

Churned Prediction System for Telecommunications Industry

Dhruv Dasadia (A20536411), Dharmik Patel (A20526771)

1 April, 2024

1 Introduction

As the competition has become more intense, it has become more crucial to retain the existing customers then onboarding the new ones. As the demand of customer increases, so to meet the requirements the service provider makes innovative strategies to lower the customer churn. Through Machine Learning Algorithmic Models, it can analyze and visualize datasets to find the patterns which can be helpful, machine learning models can predict which customers is about to leave. By applying the training and validation process, aim to build a model which can help the telecom's o reduce the churn.

Churn Customers are the numbers of existing customers who may leave the current service provider in a span of time and likely to join other service provider. The main goal of churn is to predict the customers who are likely to churn earliest and to identify the reason for churning. By doing so this will rectify the problems faced by the customers.

The Objective of this project is to enhance the productivity of Machine Learning Algorithms to create a Churn Prediction System which identifies which customer are about to leave. The project will not be only focused on customers which are about to leave but will also be aimed at improving the retention rates and overall customer satisfactions. Through a diligent process of training and validation, the project aims to develop a model which help the telecom companies to decrease the churn. These type of models help telecom industry by making them profitable.

2 Dataset

The dataset used for Training the model contains of Telco Customer Churn dataset which is obtained from Kaggle website, which is of IBM. Each row in the dataset represent the customers and columns describes the attributes, each one of them has features. One of attribute contains two classes in which it is represented by 'T' which indicates that these customers are churning and rest of them as 'F' which are likely to stay with the service provider. The dataset comprises of 16 categorical and 5 numerical columns.

The dataset contains many independent variables which can be classified into three categories.

- Demographic Information
 - Gender
 - Senior Citizen
 - Partner
 - Dependents
- Account Information
 - Tenure

```
import pandas as pd
import geopandas as gpd
import numpy as np
df = pd.read_csv('/content/tabular-actgan-6600dbcd4ba7a54b31c0355a-data_preview.csv')

df.columns

Index(['CustomerID', 'Count', 'Country', 'State', 'City', 'Zip Code',
      'Lat Long', 'Latitude', 'Longitude', 'Gender', 'Senior Citizen',
      'Partner', 'Dependents', 'Tenure Months', 'Phone Service',
      'Multiple Lines', 'Internet Service', 'Online Security',
      'Online Backup', 'Device Protection', 'Tech Support', 'Streaming TV',
      'Streaming Movies', 'Contract', 'Paperless Billing', 'Payment Method',
      'Monthly Charges', 'Total Charges', 'Churn Label', 'Churn Value',
      'Churn Score', 'CLTV', 'Churn Reason'],
      dtype='object')
```

Figure 1: Dataset Columns

- Contract
- Payment Method
- Paperless Billing
- Service Information
 - Phone Service
 - Device Protection
 - Tech Support
 - Online Backups
- Above are the few one's which are put into the categories.

3 Methodologies

- **Data Processing:** In this phase, Handling of raw data, and implementing the methods for cleaning, normalization and transformation. Basically it covers the techniques for missing values, outlier identification and normalization to standardize the data, mitigating it for analysis. By doing so a high-quality of dataset which represents the attributes useful for predictive accuracy of the model.
- **Exploratory Data Analysis:** EDA is responsible for gaining insights from dataset's structure and identifying the patterns of churning the customers. During this it will highlight on the summaries, analysis and visualization methods to find the relationship.
- **Feature Selection:** It mainly focuses on identifying the attributes or variables which contributes on predicting the customers churn. It delves into the application of both filter and wrapper methods through assisting the impact on complexity of model.
- **Model Developing:** Main integral part of the project is developing the model which will predict the customers likely to churn, using the specific machine learning algorithms. It focuses on training process which includes techniques such as partitioning of data, cross validation and tuning of parameters, through which it optimizes the model performance.
- **Evaluation:** This module is for measuring the performance of the churn prediction model, by analysis of performance metrics. The evaluation extends beyond the accuracy to consider other metrics which offers a comprehensive view of model performance.
- **Validation:** Validation involves testing the model on a distinct dataset from the one used during training. These process measures the model's performance to ensure it's reliability and its ability to generalize to unseen data. It is designed to prevent the overfitting and to ensure model's robustness.
- **Comparison of Results:** An analysis of various model based on the performance metrics which highlights on Strengths and weakness of each model. The most effective model for prediction can be chose based on the evaluation.
- **Conclusion:** Concluding, our model predicting ability to identify customers at risk of leaving. Based on the findings the customer retention strategies can be improved and put into the consideration. It suggests the improvements in churn prediction methods.

4 Milestone

Below are the Milestones for the project.

1. Data Preprocessing

- **Data Extraction:** First Step where data is gathered and loading the necessary data into a single format to present a comprehensive view.
- **Data Augmentation:** Increasing the diversity of data by generating synthetic data or creating new features to enhance the robustness of Model.
- **Data Balancing:** To prevent the bias in the Model's Prediction, which in general favors the majority class. So techniques such as oversampling the minority or undersampling the majority class.

2. Exploratory Data Analysis

- **Data Exploration with Visual:** Creating visualizations to understand the dataset, it is done by identifying the patterns, detecting the outliers
- **Finding Correlation:** Plotting the correlation matrix and calculating the Coefficients and finding the relationship between the variable and identify which one has the strong relationship with customer churn.
- **Feature Extraction:** Developing new features on the context of the existing one which can help the model's ability to predict outcomes.
- **Preparing Final Dataset split for Modelling:** This step involves splitting the dataset into training, testing and validation. By doing so it helps the model to tune the parameters and evaluate the model's performance to determine how well it generalizes to unseen data.

3. Modelling

- **Algorithm Selection:** Selecting the right ones which are best fitted for the model, which provides good interpretability and performance.
- **Ensembling Multiple Models:** Integrating multiple models which significantly boost precision and reliability. Techniques are used to deal with the underfitting or overfitting by changing the variance and bias.
- **Determining the best Model:** Once the development and integrating the models, next step is to assess their performance through the evaluation metrics.

4. Model Evaluation

- Evaluation is not only judged on the overall accuracy of the model, but also how well it identifies True Positives and True Negatives; customers who are going to churn and the retained customers.

5. Optimization

- Adjusting or Tuning the model's parameters to improve the accuracy of model on unseen data and generalizes it well by preventing the overfitting of data, making the model reliable.

6. Conclusion

- It implicates all the steps from Data Preprocessing to Evaluation of Model by stating the predictive performance and enhancing the Customer Retention Strategies.

5 Future Work

Based on our current progress, we have completed the Exploratory Data Analysis phase. Next steps is to begin with the Preparing Final Dataset split for Modelling, which marks the final step of EDA process.

Further we will proceed with Modelling phase, we will start from Algorithm selection and select the most appropriate machine learning algorithm.

Finally, once the entire process has been completed we will get understanding of model's ability to predict and provide the insights which customers are likely to churn.

6 References

- 1 SharmilaK.Wagh,AishwaryaA.Andhale,KishorS.Wagh, Jayshree R. Pansare, Sarita P. Ambadekar, S.H. Gawande, Customer churn prediction in telecom sector using machine learning techniques.
- 2 AbhishekGaur,RatneshDubey,PredictingCustomerChurn Prediction in Telecom Sector Using Various Machine Learning Techniques.
- 3 Kaggle Dataset Link