Churn No More!

Churn Prediction in Telecommunication Using machine learning

Anwar Hussain Sofi

CONTENTS

INTRODUCTION

Telecom Customer Churn Prediction

Telecom customer churn prediction is an important topic in the industry as it can help companies retain customers and increase revenue. According to a study by Accenture, telecom companies lose $756 billion annually due to customer churn. By accurately predicting which customers are likely to churn, companies can take proactive measures to prevent it from happening.

One of the main reasons for customer churn in the telecom industry it's because poor customer service. Other factors include pricing, network quality, and competition. It is important for companies to identify these factors and address them in order to retain customers.

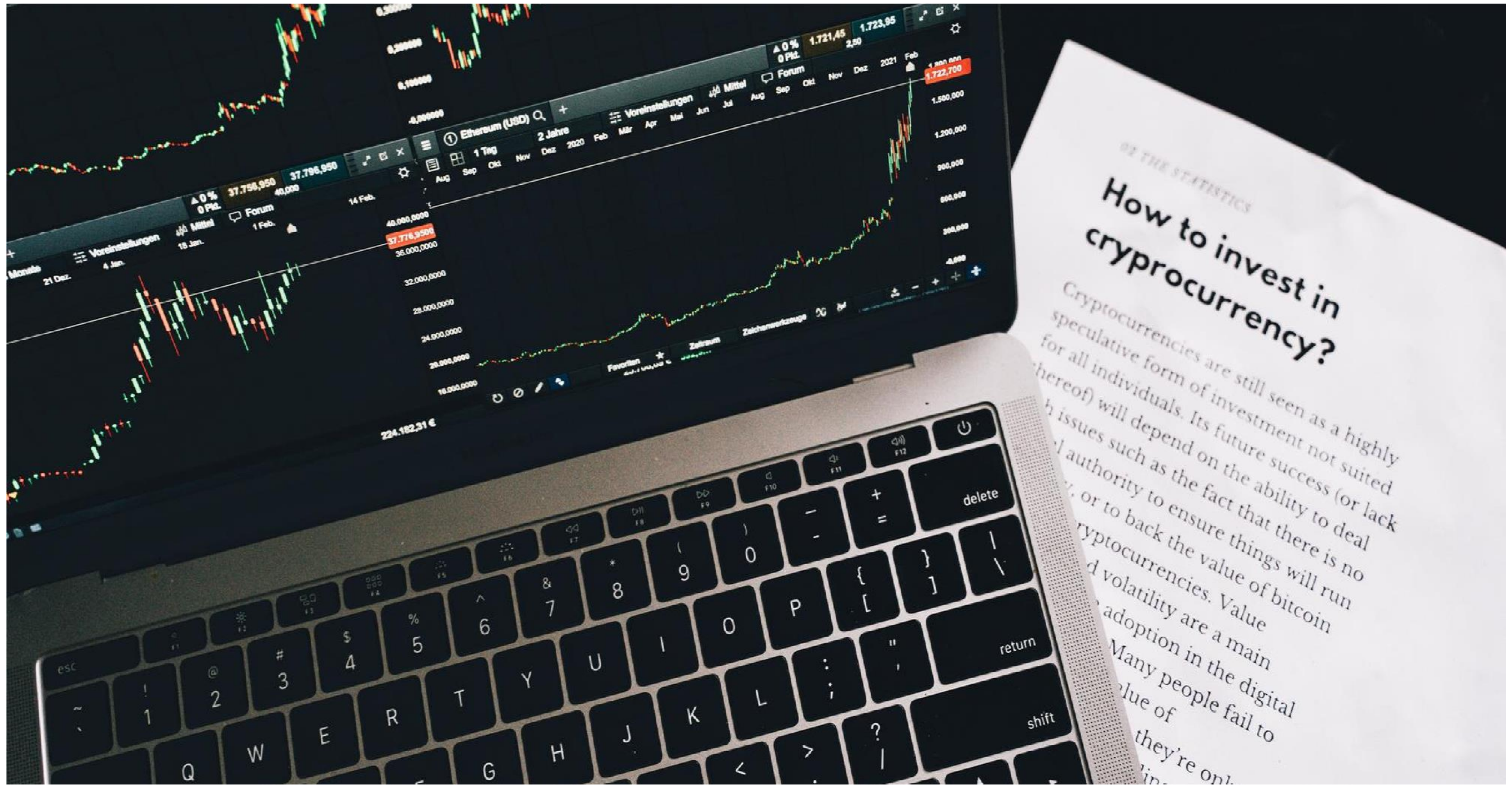
What is Customer Churn?

Customer churn refers to the rate at which customers stop doing business with a company. In the telecom industry, customer churn is a concern because it can lead to a loss of revenue and market share. When customers switch to another competitor, they take their revenue with them and reduce the company's market share.

Since keeping existing customer’s costs less than recruiting new ones, customer churn is a major problem for telecom firms. We wanted to create a reliable predictive model for telecom customer attrition with this project. We developed a model that can detect customers who are likely to churn, enabling the business to take preventative efforts to retain them.

About the data

The data used to develop a churn prediction model is the "WA_Fn-UseC_-Telco-Customer-Churn.csv" dataset from Kaggle. This dataset provides valuable information about telecom customer behavior, including call duration, network usage, and customer demographics. The data is collected from various sources such as customer interactions, billing records, and network performance metrics.



The "Telecom Customer Churn" dataset includes details on telecom customers' characteristics. Every customer is represented by a row, and every attribute or quality of each customer is represented by a column. The dataset has 21 columns that reflect distinct attributes and 7043 rows that represent various clients. This dataset's primary goal is to forecast client attrition, as seen by the "Churn" column. Customers who have stopped doing business with the telecom firm are referred to as churning. The target variable for predicting customer attrition is the "Churn" column.

Implementat ionmethode

K-Nearest Neighbors (KNN)

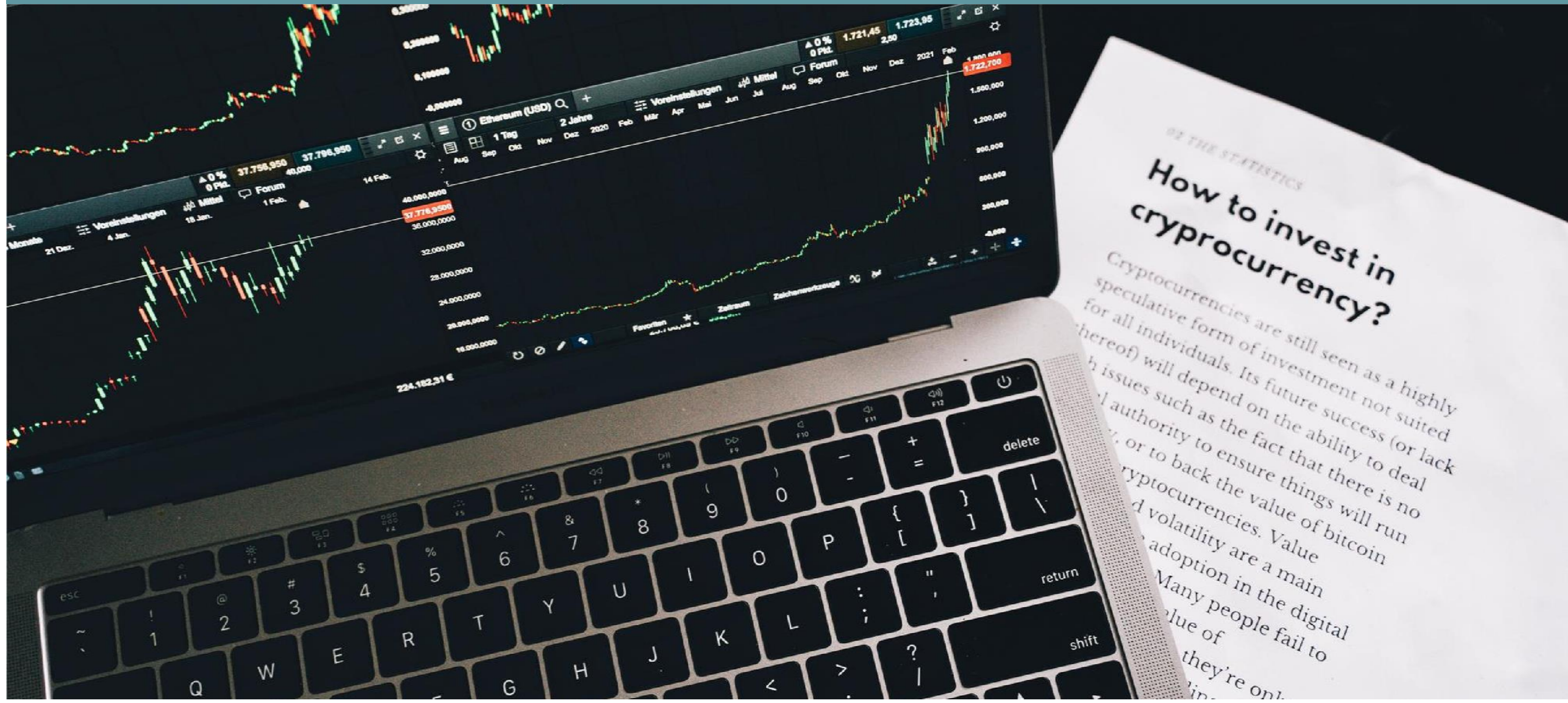
The KNN algorithm is a machine learning algorithm that can be used for customer churn prediction. It works by finding the k-nearest neighbors of a given data point and using their class labels to predict the class label of the new data point.

Decision Tree Classifier

creates a collection of decision trees, where each tree is trained on a different subset of the training data and a random subset of features. It leverages the principle of "wisdom of the crowd" by aggregating the predictions of individual trees to make the final prediction.

Random Forest

The Random Forest algorithm is another machine learning algorithm that can be used for customer churn prediction. It works by creating multiple decision trees and combining their predictions to make a final prediction. In the context of customer churn prediction, the Random Forest algorithm can be used to identify the most important features for predicting churn.

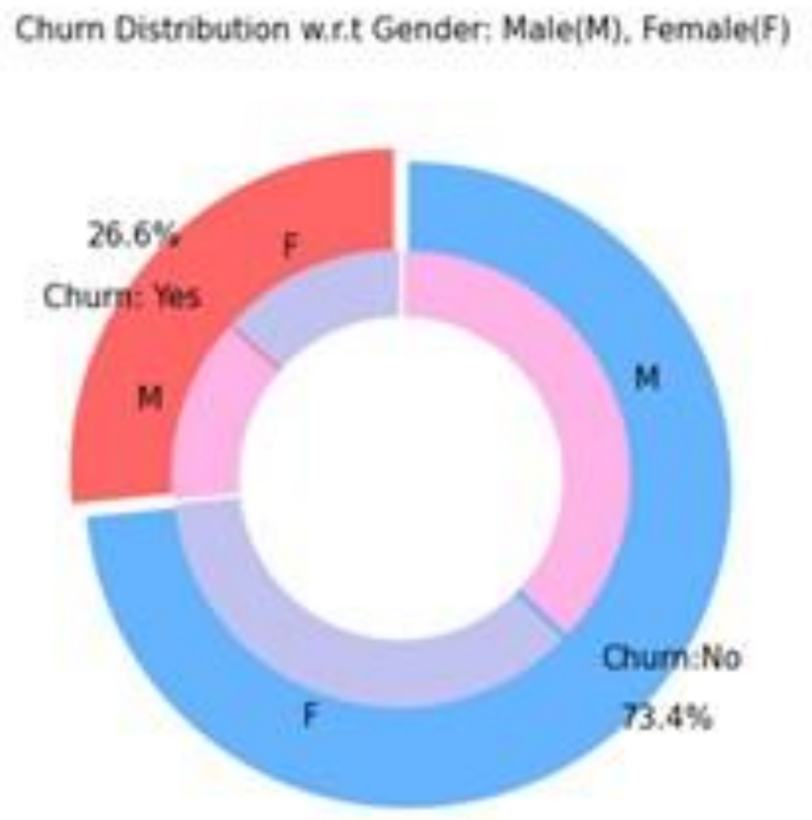


AdaBoost Classifier

The AdaBoost classifier is an ensemble learning algorithm that combines multiple weak classifiers to create a strong classifier. In the context of customer churn prediction, the AdaBoost algorithm can be used to identify the most important features for predicting churn and to make accurate predictions. It works by assigning weights to each instance in the training data and adjusting these weights based on the accuracy of the previous iterations.

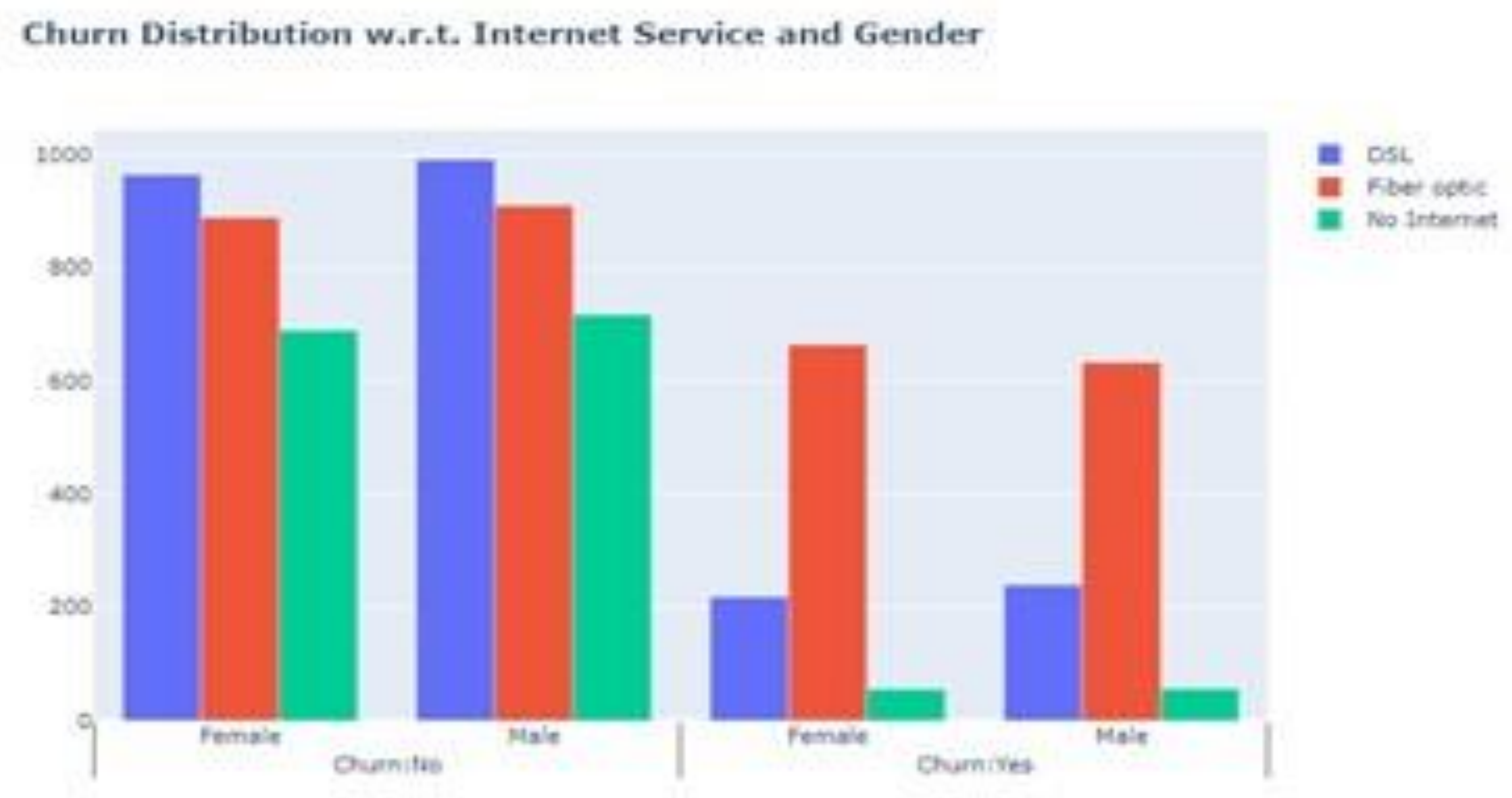
Data Visualization

From the data as we can see a costumers are mostly male with 50.5% and 49.5% female, and the costumers who decide to churn 26.6% switched to another firm.



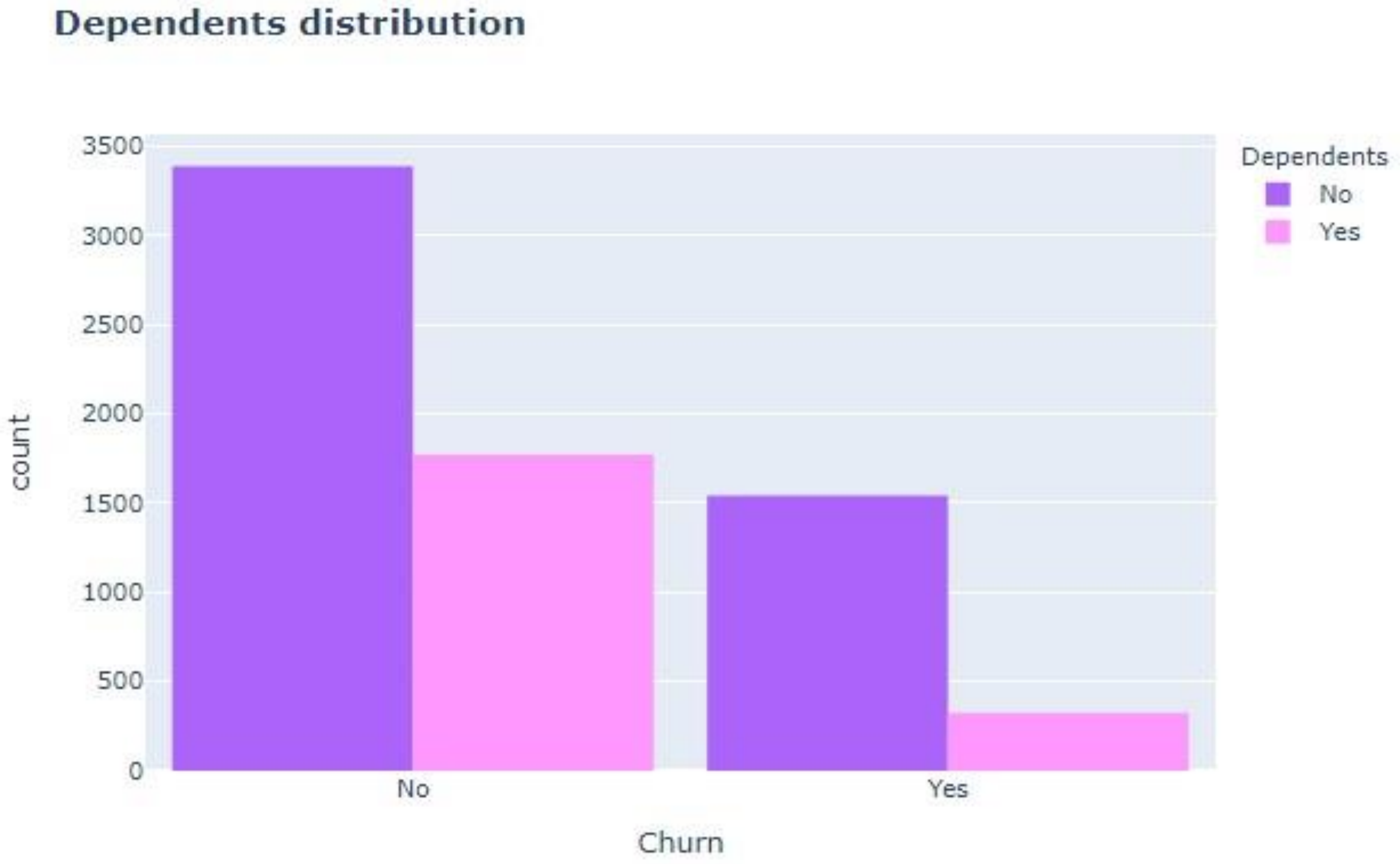
The percentage or number of customers that switched service providers differs just little. When it came to switching to a different service provider or organization, all genders acted similarly.

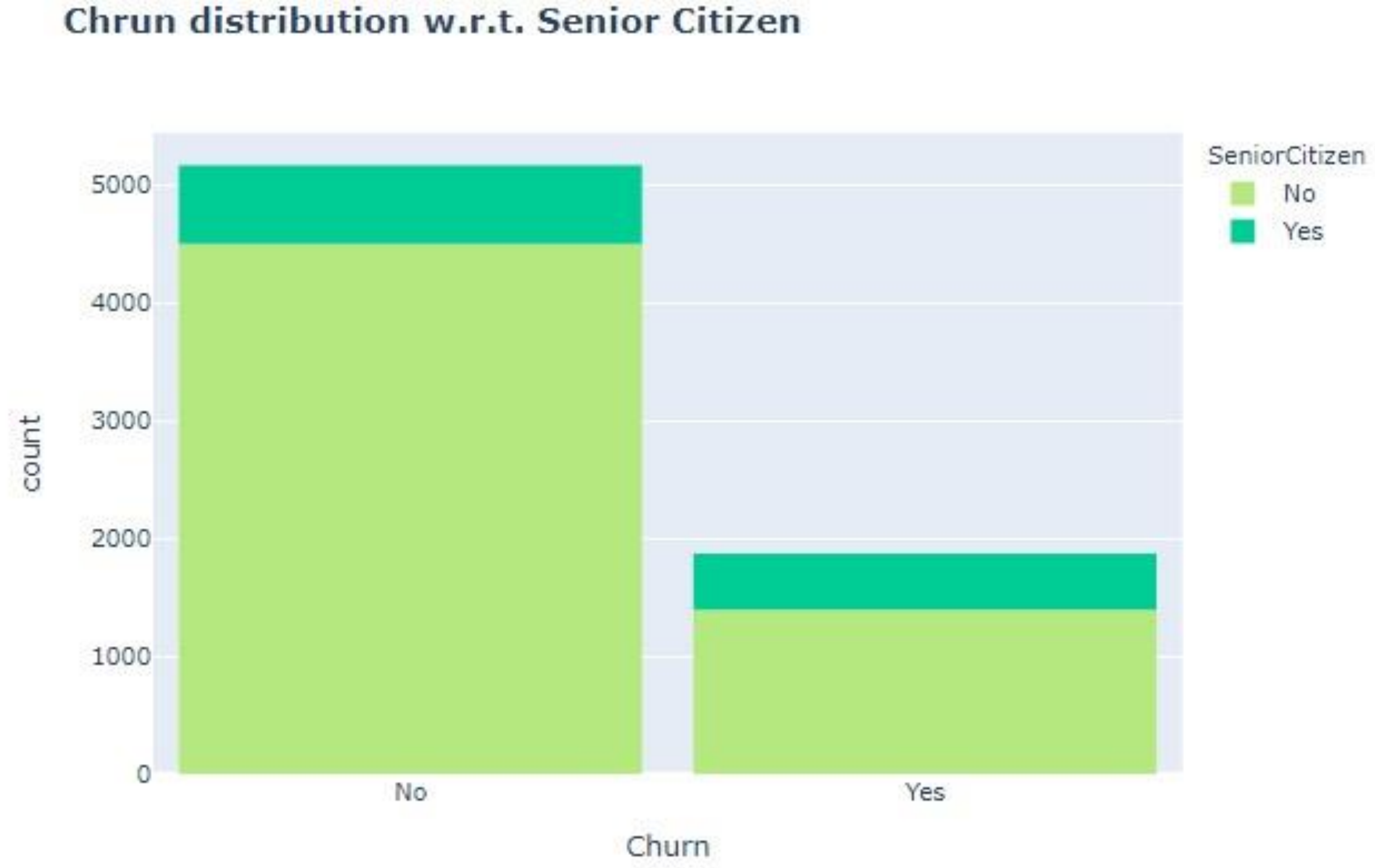
Major customers that left, choose to pay with electronic checks. Customers were less likely to left when they chose bank transfers, mailed checks, or automatic creditcard transfers as their payment option.



Numerous people pick the Fiber optic service, and it is also clear that these customers have a high turnover rate, which may indicate that they are not happy with this kind of internet service. The majority of customers have DSL service, which has a lower turnover rate than fiber optic service.

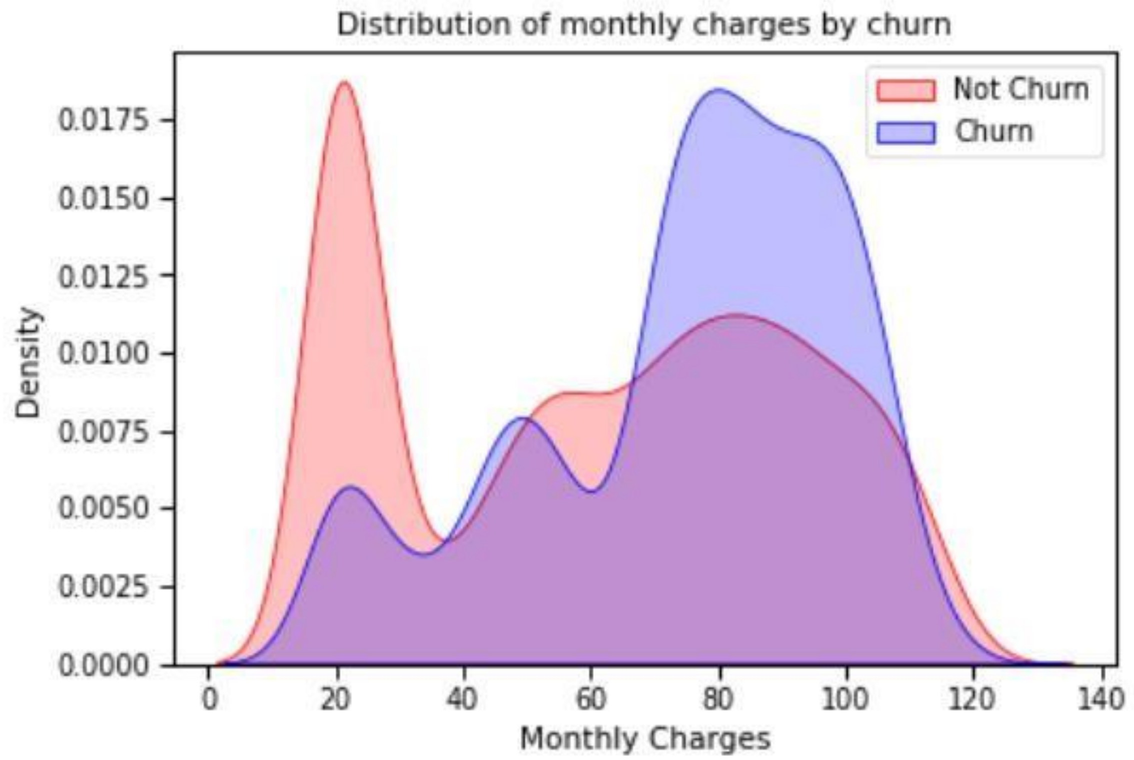
Customers without dependents are more likely to churn



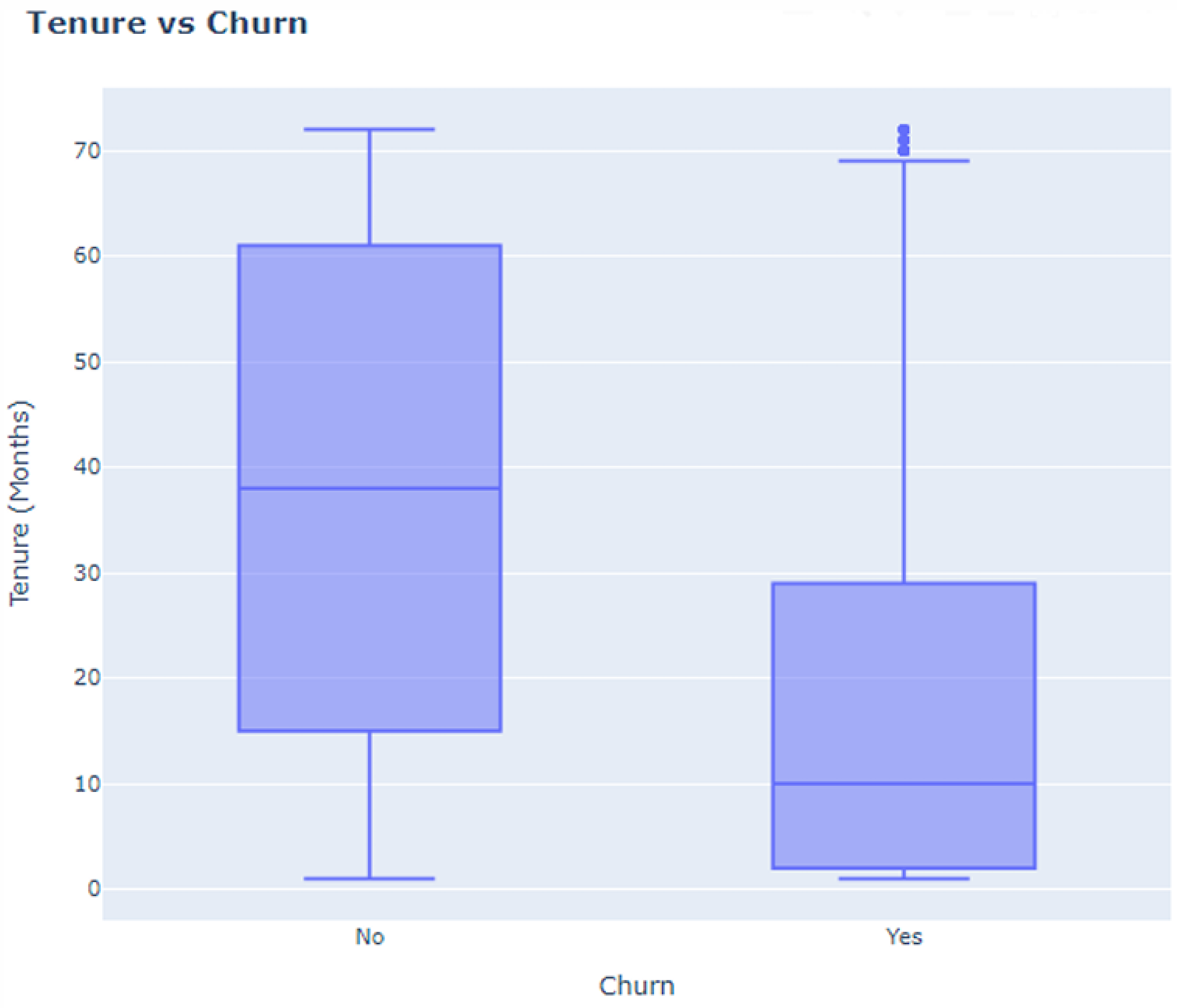


It can be observed that the fraction of senior citizen is very less. Most of the non senior citizens churn

Customers with higher Monthly Charges are also more likely to churn



New customers



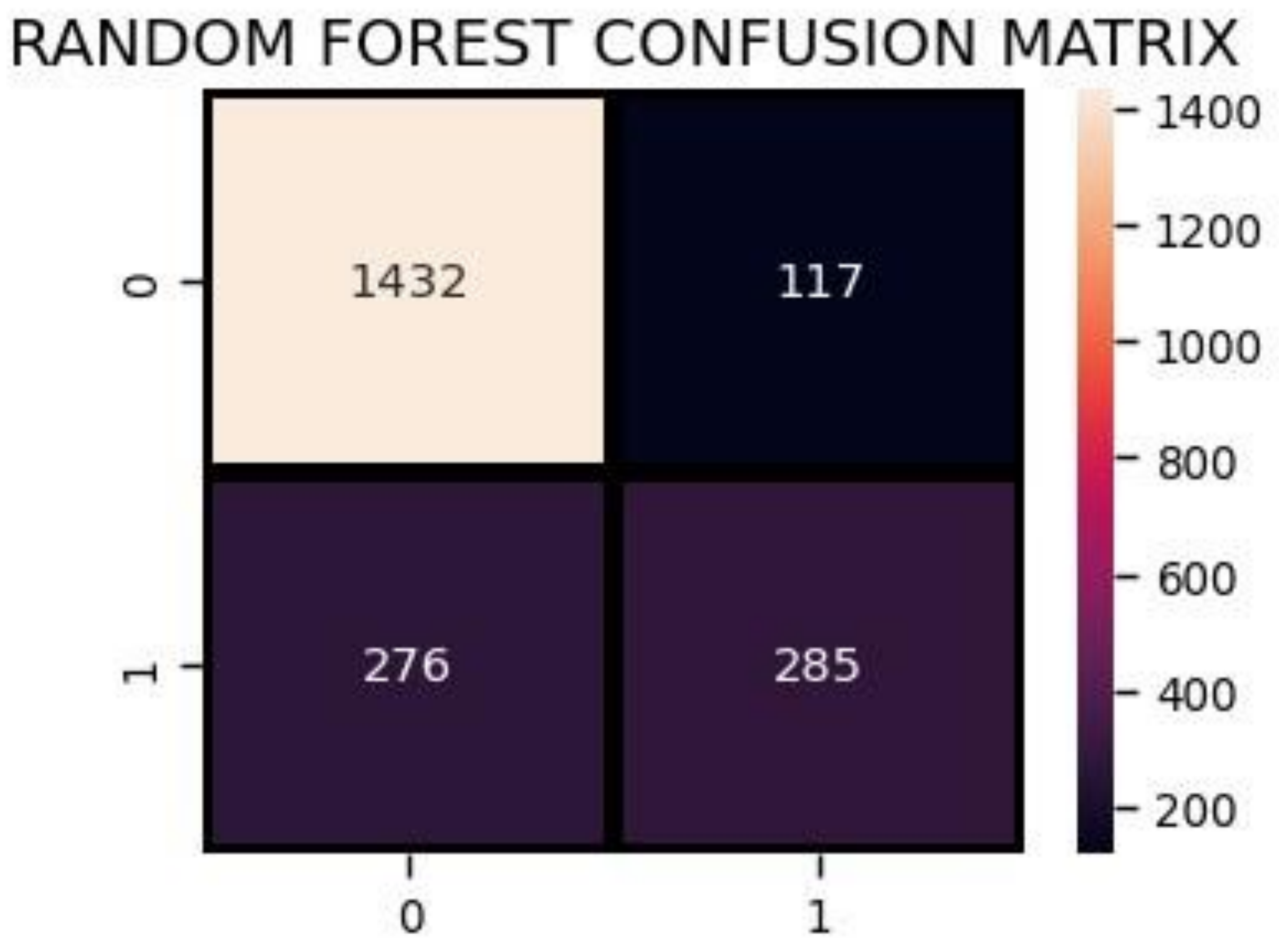
New customers are more likely to churn

Machine Learning Model Evaluations and Predictions

KNN	RANDOM FOREST	DECISION TREE	ADABOOST
KNN ACCURACY : 0.775 F1 ACCURACY : 0.78	RANDOM FOREST ACCURACY : 0.813 F1 ACCURACY : 0.81	DECISION TREE ACCURACY : 0.725 F1 ACCURACY : 0.73	ADA BOOST ACCURACY : 0.807 F1 ACCURACY : 0.81

Result Machine Learning

Based on our evaluation metrics, the Random Forest Classifier Model algorithm outperformed compare to other models in terms of accuracy and F1-score. The selected model achieved an accuracy of 0.8137440758293839 and an F1-score of 0.81 on the test dataset. From the confusion matrix we can see that: There are total 1432+117= 1549 actual non-churn values and the algorithm predicts 1432 of them as non churn and 117 of them as churn. While there are 276+285=561 actual churn values and the algorithm predicts 276 of them as non churn values and 285 of them as churn values.



Conclusion

Customer churn has a negative impact on a company's profitability. There are numerous tactics that can be used to reduce client churn. Knowing a company's customers well is the best strategy to prevent customer churn. This entails identifying clients who run the danger of leaving and trying to increase their contentment. Naturally, the primary priority for solving this problem is to improve customer service. Another tactic to lower customer churn is to increase customer loyalty through meaningful experiences and personalized service and also we can give a costumers a promotion as well..


Thank you