Customer Churn Prediction using Data Mining Techniques

Milestone: Initial Project Proposal

Group 16

Amitesh Tripathi
Sayali Lad

857 376 1991(Tel:Student 1)
857 277 4326(Tel:Student 2)

tripathi.am@northeastern.edu
lad.sa@northeastern.edu

Percentage of Effort Contributed by Student 1: 50 %

Percentage of Effort Contributed by Student 2: 50 %

Signature of Student 1: Amitesh Tripathi

Signature of Student 2: Sayali Lad

Submission Date: 01/27/2023

PROJECT PROPOSAL TITLE : "Customer Churn Prediction Using Data Mining Techniques"IE 7275: DATA MINING IN ENGINEERING

Problem Setting:

Customer churn is a major concern in the business world because it can have a significant influence on a company's revenue. Churn refers to the loss of customers over a specific time period. Businesses must anticipate and avoid customer behavior in order to retain and expand their base of customers. The challenges that can be encountered in developing an effective customer churn prediction model include:

- Erroneous or untidy customer data
- Data imbalance
- The selection of KPIs to assess churn model performance
- The selection of a suitable modeling approach

Problem Definition:

The purpose of this project is to create a machine learning model that can forecast customer attrition for a certain organization. The model will be trained on historical customer data to discover patterns and factors that contribute to churn. The following are the specific questions that this initiative seeks to address:

What are the primary causes of client churn?

Can a machine learning model reliably forecast customer churn?What strategies may be used to reduce client churn?

Data Source:

The data "Bank Customers Data for Churn" for this project is obtained from Kaggle website. It will include information such as customer demographics, purchase history, and engagement with the Bank's products and services.

https://www.kaggle.com/datasets/adammaus/predicting-churn-for-bank-customers

Data Description:

The dataset consists of customer information such as customer ID, age, gender, income, location, and tenure with the company. It includes information on the customer's purchase history and engagement with the company, such as the number of purchases made, the amount spent, and the customer's level of engagement with the company's marketing campaigns. The dataset has approximately 10,000 rows and 20 columns. A sample of variable names include "customer_id", "age", "gender", "income", "location", "tenure", "no_of_products", “churn”.