

Telecom Churn Prediction

Submitted

by

Abhay Pratap Singh - 20113004

Mukund Shah - 20121014

Shiv Natani - 20115133

Shubham Jha - 20115140

Yash Gupta - 20117144

Supervisor

Prof. Sumit Kumar Yadav



Department of Management Studies
Indian Institute of Technology Roorkee

2023

Introduction

The telecommunications sector has become one of the leading industries in developed countries. The technical progress and the increasing number of operators raised the level of competition. Companies work hard to survive in this competitive market depending on multiple strategies. Three main approaches have been proposed to generate more revenues: acquire new customers, upsell existing customers, and increase the retention period of customers. However, comparing these strategies taking the value of return on investment (RoI) of each into account has shown that the third strategy is the most profitable strategy, proving that retaining an existing customer costs much lower than acquiring a new one, in addition to being considered much easier than the upselling strategy. To implement the third strategy companies need to reduce the customer's churn. If done early on, identifying the clients who are most likely to depart the business could represent a sizable extra revenue source. Whole data is divided into training and test data, training data is used for training purposes, and thereafter for checking the actual performance, we use test data.

Data Collection:

The Data for this project is used from Kaggle Customer Churn Prediction 2020 Competition.

Here is the link for the same:-

<https://www.kaggle.com/competitions/customer-churn-prediction-2020/data>

The dataset is available in .csv format. It has 19 features and one boolean variable to predict the class of the sample. The dataset is labeled perfectly and we have used PCA to reduce its dimensionality and to reduce the effect of noise in the data, if any.

Algorithms used:

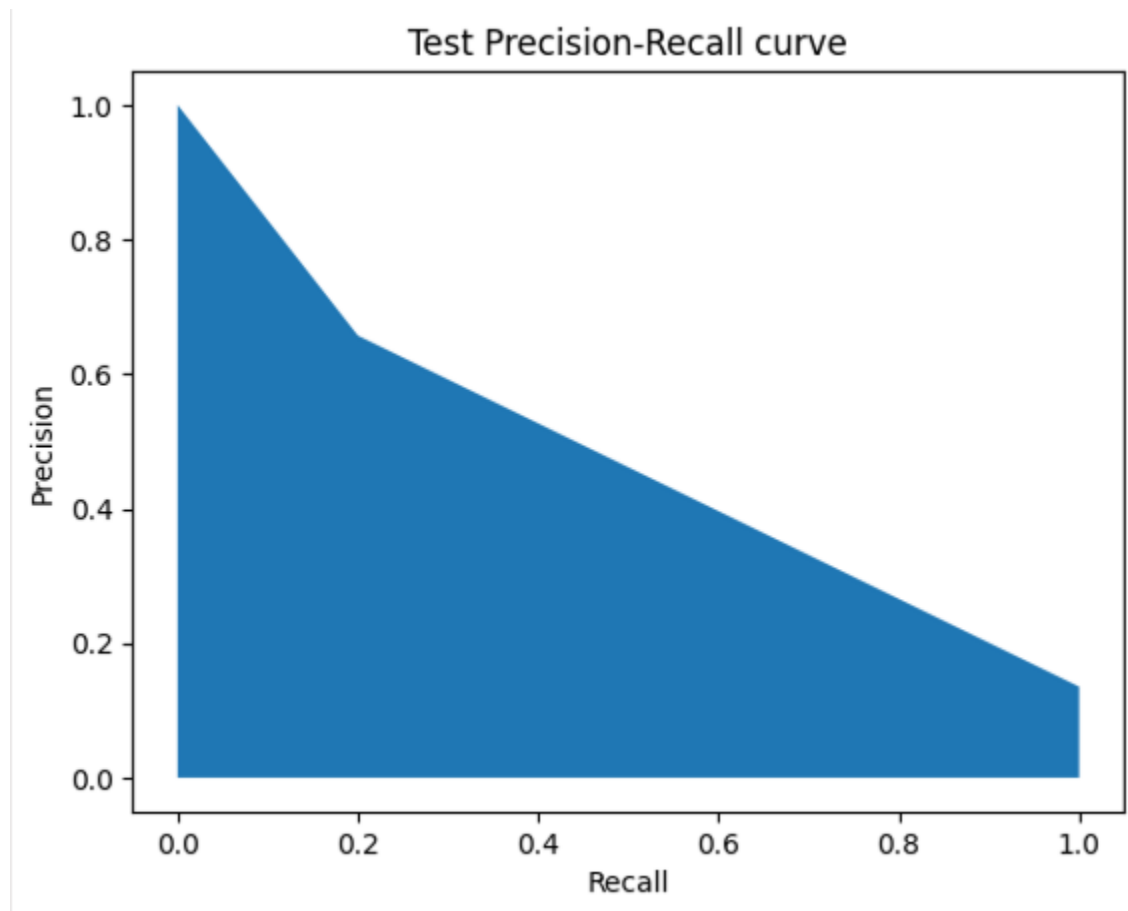
We used two major Machine Learning algorithms in this project which are PCA(principal component analysis) and Logistic Regression.

Principal Component Analysis is an unsupervised learning algorithm that is used for dimensionality reduction in machine learning. It is a statistical process that converts the observations of correlated features into a set of linearly uncorrelated features with the help of orthogonal transformation. These new transformed features are called the Principal Components. We used PCA to reduce the dimension and remove irrelevant features.

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, True or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1. We used logistic regression after transforming the data using PCA to predict the output as 0 or 1.

Results:

The trained model's accuracy for test data is 87.765%, and for those data, there were 723 true positives, 92 false positives, 12 false negatives, and 23 false negatives. The precision-recall curve is also shown in the graph below.



Discussion:

Customer churn prediction is a very important task in the business world and is a very crucial factor for the growth and success of many

companies like telecommunication, banking, etc. In this project, using ML algorithms, we found possible customers who can change their minds and can switch to some other company for service.

The main objective of this project is to identify at-risk customers and reduce churning rates by taking appropriate actions. If you want to grow as a company, you have to invest in acquiring new clients. Every time a client leaves, it represents a significant investment loss. Both time and effort need to be channeled into replacing them. Being able to predict when a client is likely to leave and offer them incentives to stay, can offer huge savings to a business. By identifying at-risk customers they can be targeted with retention strategies and improved customer service. By identifying and addressing their complaints, companies can reduce churn rates by improving their satisfaction.

Overall, the results of customer churn prediction can provide valuable insights and information for businesses to proactively address customer churn and improve customer retention. By leveraging these insights, companies can optimize their marketing efforts, reduce churn rates, and ultimately result in improved business performance and profitability.

Conclusion:

The purpose of this kind of study in the telecom industry is to assist businesses in increasing their profits. It is well known that one of the most significant revenue streams for telecom firms is churn prediction. Therefore, the goal of this research was to develop a system that could forecast customer attrition in the telecom industry. The sample data is split into 80% for training and 20% for testing in order to test and train

the model. We have transformed the data using PCA to apply machine learning algorithm logistic regression on it. The accuracy we reached is 87.765%.