Telecom Churn Project

PREPARED BY: RENUKA BHADAKWADE

1. Data Understanding

File Name: Telecom_churn_data.csv

The dataset contains mobile_no, average revenue,minutes of usage, age of network,usage pattern,recharge behavior and other matrics over 4 months ie June to September

Purpose: The dataset is likely used for analyzing customer behavior, predicting churn, and understanding usage patterns.

1. Data Understanding

- Exploratory Data Analysis (EDA) using visualizations
- Check that data set using read_csv using pandas, then using head() to check first 5 rows.
- Data Defination of data set given below on next page…

Sr.No.	Columns	Descriptions
1	MOBILE_NUMBER	Customer phone number
2	CIRCLE_ID	Telecom circle area to which the customer belongs to
3	LOC	Local calls - within same telecom circle
4	STD	STD calls - outside the calling circle
5	IC	Incoming calls
6	OG	Outgoing calls
7	T2T	Operator T to T, i.e. within same operator (mobile to mobile)
8	T2M	Operator T to other operator mobile
9	T2O	Operator T to other operator fixed line
10	T2F	Operator T to fixed lines of T

Sr.No.	Columns	Descriptions
11	T2C	Operator T to it's own call center
12	ARPU	Average revenue per user
13	MOU	Minutes of usage - voice calls
14	AON	Age on network - number of days the customer is using the operator T network
15	ONNET	All kind of calls within the same operator network
16	OFFNET	All kind of calls outside the operator T network
17	ROAM	Indicates that customer is in roaming zone during the call
18	SPL	Special calls
19	ISD	ISD calls
20	RECH	Recharge

Sr.No.	Columns	Descriptions
21	NUM	Number
22	AMT	Amount in local currency
23	MAX	Maximum
24	DATA	Mobile internet
25	3G	3G network
26	AV	Average
27	VOL	Mobile internet usage volume (in MB)
28	2G	2G network
29	PCK	Prepaid service schemes called - PACKS
30	NIGHT	Scheme to use during specific night hours only

Sr.No.	Columns	Description
31	MONTHLY	Service schemes with validity equivalent to a month
32	SACHET	Service schemes with validity smaller than a month
33	*.6	KPI for the month of June
34	*.7	KPI for the month of July
35	*.8	KPI for the month of August
36	*.9	KPI for the month of September
37	FB_USER	Service scheme to avail services of Facebook and similar social networking sites
38	VBC	Volume based cost - when no specific scheme is not purchased and paid as per usage

2. Data Cleaning

Data Size

Rows : 99999 Columns: 226

- Handling missing values – Dropped columns with **more than 75% missing values** to reduce noise.
- Identified and removed duplicate entries to avoid data redundancy.
- Treating outliers – used IQR method

3. Data Preprocessing

- Filtering high-value customers – for that we are creating a cost metric for 6th and 7th month using quantile method
- Feature engineering: Creating new derived churn variables.& creating derived matrics
- Standardization and normalization of numerical data
- Visualize data using pie, count, box,bar,hist & heatmap plots

4. Model Building

We will create 2 types of models:

1. Accuracy Models: which will focus on the model accuracy (here: recall since we want to identify more and more churn customers (Logistic Regression with PCA, Random Forest & Gradient Boosting)
2. Interpretability Models: which can suggest the important variables. (Logistic Regression without PCA & Decision Tree without PCA)

4. Model Building

- **Logistic Regression**
- **Random Forest Classifier**
- **Gradient Boosting Classifier using random search**
- **PCA for dimensionality reduction**

5. Model Evaluation

- Accuracy, Precision, Recall, F1-Score
- Confusion Matrix Analysis
- ROC Curve and AUC Score
- Hyperparameter Tuning using RandomSearchCV
- High Accuracy model Gradient Boosting gives AUC 0.81 of and Recall of 71.6%.

6. Business Insights & Recommendations

- Identifying key factors leading to churn
- Business actions to reduce churn rate
- Customer retention strategies

Final Insights

Recommendations based on the columns achieved from Feature Importance:

- Provide cheap Local plans to the customers (loc_ic_mou_8).
- Watch out for the sudden drop in the ARPU (arpu_diff).
- Providing the combo plans to the customer based on their usage (max_rech_amt_8).
- Provide cheap Roaming plans to the customers (roam_og_mou_8 & roam_ic_mou_8).