Churn Prediction in Telecom using Machine Learning

Estimating churners before they discontinue using a product or service is extremely important. In this ML project, you will develop a churn prediction model in telecom to predict customers who are most likely subject to churn.

Project template outcomes

- Understanding the business problem.

- Understanding what churn is, why to solve it and how to solve it.

- Importing the dataset and required libraries.

- Performing basic Exploratory Data Analysis (EDA).

- Data cleaning and missing data handling if required, using appropriate methods.

- Checking for class distributions

- Outlier treatment

- Splitting dataset as per requirement

- Sampling the dataset

- Perform Feature Engineering

- Building model using all ML algorithms

- Understanding different performance metrics like Confusion matrix, AUC, Recall, F1-score

- Building feature importance

- Analysing the correlation plot and gaining insights

- Model refinement and implementation

Description

Business Objective

In any service providing industry, when a customer decides to stop using the service either by cancelling the subscription or not paying for the service, we call this customer churn.

Churn is defined as how many customers are not using the service for a certain period.

Hence, customer churn is one of the essential metrics that every business must evaluate to grow. The churn rate is calculated by dividing the number of lost customers by the last number of customers. Thus, a company churn rate must be as low as possible, ideally 0%.

But why is it so important to calculate the churn rate? Does it affect the business if you lose around 5% of customers? Yes, the answer is that it costs more to acquire a new customer than retain the existing customers. Retaining the current customers, any company can spend less on operating costs needed to reach new customers.

So, we will use advanced machine learning techniques to predict the potential churners who are about to leave a company's service and take the necessary steps to prevent it.

This project aims to build a deep learning model that will help predict customers who are likely to churn in the next N months and facilitate in taking business actions for reducing the churn.

Data Description

The available dataset is:

Telco-Customer-Churn – This dataset has 7043 rows and 21 columns present.

The 21 features of this dataset are as follows:

1. Churn – the target variable, if the customer is churned or not (Yes / No)
2. customerID – The unique identification of every customer
3. gender- If the customer is a male or a female (Female / Male)
4. SeniorCitizen – If the customer is a senior citizen or not (0 / 1)
5. Partner – If the customer has a partner or not (Yes/No)
6. Dependents – If the customer has any dependents (Yes / No)
7. Tenure – The time period(months) the customer has stayed with the company.
8. PhoneService – If the customer has a phone service or not (Yes/No)
9. MultipleLines – If the customer has multiple lines or not (Yes/No/No Phone service)
10. InternetService – If the customer has any internet service or not (DSL/ Fibre optics/ No)
11. OnlineSecurity – If the customer has any online security (Yes/No/No internet service)
12. OnlineBackup – If the customer has any online backup (Yes/No/No internet service)
13. DeviceProtection – If the customer has device protection (Yes/No/No internet service)
14. TechSupport – If the customer has tech support (Yes/ No/ No internet service)
15. StreamingTV – If the customer has any streaming TV (Yes/ No/ No internet service)
16. StreamingMovies – If the customer has streaming movies (Yes/ No/ No internet service)
17. Contract – The customer term period with the company (Month-to-month, One year, Two years)
18. PaperlessBilling – If the customer has paperless billing or not (Yes/ No)
19. PaymentMethod – The payment mode of each customer (Electronic check, mailed check,Bank transfer, Credit card)
20. MonthlyCharges – The amount that is charged to the customer every month
21. TotalCharges – The total amount charged to the customer

Aim

To predict the customers who are likely to churn in the next N months & facilitate in taking business actions for reducing the churn