Case Study: Predicting Customer Churn in Subscription-Based Services

Introduction

Customer churn is a major challenge for subscription-based services. Losing customers not only affects revenue but also increases the cost of acquiring new ones. This case study aims to predict customer churn using data analysis and machine learning techniques, offering actionable insights to help companies retain their customers.

Objectives

1. Predict which customers are likely to churn.
2. Identify the key factors that influence customer churn.
3. Provide actionable recommendations to reduce churn rates.

Data Sources

To achieve our objectives, we will gather data from various sources:

1. **Customer Usage Data**:
    o Records of customer interactions with the service (e.g., login frequency, usage duration).
    o Transactional data (e.g., purchase history, subscription renewals).
2. **Demographic Information**:
    o Customer age, gender, and location.
    o Socioeconomic status and employment information.
3. **Customer Service Interaction Logs**:
    o Records of customer support tickets and their resolutions.
    o Customer feedback and satisfaction scores.
4. **Historical Churn Data**:
    o Historical records of customers who have churned.
    o Reasons for churn if available.

Data Understanding
Exploratory Data Analysis (EDA)

To understand our data, we will perform an initial exploratory data analysis (EDA). This involves:

- **Data Structure**: Understanding the types and distribution of data.
- **Key Variables**: Identifying significant variables related to customer behavior and churn.

Initial Insights

We will create visualizations and statistical summaries to uncover initial patterns and anomalies, providing a foundational understanding of the data.

Data Cleaning and Preprocessing

Cleaning and preparing the data is crucial for accurate analysis. Our steps include:

- **Handling Missing Values**: Imputing or removing missing data to ensure completeness.
- **Outlier Treatment**: Identifying and addressing outliers that could skew results.
- **Normalization/Standardization**: Scaling features to a standard range to ensure consistency.
- **Feature Engineering**: Creating new features from existing data, such as customer tenure and interaction frequency, to enhance model performance.

Data Analysis and Visualization

Behavioral Analysis

We will examine customer usage patterns and service interactions to identify trends that may signal churn risk.

Demographic Analysis

Assessing the impact of demographic factors, such as age and location, will help us understand which groups are more likely to churn.

Visualizations

- **Trend Analysis**: Visualizing churn rates over time.
- **Correlation Heatmaps**: Identifying relationships between variables.
- **Distribution Plots**: Understanding the distribution of key features.

Model Building and Evaluation

Data Splitting

We will divide the data into training and testing sets to validate our models.

Model Selection

We will build several predictive models, including:

- Logistic Regression
- Decision Trees
- Random Forests
- Gradient Boosting

Model Evaluation

Our models will be evaluated using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to ensure reliable performance.

Results Interpretation and Documentation

Model Interpretation

Analyzing the output and feature importance from our models will help us understand the key drivers of churn.

Key Findings

We will highlight significant insights and patterns discovered during our analysis.

Recommendations

Based on our findings, we will provide actionable strategies for reducing churn, such as targeted marketing campaigns or improving customer service for at-risk groups.

Presentation and Sharing

Report

We will create a comprehensive report detailing our methodology, analysis, and findings.

Presentation

Preparing a set of slides will help summarize our findings and insights for stakeholders.

LinkedIn Post

We will share the case study on LinkedIn, using relevant hashtags (#DataScience, #CustomerChurn, #SubscriptionServices) and tagging appropriate individuals and organizations to reach a broader audience.

Expected Outcomes

- Identification of key factors influencing customer churn.
- Development of a reliable predictive model for churn.
- Actionable insights and recommendations to help reduce churn rates.

Appendices

- **Data Dictionaries**
- **Modeling Code**
- **Additional Visualizations**