Project Synopsis

Name	Pavan Raj K G
USN	221VMTR02085
Elective	MCA, Data Science
Date of Submission	29-march-2024

Churn Prediction Model for E-commerce Customer Retention

Executive Summary

In today's competitive E-commerce landscape, retaining customers is critical. This project proposes a churn prediction model for an E-commerce company to identify customers at risk of leaving. By analyzing data like tenure, service scores, payment preferences, and complaint history, the model aims to generate actionable insights for targeted marketing campaigns and improve customer retention strategies.

Objectives & Scope

- Develop a churn prediction model with high accuracy to identify accounts vulnerable to churn within the E-commerce company.
- Provide business recommendations for targeted marketing campaigns focused on retaining potential churners.
- Analyze the impact of different features (e.g., tenure, service score) on churn likelihood to understand customer behavior and inform product/service improvements.
- Explore the possibility of segment-specific churn prediction models to provide even more granular insights and targeted retention strategies.

This project will encompass the following stages:

1. Data Collection and Preprocessing
2. Exploratory Data Analysis (EDA)
3. Feature Engineering
4. Model Development
5. Model Evaluation
6. Business Recommendations

By focusing on these steps, the project aims to deliver a comprehensive churn prediction solution tailored to the company's needs, enabling effective customer retention strategies.

Methodology

1. Data Collection and Preprocessing

- **Download Data**
- **Identify and Handle Missing Values**
- **Outlier Detection and Treatment**
- **Data Normalization**
- **Encoding Categorical Variables**

2. Exploratory Data Analysis (EDA)

- **Descriptive Statistics:** Summarize key characteristics of the data using measures like mean, median, standard deviation for numerical features and frequency tables for categorical features.
- **Data Visualization:** Create visualizations like histograms, boxplots, and scatterplots to explore distributions, identify outliers, and visualize relationships between features.
- **Customer Segmentation:** Grouping customers based on shared characteristics.

3. Feature Engineering

- **Deriving New Features:** Create new features from existing data that might be more informative for churn prediction.
- **Interaction Features:** Combine existing features to create new ones that capture specific relationships.
- **Feature Selection:** Choose a relevant subset of features to avoid overfitting and improve efficiency.

4. Model Development

- **Selection of Machine Learning Algorithms:** Choosing appropriate algorithms for churn prediction based on data characteristics and project goals.
    - **Logistic Regression:** A classic algorithm suited for binary classification problems like churn prediction. It estimates the probability of churn based on a linear combination of features.

- ○ **Decision Trees and Random Forest:** These tree-based algorithms create decision rules based on feature values to classify new data points. Random Forest combines multiple decision trees for improved accuracy and robustness.
- ○ **Gradient Boosting:** This ensemble technique builds sequential models where each new model learns to improve on the errors of the previous one. XGBoost is a popular implementation known for high performance.
- **Model Training and Tuning:** Split the data into training and testing sets. Train the chosen machine learning models on the training set, adjusting hyperparameters to optimize performance. Techniques like grid search or randomized search can be used for hyperparameter tuning.
- **Model Comparison:** Evaluate the performance of different models on the held-out testing set using metrics like accuracy, precision etc., Choosing the model with the best overall performance.

5. Model Evaluation

- **Evaluation Metrics:** Assess the model's effectiveness using various metrics:
  - ○ Accuracy: The overall percentage of correct predictions
  - ○ Precision: The proportion of predicted churners who actually churned
  - ○ etc.,

6. Business Recommendations

Based on the churn prediction model's output, generate actionable recommendations for targeted marketing campaigns focused on retaining high-risk customers. Here are some examples:

- Customer Segmentation and Tailored Offers
- Proactive Customer Service
- Cost-Benefit Analysis
- etc.,

Preliminary Findings & Expected Results

Through preliminary analysis and exploration of the dataset, several key findings have emerged, providing insights into the factors influencing customer churn within the E-commerce company. These findings serve as a foundation for understanding customer behavior and guiding the development of the churn prediction model.

Preliminary Findings:

- Initial data analysis revealed variations in churn rates among customer segments.
- Features like tenure, service score, and complaint history show significant association with churn.
- Newer customers and those with lower satisfaction scores exhibit higher churn risks.
- Payment preferences and complaint history show connections to churn likelihood.
- Specific customer segments, like "Regular Plus", have higher churn probabilities compared to others.

Expected Results:

- Development of an accurate churn prediction model with high predictive power to identify customers at risk of churn.
- Generation of actionable insights and recommendations for improved customer retention strategies like personalized offers, loyalty programs, and service enhancements.
- Enhanced revenue assurance and profitability through effective churn mitigation and customer retention.

Overall, this project aims to empower the E-commerce company with data-driven insights and predictive capabilities to effectively address customer churn, enhance customer loyalty, and drive sustainable business growth.