Predicting Customer Churn for Subscription Services

Jackson Baxter

Participation Interest 8/10

Project Description

The goal of this project is to forecast customer churn for subscription-based services including membership-based apps, SaaS solutions, and streaming platforms. Because it's usually less expensive to keep existing customers than to get new ones, churn prediction is important for organizations. Creating a system that recognizes users who are most likely to terminate their subscriptions within a given time frame is the aim of this project. Businesses can tailor retention measures, such discounts or special offers, by anticipating these users' needs. Sorting clients into two groups—those who have churned or those who have not—based on their interactions with the service as well as additional behavioral and demographic characteristics is the task at hand.

What features the data set might include

Data on user interactions, such as frequency of use, amount of time spent on the platform, number of logins, kind of subscription plan, and exchanges with customer care, may be included in the dataset. Demographic information such as age, gender, location, and payment history (missed payments, mode of payment, etc.) could be added as extra features. Whether or not the consumer churned, or canceled their membership, usually within the next 30 or 60 days, would be the target variable. The dataset would consist of a mix of categorical (like payment method), binary (like churned or not), and continuous (like usage duration) elements.

Example of a data instance:

Usage Time (hrs/month)	Logins (per month)	Subscription Plan	Payment Method	Customer Support Calls	Age	Gender	Chu (tar
12	5	Premium	Credit Card	2	28	F	Yes

How and from where would the data set be gathered and labeled

The dataset might be obtained by utilizing publicly accessible datasets from open sources like Kaggle or research datasets on customer attrition, or it could be gathered from subscription service providers willing to share anonymized user data. By scraping or merging data with other services that offer demographic information based on geography or economic variables, data could be improved. Preprocessing the data, filling in missing information, managing unbalanced data (because the percentage of churning users may be much lower than that of non-churners), and feature engineering will be the main challenges. Churn labels are derived from past data, designating a user who cancels their subscription as "churned." The goal of this research is to determine how various demographics and behaviors connect to a user's propensity to discontinue service and how this information can be used to increase customer retention.