Telecom Churn Prediction Project

By:- Shiv Om Mishra

1. Introduction and Problem Sta

Business Problem Overview

In the telecom industry, customers have the flexibility to choose from multiple service providers and can easily switch from one operator to another. This results in an annual churn rate of 15-25%. Given that acquiring a new customer costs 5-10 times more than retaining an existing one, customer retention has become a critical focus for telecom companies.

The objective of this project is to analyze customer-level data from a leading telecom firm, build predictive models to identify customers at high risk of churn, and determine the key factors influencing churn.

2. Understanding Data

Dataset Overview

The dataset contains customer-level information for four consecutive months: June, July, August, and September. The business goal is to predict churn in the ninth month (September) using data from the first three months.

Customer Behavior Phases

Customers typically do not churn instantly; the process happens in phases:

- **Good Phase (Months 6 & 7):** Customers are satisfied and exhibit normal usage behavior.
- **Action Phase (Month 8):** Customers may start reducing usage, indicating dissatisfaction.
- **Churn Phase (Month 9):** Customers stop using the service entirely.

3. Data Preparation

Step 1: Filtering High-Value Customers

Since 80% of revenue comes from 20% of customers, we define high-value customers as those who have recharged an amount greater than or equal to the 70th percentile of the average recharge amount over the first two months.

Step 2: Identifying Churners and Removing Churn Phase Data

Customers are tagged as churners if they have not made any outgoing or incoming calls and have not used mobile internet in the churn phase. The key indicators used for this are:

- `total_ic_mou_9`
- `total_og_mou_9`
- `vol_2g_mb_9`
- `vol_3g_mb_9`

Once churners are tagged, all data from month 9 is removed to ensure the model only uses past data for prediction.

Step 3: Handling Missing Values

- Numerical columns are filled with their median values.
- Categorical columns are filled with their mode.
- Any duplicate records are removed.

Step 4: Handling Outliers

- Outliers are detected using the Z-score method and removed if their Z-score is greater than 3.

4. Feature Engineering

- **Trend-based Features:** Change in ARPU, call minutes, and data usage over months.
- **Aggregated Features:** Total recharge amount, total call duration, and internet usage.
- **Customer Segmentation:** Categorizing customers based on voice/data usage patterns.

5. Model Building

Step 1: Handling Class Imbalance

Since churners are typically a small fraction (5-10%), we use:

- Oversampling techniques like SMOTE
- Class-weighted models to balance the dataset

Step 2: Model Selection

We train multiple models and compare performance:

- **Logistic Regression:** For interpretability and identifying important features.
- **Random Forest:** For handling complex interactions in data.
- **XGBoost:** A high-performance model for predictive tasks.

Step 3: Model Evaluation

- **Confusion Matrix:** To understand misclassification.
- **Precision-Recall Curve & AUC Score:** Due to class imbalance.
- **Feature Importance Analysis:** Using SHAP values for model interpretability.

6. Insights and Recommendations

Key Predictors of Churn

- **Decrease in Recharge Amount:** A significant drop in recharge over months is a strong indicator.
- **Reduction in Call Minutes:** If a customer reduces call duration, they might be moving to another network.
- **Decline in Data Usage:** Customers who decrease data usage over time are likely to churn.

Business Actions

- **Proactive Customer Retention:** Offer discounts or better plans to high-risk customers in the action phase.
- **Improve Service Quality:** Address complaints proactively to reduce churn triggers.
- **Loyalty Programs:** Incentivize long-term customers with exclusive benefit

7. Conclusion

This project successfully built a churn prediction model that enables telecom companies to take preemptive actions. By identifying at-risk customers early, telecom providers can improve retention strategies and reduce revenue leakage.