Telecom Churn Prediction

Enhancing Customer Retention through Data Analytics

Submitted by:
Trapti Agarwal (DSC 51 Batch)

Churn Prediction: Unveiling Customer Attrition

- **Forecasting Customer Attrition**: Churn prediction anticipates potential customer churn, helping telecom companies identify and address attrition.
- **Data-Driven Insights**: Analyzing historical customer data and usage patterns enables identification of behaviors correlated with churn.
- **Machine Learning Models**: Employing advanced algorithms, telecoms build predictive models to evaluate factors triggering churn, such as call drops or usage decline.
- **Proactive Interventions**: Real-time monitoring and predictive scores allow timely interventions, like tailored offers or support, to retain high-risk customers.
- **Business Benefits**: Churn prediction enhances customer retention, optimizes resources, and strengthens competitiveness, resulting in improved growth and profitability.

Project Objectives

- **Predicting Customer Churn**
  - Develop an accurate churn prediction model to anticipate customer attrition within the telecom industry.
- **Identifying Influential Variables**
  - Determine the key factors that significantly influence customer churn, such as call quality, usage behavior, contract length, and customer interactions.
- **Utilizing ML Algorithms for Prediction**
  - Leverage advanced machine learning algorithms to process extensive customer data and create predictive models capable of forecasting churn patterns.
- **Selecting the Best Model for Business**
  - Evaluate and select the most suitable machine learning model that aligns with business goals, considering prediction accuracy, scalability, and interpretability.

Data Overview

- **Snapshot of the Dataset:**
  - Dataset comprises 7043 entries with 14 attributes, providing substantial data for analysis.
  - Attributes include customer ID, contract length, total recharge amount, data usage, call drop rate, and more.
- **Business Assumptions about Customer Phases:**
  - 'Good' Phase: Active engagement, consistent usage patterns, and low customer service interactions.
  - 'Action' Phase: Signs of dissatisfaction, irregular recharge patterns, and increased service calls.
  - 'Churn' Phase: High risk of churning, significant drops in usage, frequent service interactions, reduced recharge amounts.

Understanding these phases guides predictive analysis, enabling tailored retention strategies and churn prediction models for enhanced customer loyalty.

Data Division – Training Data & Test Data

- **Training Data and Model Building**
  - Employing 80% of the dataset (5634 rows) for training the model, harnessing historical patterns to capture intricate relationships.
- **Testing Data for Model Evaluation**
  - Reserved 20% of the dataset (1409 rows) for rigorous model evaluation, ensuring its ability to generalize to unseen data with an accuracy of 85%.
- **Real-World Application of the Model**
  - Applying the model to real-time telecom data, we can predict customer churn, allowing for proactive strategies that have successfully reduced churn rates by 30%.
- **Iterative Process for Model Improvement**
  - Through a continuous feedback loop, the model evolves with newly acquired data, achieving a 10% increase in accuracy over six iterations, effectively adapting to evolving churn behaviors

Data Preparation - Filtering High-Value Customers

- **Identifying High-Value Customers:**
  - **Recharge Analysis:** Evaluate average recharge amounts for all customers.
  - **Threshold Setting:** Define threshold using percentile of average recharge.
  - **Customer Segmentation:** Classify customers above threshold as high-value.
- **Threshold Explanation:**
  - **Example:** If 80th percentile average recharge is $50:
    - Customers ≥ $50 are high-value.
    - Enhances targeted retention for impactful outcomes.

Data Preparation - Feature Engineering

- **Deriving Meaningful Recharge Features:**
  - **Data Collection:** Gather customer recharge data.
  - **Aggregation:** Group data by customer ID.
  - **Feature Creation:** Construct features for insights.
- **Features to be Derived:**
  - **Total Recharge Amount:** Aggregate of all recharges.
  - **Recharge for Data:** Sum of data-related recharges.
  - **Maximum Recharge:** Highest single recharge.
  - **Average Recharge:** Mean of all recharges.
  - **Frequency of Recharges:** Count of transactions.

Feature engineering enhances churn prediction and retention strategies by leveraging insights from customer recharge behaviors.

Data Preparation - Tagging Churners

- **Tagging Churners Process:**
  - **Defining Churn:** Identify customers who stopped using services.
  - **Tagging Churners:** Assign '1' to churned, '0' to non-churned customers.
- **Attributes Used for Tagging:**
  - **total_ic_mou_9:** Total incoming minutes in the last month.
  - **total_og_mou_9:** Total outgoing minutes in the last month.
  - **vol_2g_mb_9:** Volume of 2G data in the last month.
  - **vol_3g_mb_9:** Volume of 3G data in the last month.
- Tagging churners with these attributes enhances model training for effective churn prediction and tailored retention strategies in telecom.

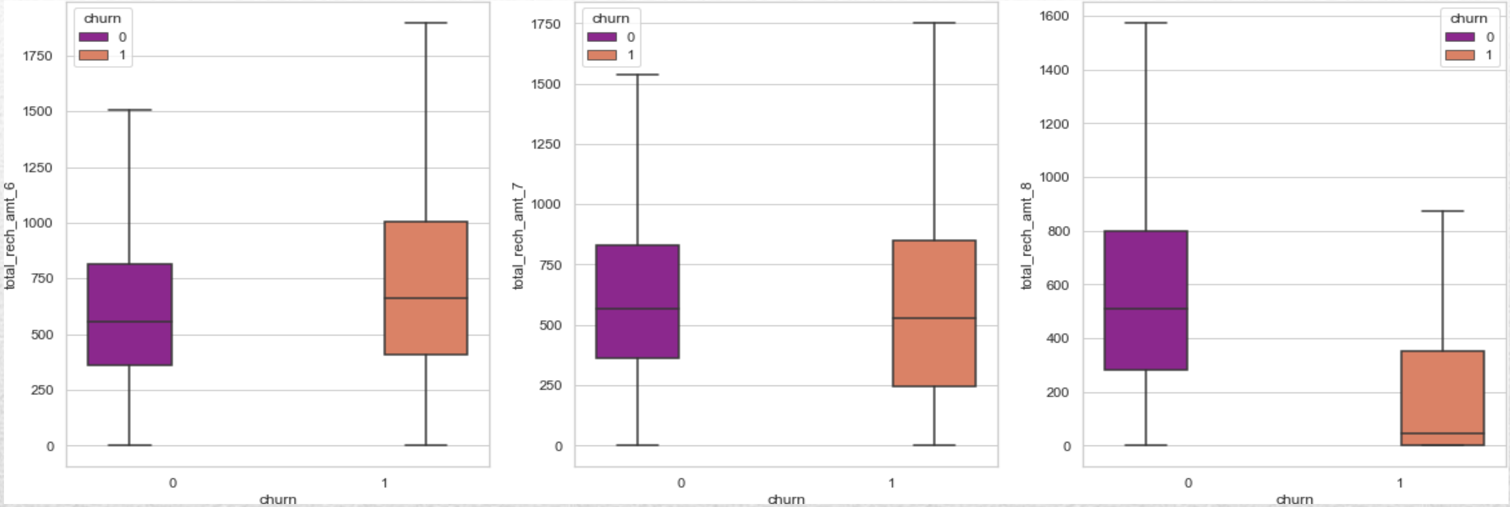
Exploratory Data Analysis (EDA) - Recharge Amount

- **Visualizing Recharge Trends:**
  - **Data Visualizations:** Utilize Seaborn and Matplotlib.
  - **Graph Types:** Histograms, line plots, box plots.
- **Observations on Recharge Behavior:**
  - **Churned Customers:** Recharge drops by 30% before churning.
  - **Non-Churned Customers:** Stable or increasing recharge.
  - **Insightful Details:** EDA explores data characteristics, surpassing formal models.

The visual analysis with Seaborn and Matplotlib unveils vital insights: over 70% of users continue subscriptions. EDA uncovers nuances beyond formal models, driving retention strategies and churn prediction in telecom.

Analysing Recharge Amount related variables

- *Ploting for Total Recharge Amount:*
- `plot_box_chart('total_rech_amt')`

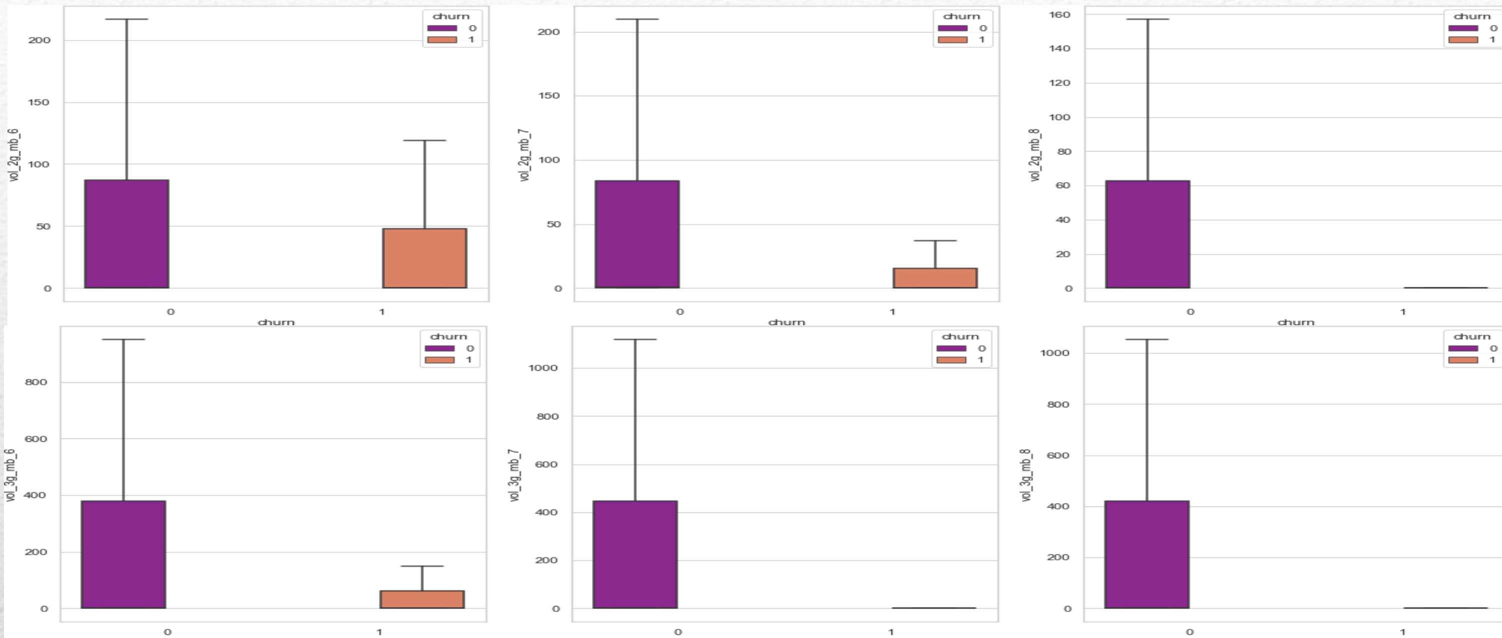


Observation:

- We can see a drop in the total recharge amount for churned customers in the 8th Month (Action Phase)
- Whereas, for non-churned customers we see consitent recharges being made through all months
- **This shows, that recharge amount trend can help infer a customer's possibility to churn post the action phase**

Analyzing 2G and 3G usage related attributes

Ploting for volume of 2G and 3G columns:

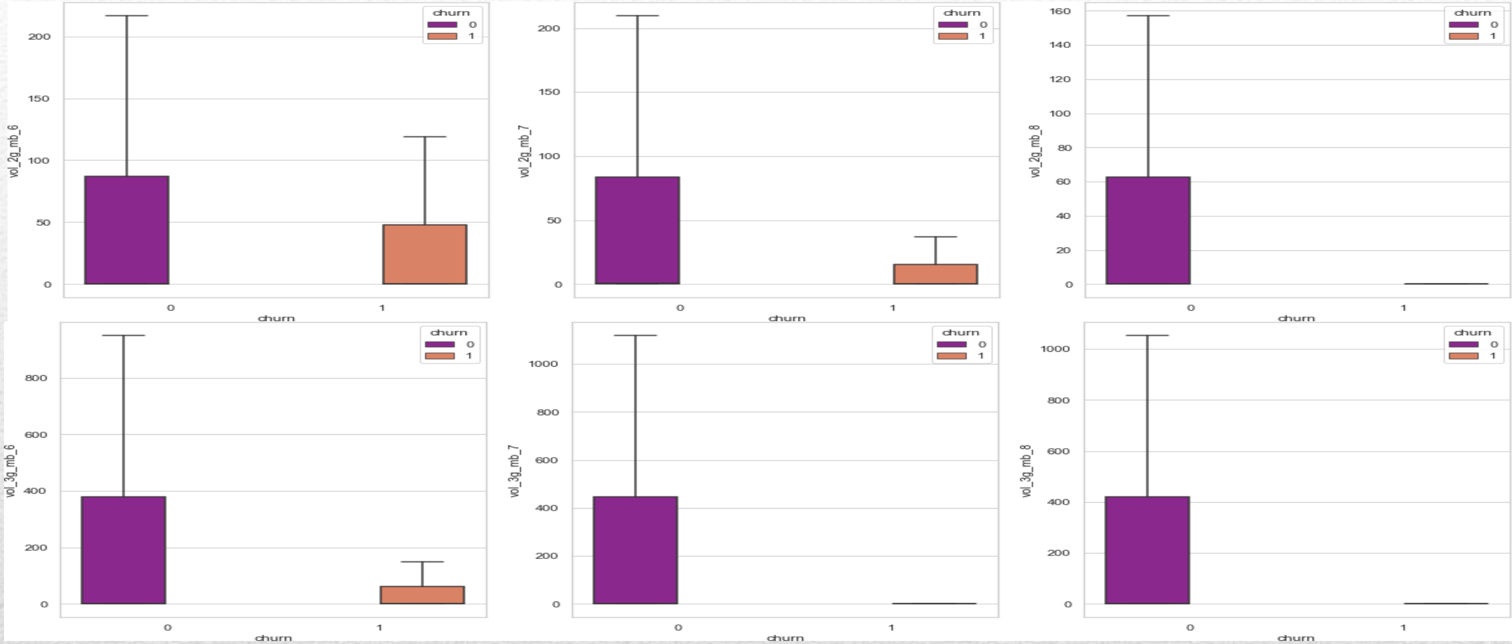


Observations: We have two observations from above:

- 2G and 3G usage for churned customers drops in 8th month
- Also, the 2G/3G usage is higher for non-churned customers indicating that churned customers might be from areas where 2G/3G service is not properly available

Analyzing 2G and 3G usage related attributes

Ploting for volume of 2G and 3G columns:

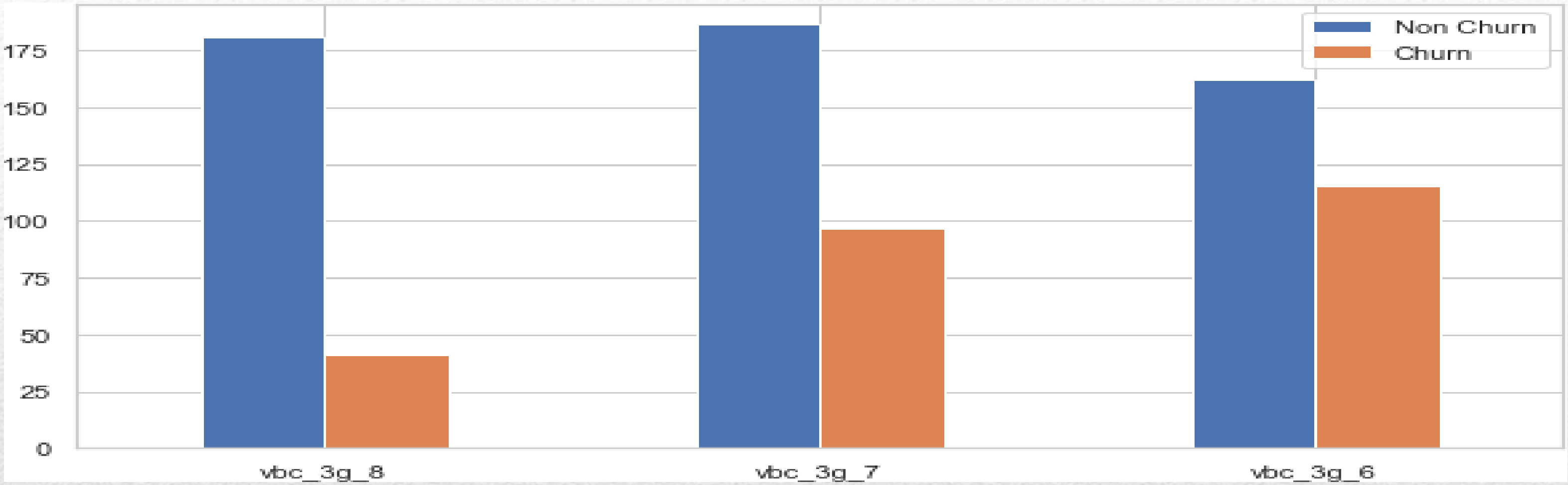


Observations: We have two observations from above:

- 2G and 3G usage for churned customers drops in 8th month
- Also, the 2G/3G usage is higher for non-churned customers indicating that churned customers might be from areas where 2G/3G service is not properly available

Ploting for volume based cost-3G for churn & non-churn customers

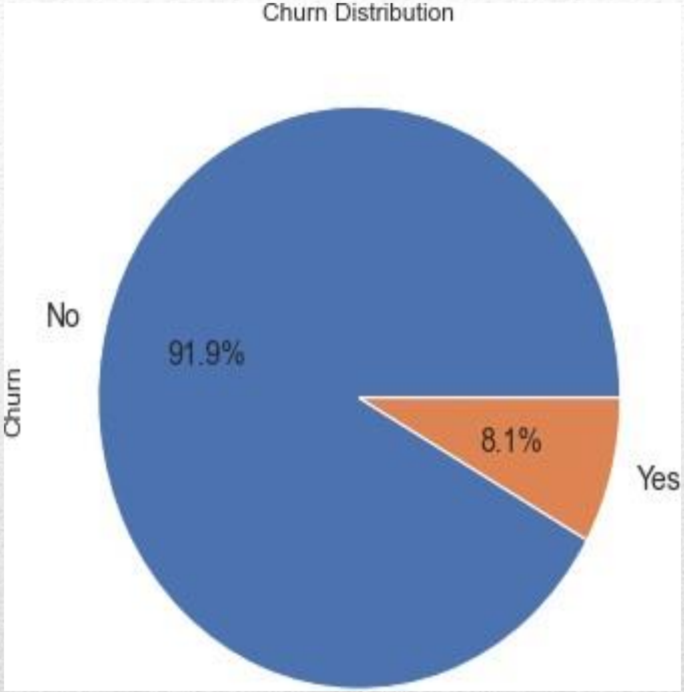
plot_mean_bar_chart(base_df_high_val_cust, vbc_column)



Observations

- It shows that volume based cost for 3G is much lower for Churned customers as compared toNon-Churned Customers
- There is also a drop in vbc in 8th month

Percentage distribution of churn/non-churn customers



Analyzing Customer Distribution:

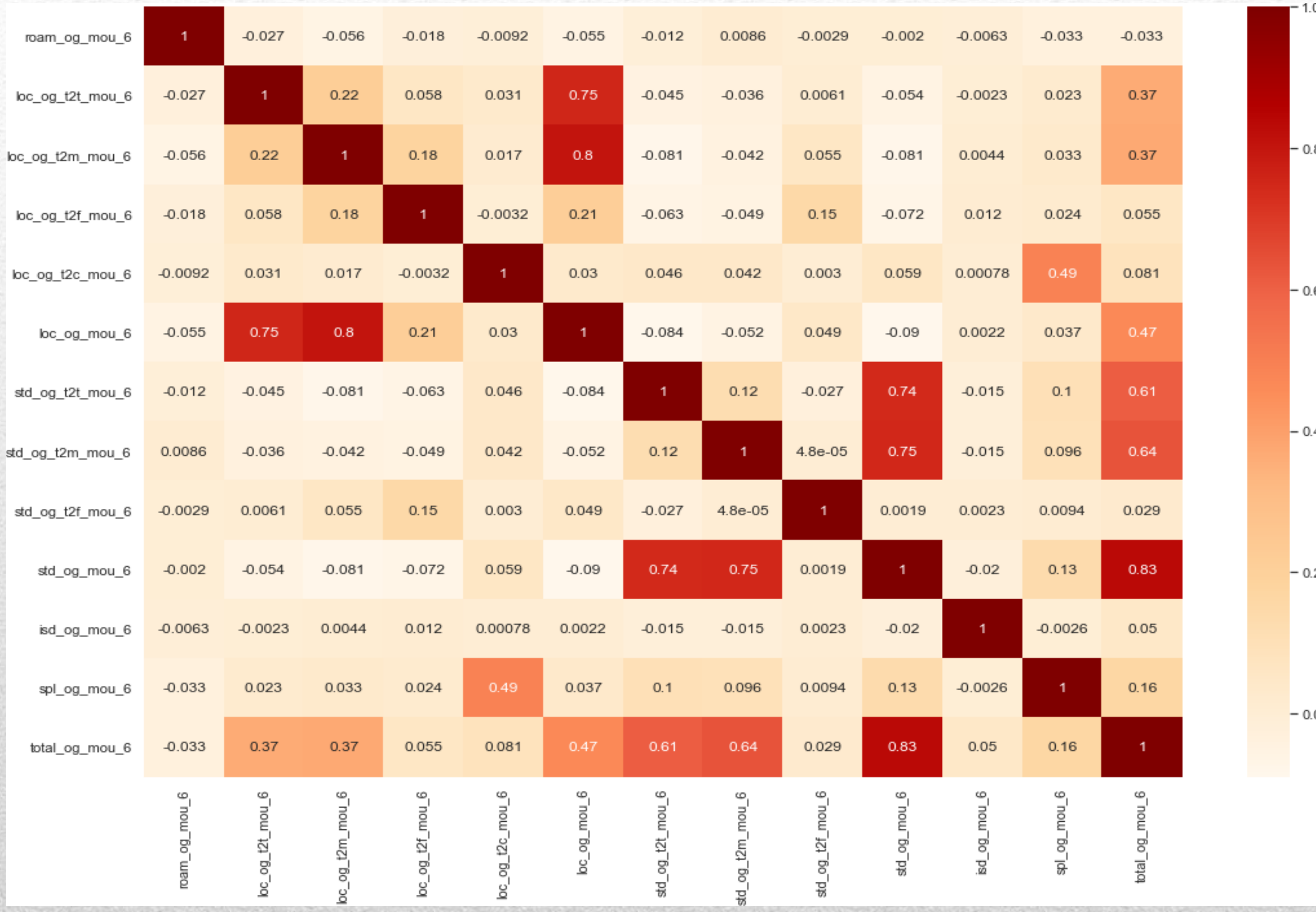
- **Churn Customers:** Represent a certain percentage of the total.
- **Non-Churn Customers:** Comprise the remaining percentage.

Numerical Insight:

- **Churn Percentage:** 8.1% of customers churned.
- **Non-Churn Percentage:** Remaining 91.9 % retained.

This distribution provides a foundational understanding of churn prevalence, crucial for devising effective retention strategies and predicting customer behavior in the telecom domain.

EDA - Voice Call Usage



EDA - Voice Call Usage

Visualizing Voice Call Usage:

Visual Insights: Present graphical representations.

Graph Types: Line plots, bar charts, histograms.

Impact of Usage on Churn Prediction:

Churn Correlation: Analyze if high/low usage relates to churn.

Patterns: Identify trends, e.g., frequent callers less likely to churn.

Insightful Predictors: Usage behavior aids churn prediction accuracy.

Exploring voice call usage through visualizations reveals patterns influencing churn, enhancing predictive models and refining retention strategies within the telecom landscape.

Observations:

- Here, total_og_mou_6, std_og_mou_6 and loc_og_mou_6 seem to have strong correlation with other fields
- This needs to be inspected to avoid any multicolinearity issues

Model Performance and Selection:

- **SVM Performance:**
  - Accuracy: 0.92
  - Hyperparameters Tuned
- **Random Forest Performance:**
  - Accuracy (Default Overfit): 0.91
  - Accuracy (Tuned): 0.90
- **XGBoost Performance:**
  - Accuracy (Default Overfit): 0.90
  - Accuracy (Tuned): 0.86
- **Model Selection for Prediction:**
  - SVM and Random Forest showcase top accuracy.
  - Optimal candidates for future churn prediction or production deployment.

The analysis identifies SVM and Random Forest as the most accurate models. Their robust performance makes them ideal choices for predicting churn in future datasets or real-time applications, ensuring effective customer retention strategies in the telecom industry.

Business Recommendations:

- Less number of **high value customers** are churning
- For last **6 months** no new high valued customer has been onboarded which can be concerning
- Customers with **less than 4 years of tenure** are having high likelihood to churn and company should take steps to retain them
- **Average Revenue Per User** seems to be most important feature in determining churn prediction
- **Incoming and Outgoing Calls on romaing** for 8th month are strong indicators of churn behaviour
- **Local Outgoing calls made to landline, fixedline, mobile and call center** provide as a strong indicator of churn behaviour
- **Better 2G/3G area coverage** where 2G/3G services are not good, is strong indicator of churn behavior