Telecom CHURN

A case study on analysis in Telecom churn and build predictive models to identify customers at high risk of churn.

Ritik Mehta, Hemanshu Singh & Komal Sharma

September 2023

CONTENT

- *Business **PROBLEM** Overview*
- *Project **APPROACH***
- *Selecting **PREDICTION** models*
- ***PERFORMANCE** evaluate*
- *Case Study **CONCLUSION***

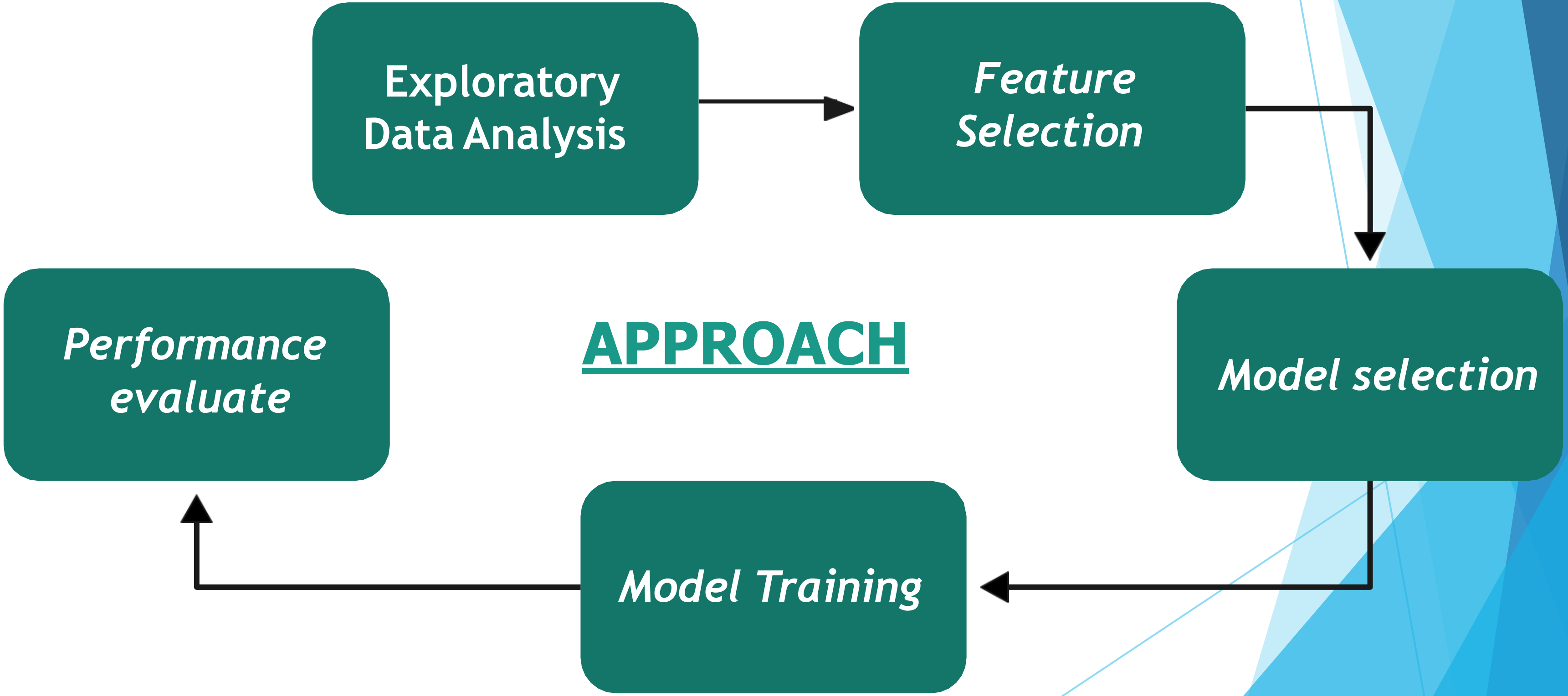
Business PROBLEM Overview

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.

For many incumbent operators, *retaining highly profitable customers is the number one business goal.*

To reduce customer churn, telecom companies need to predict which customers are at high risk of churn.

APPROACH


Exploratory Data Analysis
Feature Selection
Model selection
Model Training
Performance evaluate

Selecting PREDICTION Models

Support Vector Machine (SVM)

Versatile: SVMs can be applied to both classification and regression problems. They support different kernel functions, enabling flexibility in capturing complex relationships in the data. This versatility makes SVMs applicable to a wide range of tasks.

Effective in cases of limited data: SVMs can work well even when the training data set is small. The use of support vectors ensures that only a subset of data points influences the decision boundary, which can be beneficial when data is limited.

Ability to handle nonlinear data: SVMs can implicitly handle non-linearly separable data by using kernel functions. The kernel trick enables SVMs to transform the input space into a higher-dimensional feature space, making it possible to find linear decision boundaries.

Random Forest

Provides flexibility: Provides flexibility: Since random forest can handle both regression and classification tasks with a high degree of accuracy, Feature bagging also makes the random forest classifier an effective tool for estimating missing values as it maintains accuracy when a portion of the data is missing.

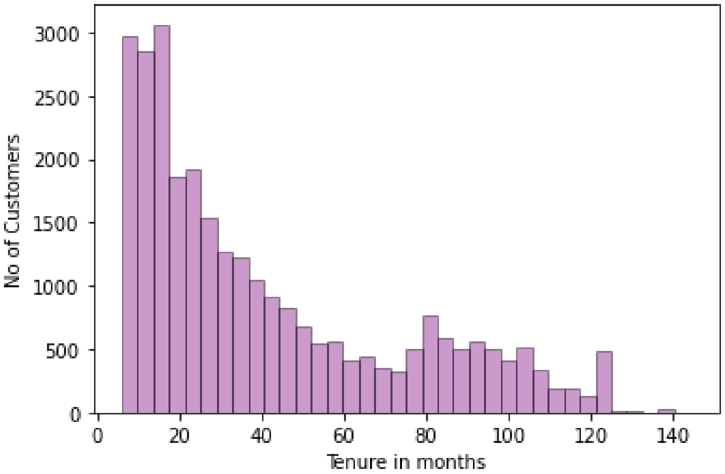
No feature scaling required: No feature scaling (standardization and normalization) required in case of Random Forest as it uses rule based approach instead of distance calculation.

Handles non-linear parameters efficiently: Nonlinear parameters don't affect the performance of a Random Forest unlike curve-based algorithms. So, if there is high non-linearity between the independent variables, Random Forest may outperform as compared to other curve-based algorithms

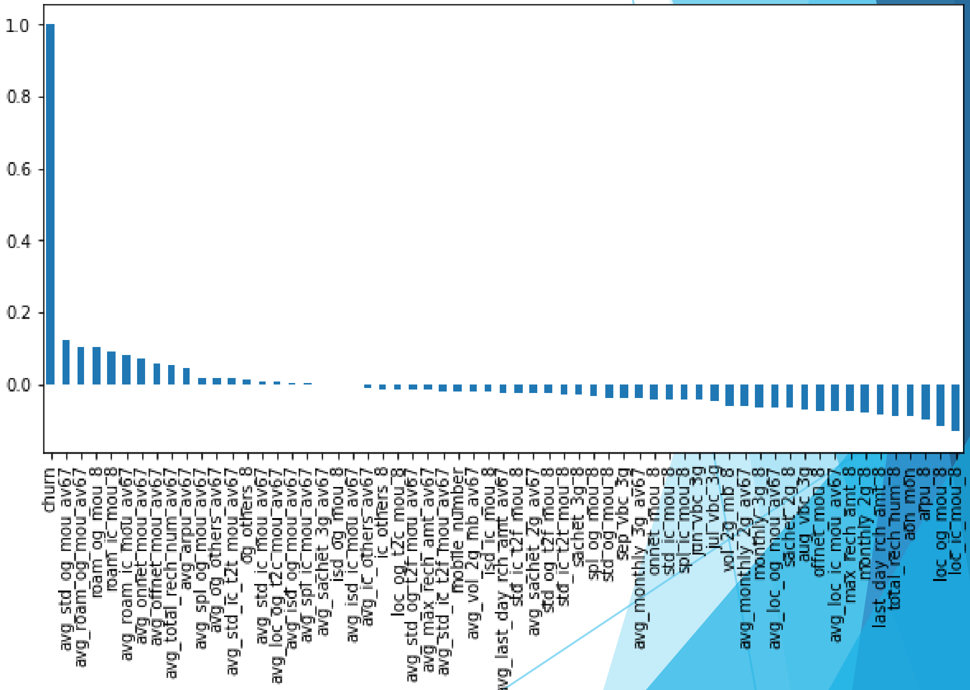
Steps followed in Telecom Churn Case Study:

- Telecom Churn - Case Study analysis and EDA
- Filtered high value Customers
- Derived new features from the existing columns
- Churn Vs other important features
- Decision Tree
- Fine tuning hyperparameters
- Random Forest & Conclusion
- Observations

Tenure of customers



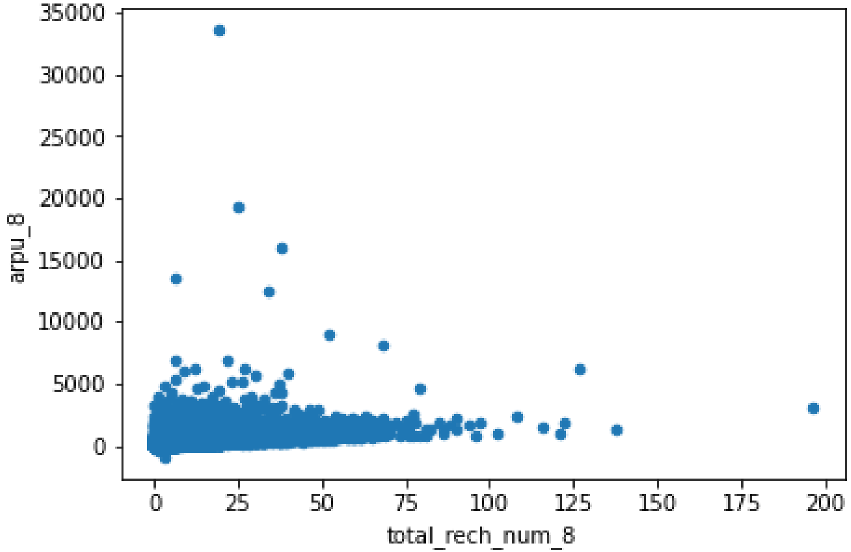
Correlation of churn with other factors

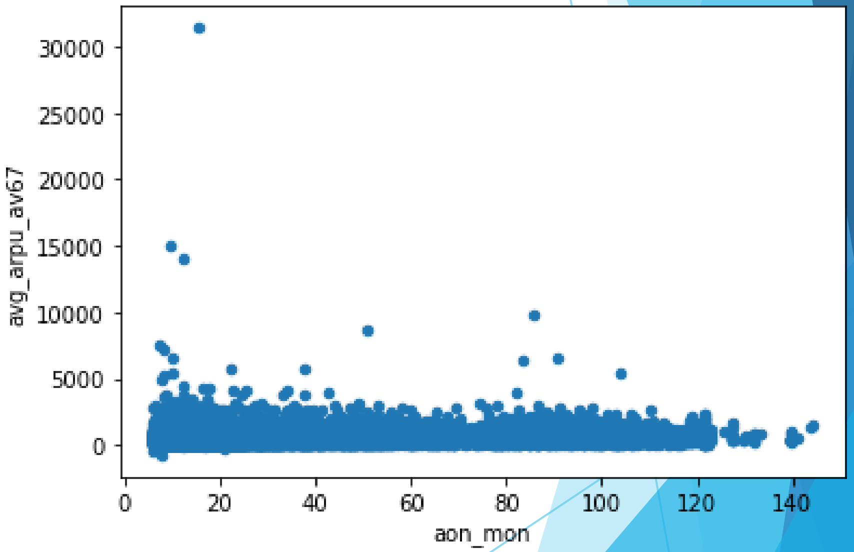


Observation: Avg Outgoing Calls & and calls on roaming for 6 and 7th months are positively correlated with churn.

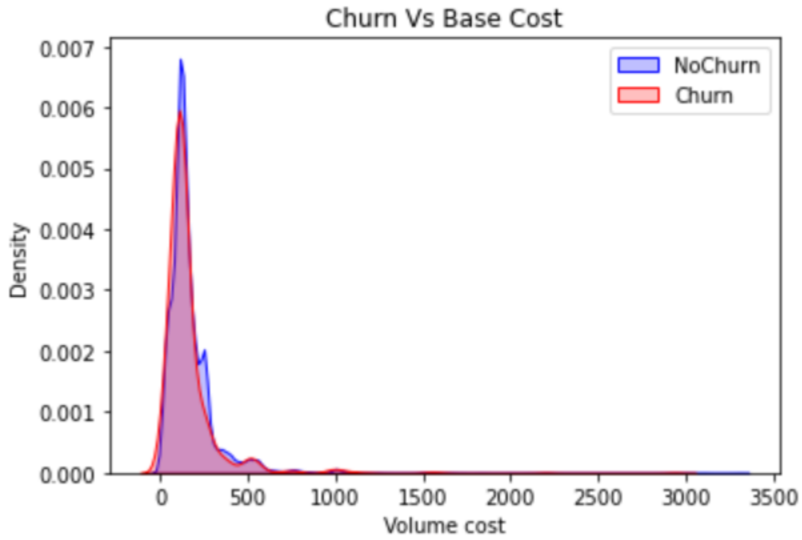
Avg Revenue, No. of Recharge for the 8th month has a negative correlation with churn.

Scatter plot between total recharge and avg revenue for the 8th month

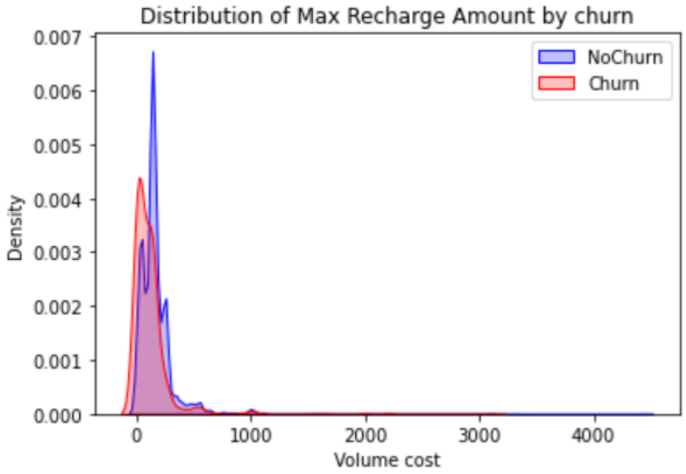




Churn Vs Base Cost

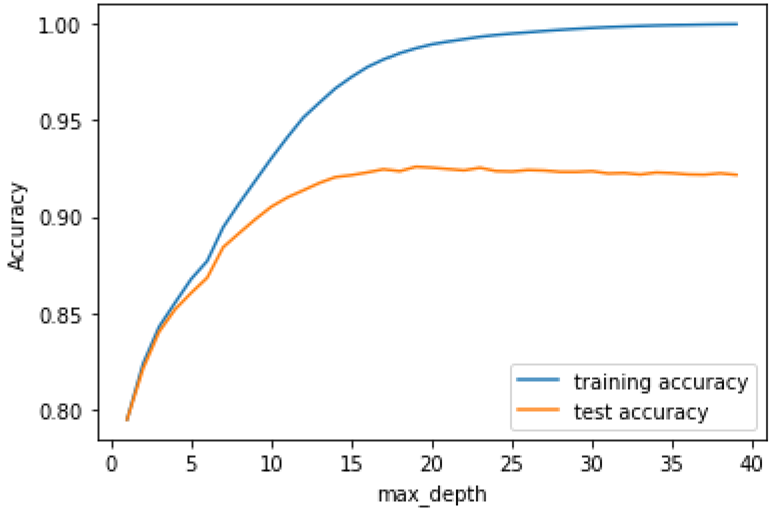


Distribution of Max Recharge Amount by churn

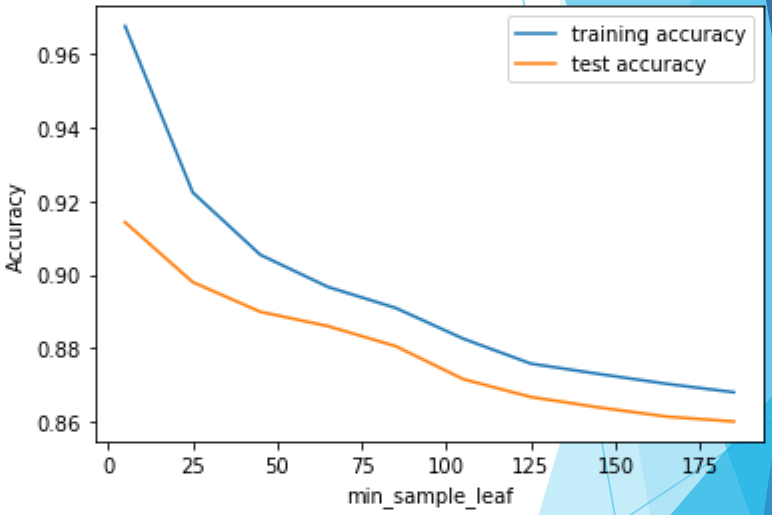


Churn/NoChurn vs Imp features comparison

Max depth chart

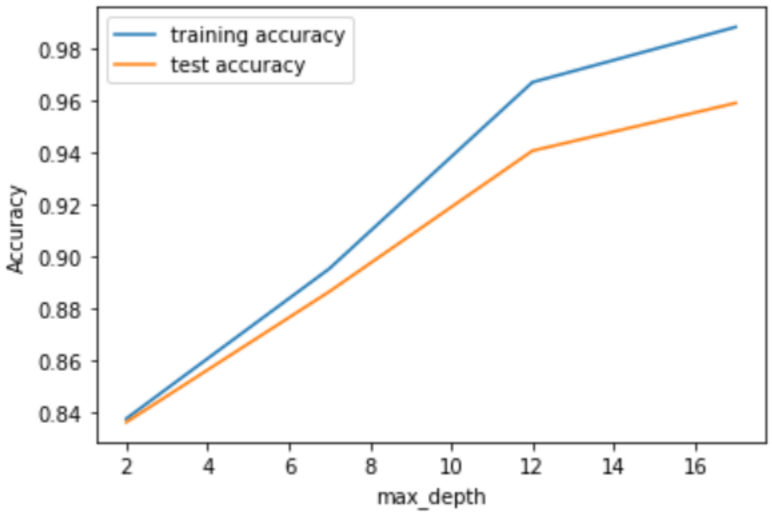


Minimum sample leaf chart



Decision Tree accuracy came out around 85%

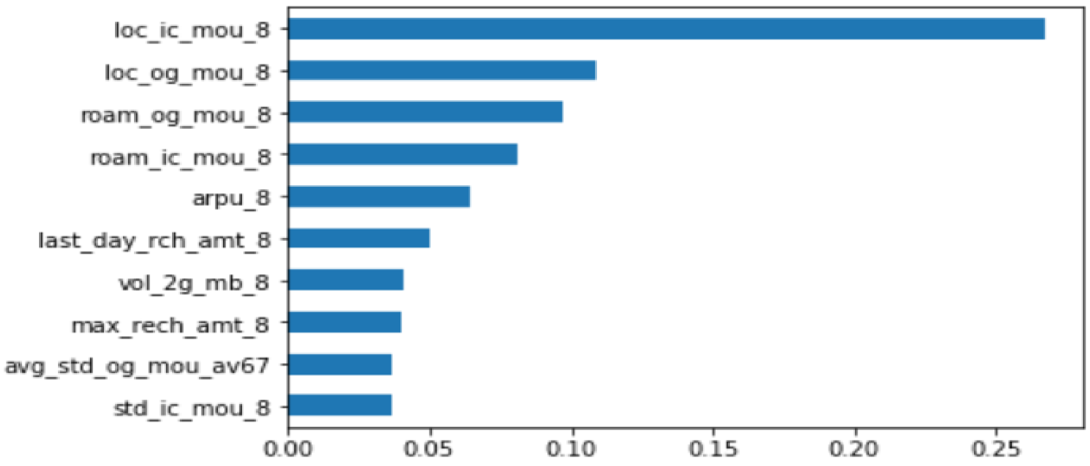
Random Forest:



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:



Case Study CONCLUTION

Std Outgoing Calls and Revenue Per Customer as Churn Indicators: Standard Deviation (Std) of outgoing calls and Revenue Per Customer are robust indicators of churn. High variations in outgoing calls or declining revenue per customer may signify dissatisfaction or reduced engagement.

Local Incoming and Outgoing Calls in 8th Month and Average Revenue in 8th Month: For predicting churn, specific columns such as local incoming and outgoing calls during the 8th month, as well as the average revenue in the 8th month, are critical. These metrics provide insights into the recent behavior and financial health of customers, both of which are pivotal for churn prediction.

After thorough analysis, Support Vector Machines (SVM) and Random Forest models have demonstrated the highest accuracy levels in predicting churn. These models outperform others in classifying churn data, making them prime candidates for future churn prediction efforts.

- **Customers with Tenure Less than 4 Years:** Customers with a tenure of less than 4 years are more likely to churn. This is because newer customers often have less attachment to the service and are more prone to exploring alternatives or switching providers.

- **Max Recharge Amount as a Strong Feature:** The maximum recharge amount serves as a robust feature for predicting churn. When customers start reducing their maximum recharge amounts, it can be a sign of declining commitment to the service, making them more likely to churn.

Thank you!