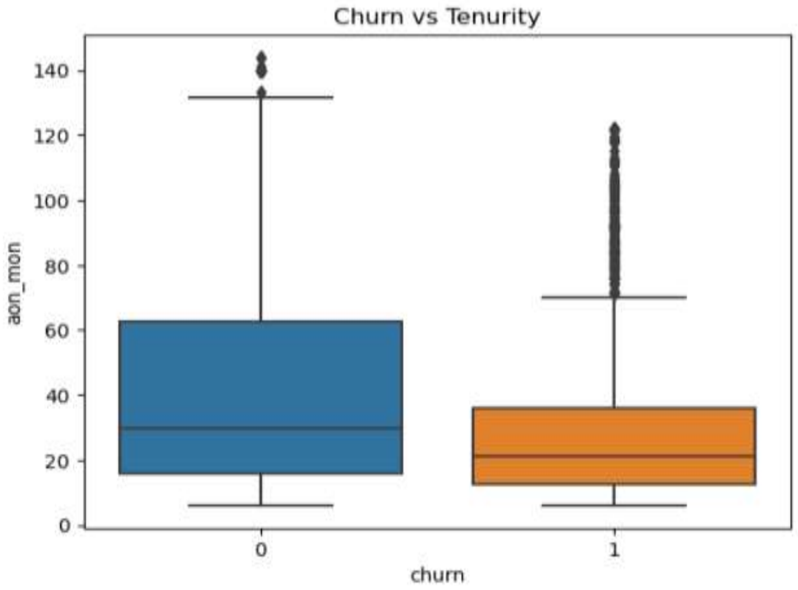
EDA

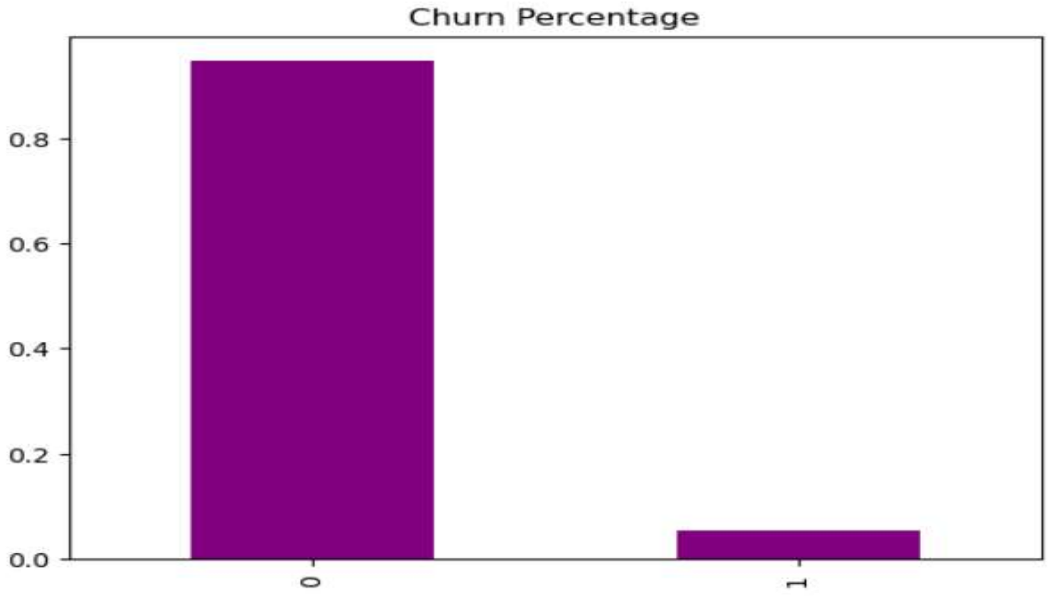

Tenure Graph
No of Customers
Tenure in months
0
500
1000
1500
2000
2500
3000
20
40
60
80
100
120
140


Correlation of Churn with other attributes
1.0
0.8
0.6
0.4
0.2
0.0
churn


Total Recharge vs Average Revenue
arpu_8
total_rech_num_8
35000
30000
25000
20000
15000
10000
5000
0
25
50
75
100
125
150
175
200

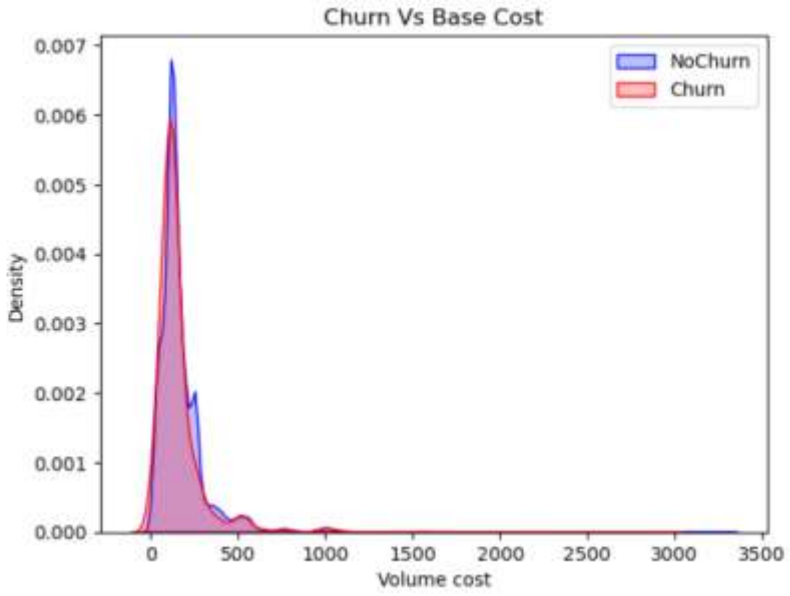
EDA contd.

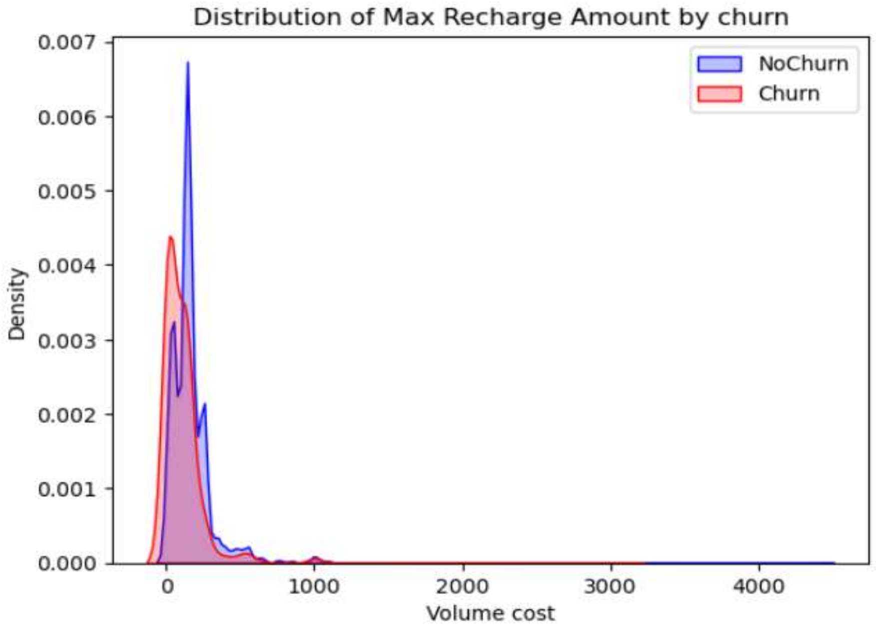




As we can see there is data imbalance in the churn column which was handled using an oversampling technique.
We can also see tenured customers churn less.

EDA contd.


Churn Vs Base Cost
NoChurn
Churn
Density
Volume cost
0.000
0.001
0.002
0.003
0.004
0.005
0.006
0.007
0
500
1000
1500
2000
2500
3000
3500


Distribution of Max Recharge Amount by churn
NoChurn
Churn
Density
Volume cost
0.000
0.001
0.002
0.003
0.004
0.005
0.006
0.007
0
1000
2000
3000
4000

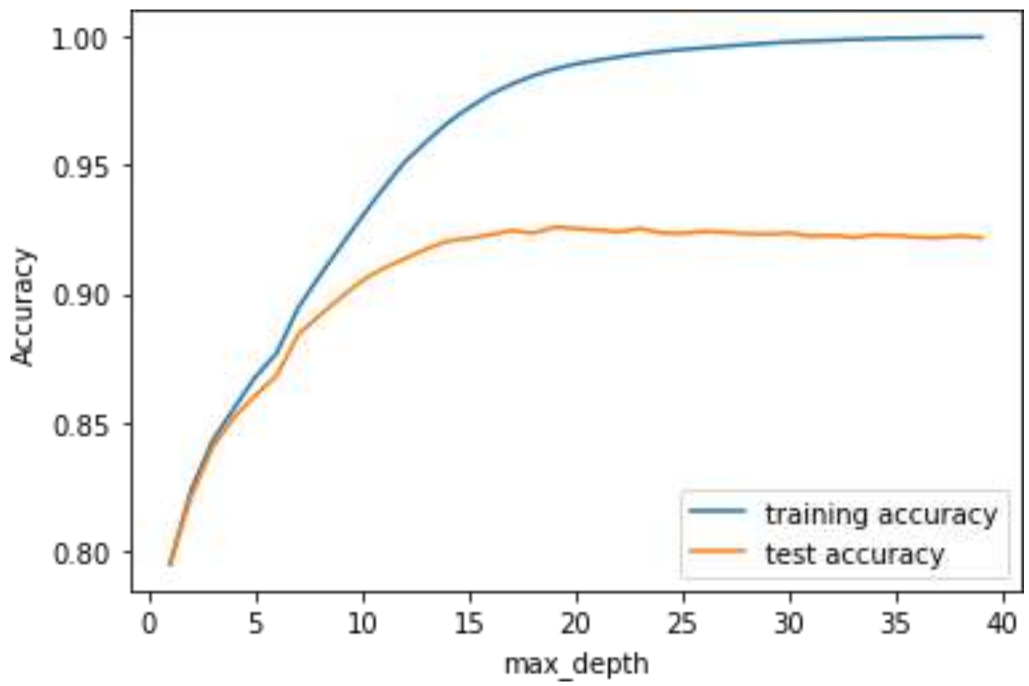
Insights

- While performing EDA, there were some distinct patterns which were noticed with customers who churn. It can be seen that newer customers are more likely to churn.
- While predicting churn, coulmns like local incoming,outgoing and the arpu for month 8 seems like important attributes. Usually, people who churn do not use the services in the preceding months.

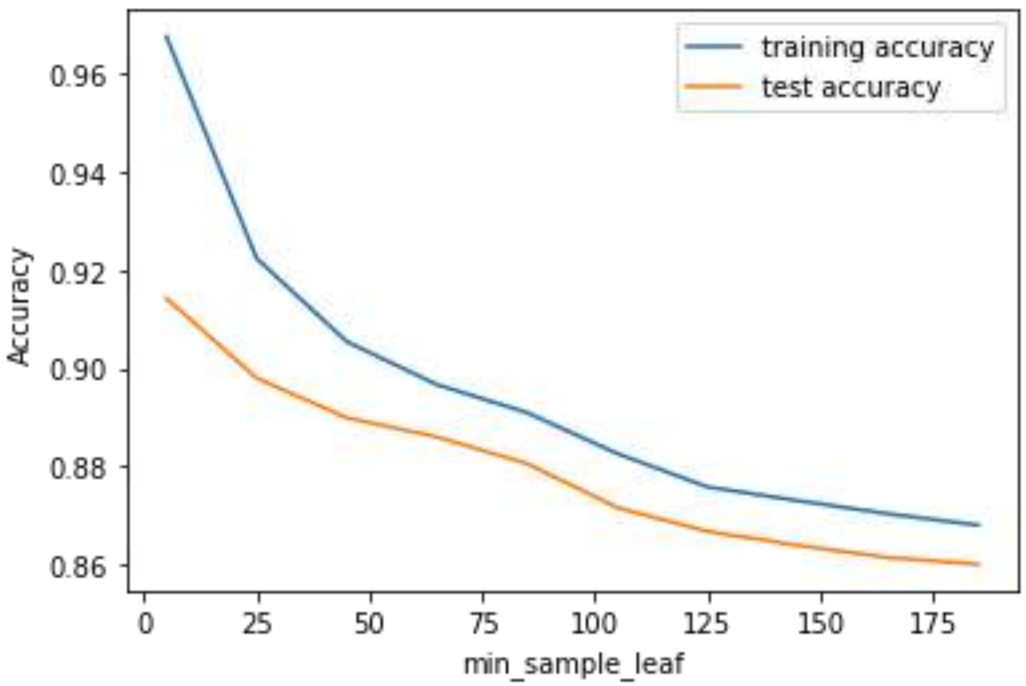
- The maximum recharge amount seems like a strong indicator of churn. When customers spend activity decreases, this could mean that customer is planning to quit service provider for another.

Data Modeling

Max depth chart



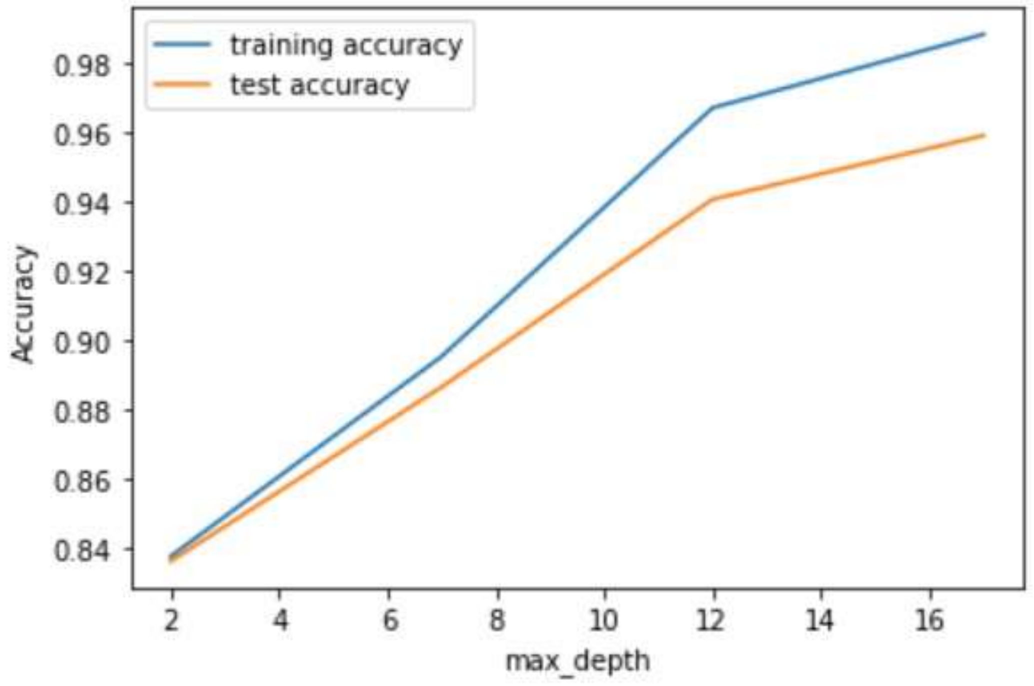
Minimum sample leaf chart



Decision Tree accuracy came out around 85%

Data Modeling

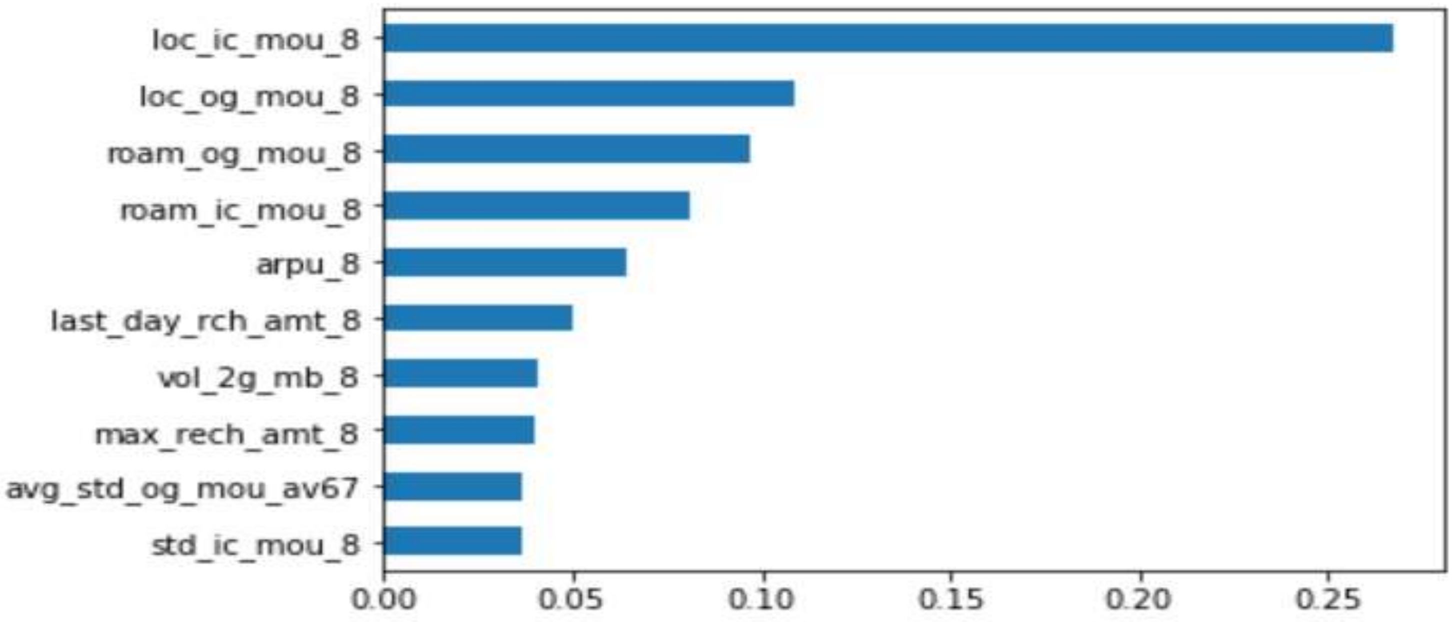
Random Forest:



Max_depth of around 11 gives optimal accuracy> 94%

Observations:

Linear SVM gave us accuracy of 94%.
Accuracy Score for Random Forest Final Model : 93%
Most Important features:



Conclusion

- The newer customers are more likely to churn. Customers with tenurity less than 4 years are more susceptible to churn hence it is important to focus on these customers and the service provider can roll out exciting offers and deals to gain their trust and loyalty.
- When the usage of customers dip, that’s an indicator that they are probably planning to switch to a different service provider. When such patterns are identified, these customers should be reached out and understand the reason behind not using the service.
- The spend activity decreases, this is also an indicator that customer is likely to churn. These customers should be targeted to understand the reasons behind their dissatisfaction with the services.

Thank You