DB6

by Vamshikrishna Namani

Submission date: 09-Mar-2023 03:44PM (UTC+1000)
Submission ID: 2032773460
File name: plagiarism_1.docx (461.7K)
Word count: 1503
Character count: 8398

Telecom Cutomer Churn Prediction

Abstract—**The wide variety of service providers are being elevated very swiftly in each business enterprise. For service providers, a rapidly expanding market in each location is leading to a larger subscriber base. Customer acquisition costs are rising as a result of increased competition, new and innovative business models, and better-suited goods. Service providers have learned the value of keeping existing clients happy in such a short period of time. Because of this, it is critical that service providers reduce churn, the occurrence when consumers of a business stop buying from or engaging with the business.**

Keywords—customer churn prediction, machine learning, random forest, support vector machine

I. Introduction

Telecom customer churn prediction is a complex process that involves the use of advanced data analytics and machine learning techniques. In order to build an effective churn prediction model, telecom companies must collect and analyze a large amount of data on customer behavior and interactions. [1]

One of the key challenges in telecom customer churn prediction is identifying the right variables to include in the model. There are many different factors that can influence whether a customer decides to churn, and companies must carefully consider which variables are most relevant for their particular customer base.

Another challenge is ensuring that the model is accurate and reliable. In order to achieve this, companies must use high-quality data and employ robust modeling techniques that are capable of handling complex, non-linear relationships between variables.

Despite these challenges, telecom customer churn prediction can provide significant benefits for companies that are able to implement it effectively. By identifying at-risk customers and taking targeted actions to retain them, companies can improve customer retention rates and reduce churn, which can lead to increased revenue and profitability.In addition, churn prediction can also help companies to better understand their customers and their needs, which can inform future product development and marketing efforts..[2] [3]

To predict customer churn, telecom companies often use data analytics and machine learning techniques on a variety of data sources, including customer usage data, billing information, and customer service interactions. By analyzing this data, companies can identify patterns and behaviors that are indicative of customers who are at risk of churning. Some of the key factors that are often considered in churn prediction models include customer demographics, usage patterns, payment history, and customer service interactions.

Once customers at risk of churning have been identified, telecom companies can take targeted actions to retain these customers. These actions might include offering discounts or promotions, improving customer service, or providing additional features or services that are tailored to the customer's needs.

Churn prediction can also be used to forecast future revenue and predict customer lifetime value. By analyzing past customer behavior, telecom companies can develop models that predict how much revenue they are likely to generate from a particular customer over their lifetime, and use this information to make decisions about how much to invest in customer retention efforts.

Overall, telecom customer churn prediction is an important tool for telecom companies looking to improve customer retention and reduce churn. By identifying at-risk customers and taking targeted actions to retain them, companies can improve customer satisfaction and build a stronger, more loyal customer base.

II. Data Preprocessing

The dataset contains 7043 observations and 21 variables extracted from a telecom company's customer database. The dataset is provided in CSV format, which is a common data format for storing tabular data.

Preprocessing helps us in machine learning in several ways:

- Handling missing values: Preprocessing can help to deal with missing values in the data. This can involve replacing missing values with an appropriate value, such as the mean or median of the variable, or removing the observation altogether if the missing data is too extensive.

- Data cleaning: Preprocessing can help to identify and correct errors and inconsistencies in the data. This might involve removing outliers, correcting data entry errors, or dealing with duplicates in the dataset.

- Scaling and normalization: Preprocessing can help to scale and normalize the data, which can be important for some machine learning algorithms.

III. PROPOSED WORK

Performing customer churn analysis using Python involves several steps. Here is an overview of the process:

Load the dataset: First, you need to load the customer churn dataset into Python. You can use libraries such as Pandas to read in the dataset from a CSV file or other format.

Explore the dataset: Once you have loaded the dataset, you need to explore the data to gain insights and identify any potential issues. This might involve looking at summary statistics, visualizing the data using graphs and charts, and checking for missing values or outliers.

Preprocess the data: After exploring the data, you need to preprocess it to prepare it for analysis. This might involve handling missing values, cleaning the data, and scaling or normalizing the variables.

Build the machine learning model: This might involve using algorithms such as logistic regression, decision trees, or random forests.[4] [5]

Evaluate the model: After building the machine learning model, you need to evaluate its performance using metrics such as accuracy, precision, and recall.

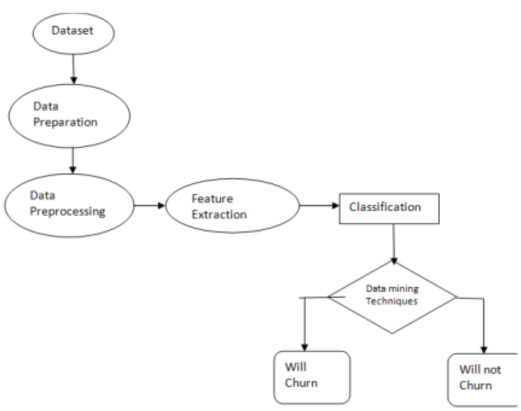

Fig 1. Churn Prediction Process

Use the model to make predictions: Once you have evaluated the model, you can use it to make predictions on new data.

1. Dataset: The dataset we have is from a telecom company and is in CSV format. The purpose of this dataset is to predict customer churn, which means identifying the customers who are most likely to leave the company. The dataset consists of 7043 observations or records, each with 21 variables or columns of data. These variables are chosen to help us understand trends in customer churn, such as demographics, usage patterns, and account details. With this dataset, we can use machine learning techniques to build predictive models that can help the telecom company to retain customers and reduce churn.

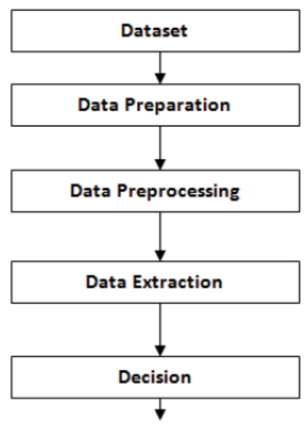

Fig 2. Analysis Steps

2. Data Preparation: Before we can apply churn prediction models to the telecom dataset, we need to prepare the data by naming each attribute. This means assigning meaningful names to the columns of data in the dataset. By doing so, we can ensure that each variable or attribute is properly identified and understood. This step is important because the raw data may contain ambiguous or unclear column names that can lead to errors in the analysis. By renaming the attributes, we can make the dataset more interpretable and easier to work with, which will ultimately improve the accuracy of our churn prediction models.

3. Data preprocessing: Data preprocessing is a crucial step in building accurate prediction models because the raw data often contains ambiguities, errors, redundancy, and transformations that can affect the quality of the analysis. By properly preprocessing the data, we can reduce errors and biases in the analysis and increase the accuracy of the prediction models. Therefore, data preprocessing is a critical step in building effective prediction models, and it

is essential to ensure that the data is of high quality before applying machine learning algorithms.

4. Data Extraction: In the churn prediction process, attributes or features are identified from the telecom dataset to be used for the classification process. These attributes can include demographic information such as age, gender, and marital status, as well as usage patterns such as call duration, internet usage, and account details such as tenure, contract type, and payment method.

5. Decision: The classification model is trained on a labeled dataset, where the churners and non-churners are identified based on historical data. The model uses the extracted attributes to learn patterns that differentiate between the two groups. Once the model is trained, it can be used to predict the churn probability for new customers based on their attributes. If the churn probability is high, the customer is identified as a potential churner, and the telecom company can take proactive steps to retain them, such as offering incentives or promotions. [6][7]

IV. Result Analysis

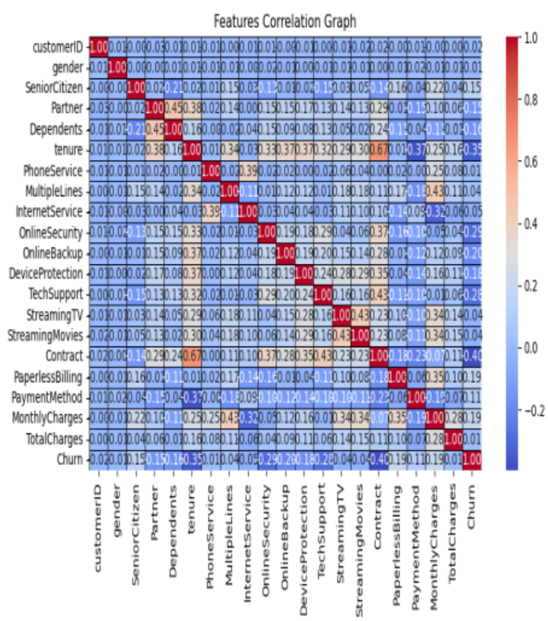

Fig 3. Correlation graph between different attributes

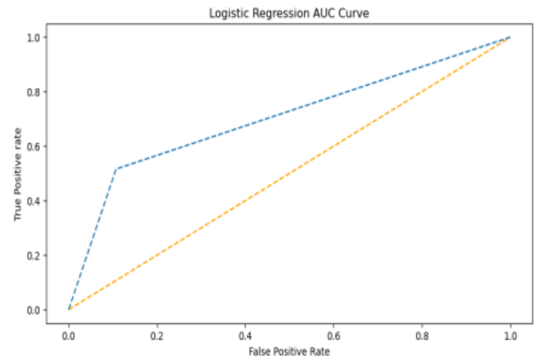

Fig 4. Logistic Regression AUC Curve

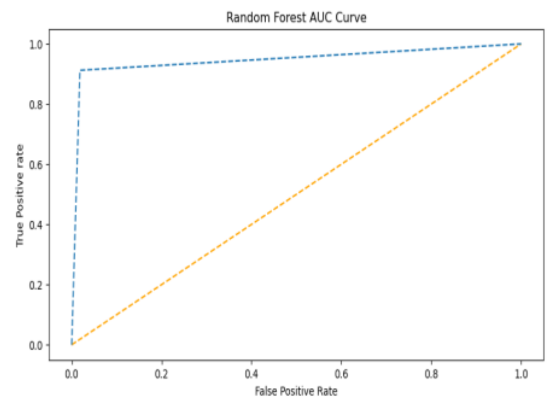

Fig 4. Random Forest AUC Curve

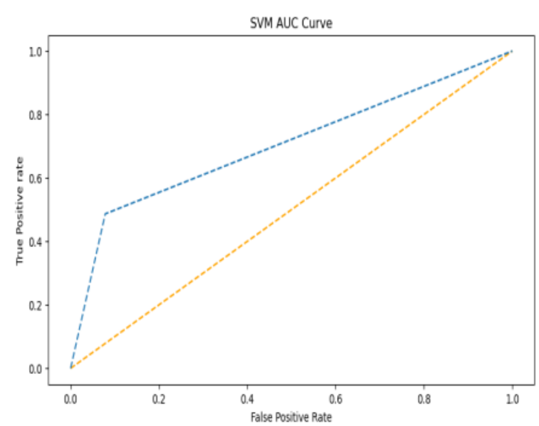

Fig 4. SVM AUC Curve

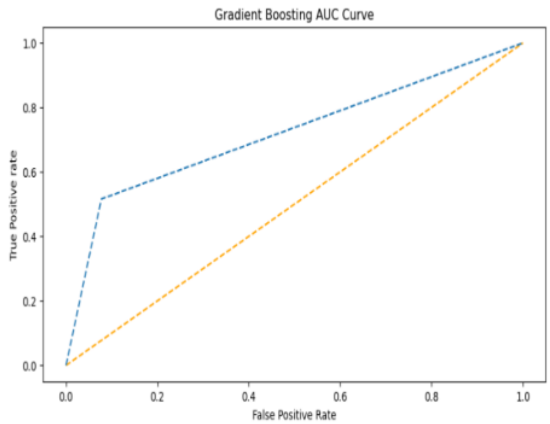

Fig 4. Gradient Boosting AUC Curve

V. CONCLUSION

As businesses continue to expand and offer more services to their customers, predicting customer turnover becomes increasingly challenging. In recent years, the telecom industry has been particularly focused on predicting customer churn, and machine learning algorithms have become an essential tool for this purpose. This study aimed to analyze the effectiveness of various classifiers in predicting customer churn using a publicly available telecom customer turnover dataset. The top three models based on experimental results were identified as Logistic Regression, Xgboost classifier, and Random Forest Classifier, SVM .[8]These models demonstrated higher accuracy, precision, recall, and AUC scores than other models.[9] While these models are effective, it is anticipated that deep learning methods will continue to improve in the future, leading to even greater success rates in predicting customer churn. Overall, the use of machine learning algorithms is essential for businesse to remain competitive and reduce customer turnover rates in the telecom industry.

DB6

ORIGINALITY REPORT

3%	1%	0%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University College London Student Paper	1%
2	Submitted to Oklahoma State University Student Paper	1%
3	www.ijirset.com Internet Source	1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		