A telecom company is facing a business problem where in customers are porting/switching over to other network providers.

There are variables in data are Churn, AccountWeeks   ContractRenewal   DataPlan   DataUsage   CustServCalls   DayMins   DayCalls   MonthlyCharge   RoamMins.  Variable definitions:

1. Churn variable represents if customer has churned or not, 1 means customer has churned and 0 means customer has not churned.

2. Account weeks: Number of weeks since account is active with company.

3. ContractRenewal: Has the customer renewed contract with us in last 5 years.

4. DataPlan: Does customer has internet data plan with us. 1 represent yes and 0 represents no.

5. DataUsage: GBs of Internet data used in each month

6. CustServCalls: Average number of customer service calls made per month.

7. DayCalls: Average Calls per day.

8. MontlyCharge: Monthly bills

9. RoamMins: Average of Roaming minutes for each day

The company would like to build a Logistic prediction model and utilize machine learning algorithms as well to identify factors which influence attrition of customers.

Workflow:

Input the data into the system, thereby checking the null values if any and converting the variables into their proper data types. Having done that built four models namely, Logistic regression with important feature selection, Decision tree, support Vector Machine and Random Forest algorithms. Lastly, compared these models against each other, and based on the Accuracy and AUC random forest outperformed the other models.

Notes: Entire data is being used for the model training.