customer Churn Prediction

-Project

Contents

# Abstract

# Churn Prediction

Churn prediction is common use case in machine learning domain. If you are not familiar with the term, churn means "leaving the company". It is very critical for business to have an idea about why and when customers are likely to churn. Having a robust and accurate churn prediction model helps businesses to take actions to prevent customers from leaving the company.

In this project, I will use "Telco Customer Churn" data set which is available on Kaggle.

There are 20 featuures (independent variables) and 1 target (dependent) variable for 7043 customers. Target variable indicates if a customer has has left the company (i.e. churn=yes) within the last month. Since the target variable has two states (yes/no or 1/0), this is a binary classification problem.

The variables are: customerID: Customer ID gender: Gender of customer SeniorCitizen: Whether the customer is a senior citizen or not (1, 0) Partner: Whether the customer has a partner or not (Yes, No) Dependents: Whether the customer has dependents or not (Yes, No) tenure: Number of months the customer has stayed with the company PhoneService: Whether the customer has a phone service or not (Yes, No) MultipleLines: Whether the customer has multiple lines or not (Yes, No, No phone service) InternetService: Customer’s internet service provider (DSL, Fiber optic, No) OnlineSecurity: Whether the customer has online security or not (Yes, No, No internet service) OnlineBackup: Whether the customer has online backup or not (Yes, No, No internet service) DeviceProtection: Whether the customer has device protection or not (Yes, No, No internet service) TechSupport: Whether the customer has tech support or not (Yes, No, No internet service) StreamingTV: Whether the customer has streaming TV or not (Yes, No, No internet service) StreamingMovies: Whether the customer has streaming movies or not (Yes, No, No internet service) Contract: The contract term of the customer (Month-to-month, One year, Two year) PaperlessBilling: Whether the customer has paperless billing or not (Yes, No) PaymentMethod: The customer’s payment method (Electronic check, Mailed check, Company transfer (automatic), Credit card (automatic)) MonthlyCharges: The amount charged to the customer monthly TotalCharges: The total amount charged to the customer Churn: Whether the customer churned or not (Yes or No)

At first glance, only customerID seems irrelevant to customer churn. Other variables may or may not have an effect on customer churn.

# What has been Provided by the customer (company):

* The customers dataset with historical data.
* From the column [Exited], we can identify the customers who exited from the company and who are continuing the services of company. The value 1 represents the customers who are exited from the company and 0 represents the customers who are continuing with the company.

# Why this problem is important:

* Using the solution to this problem, the company can easily identify the customers who are willing to exit the company soon.
* From the larger datasets, the company can easily identify the churn customers using machine learning approach, thus this can reduce the manual intervention and the cost to the company. Using machine learning solutions, the company can save processing time and manual intervention to investigate the complete records. The system can take quicker decisions with statistical models with optimal accuracy metrics.
* With the investigation of the customers who are churning soon, the company has an option to reduce the churn by further investigating the reason of leaving the company and to convince the customers by providing or improvising the services rendered to them.

# Common business applications:

* The churn model can be integrated with the call center / business software, so that the proper discounting can be provided to the identified customers. Targeted marketing strategies can be used.
* Monitoring of the customer trends and building the alerting mechanism to the business users on a daily / monthly basis.
* This problem can be applicable to any industry (for e.g., Telecom) to identify the churn customers within their organizations.

**Problem**:The classification model to be built on historical data and then used to predict the classes for the current customers to identify the churn.

All the Best!