Journal of Positive School Psychology http://journalppw.com
2022, Vol.6, No.4, 5475-5488

CUSTOMER CHURN ANALYSIS IN TELECOM ORGANIZATION

Yash Pandey
Department of Computer science and Engineering
SRM Institute of Science and Technology
Chengalpattu, India.

Roshan Jha
Department of Computer science and Engineering
SRM Institute of Science and Technology
Chengalpattu, India.

Dr. Umamaheswari
K M Department of Computing Technologies
SRM Institute of Science and Technology
Chengalpattu, India. umamahek@srmist.edu.in

Abstract

In the telecommunications industry that is working with a different number of subscribers or consumers daily, the dividends of the company are mostly dependent on the payments provided by these subscribers. Because it has been observed that the subscribers get frustrated sometimes with the services as well as the response of the company to their queries and based on those situations the subscribers decide to stop using the services of the organization or shift to using other services that might provide less revenue to the organization ultimately resulting in the organization's losses. For this problem faced by the organizations, our project intends to find out the factors that influence the subscriber's mindset while taking decisions related to the services of a particular telecom organization and use the same to predict whether new subscribers will behave in the same manner or not. Our project initially focuses on using Exploratory Data Analysis to gather important factors related to the type of customers who can churn out in the company. Followed by that after getting the insights, we will be building the model using certain machine learning algorithms that will predict whether the customer will churn out or not. The last step will be implemented by deploying the model using Flask for other users to access it and try it out.

Keywords—— *K- nearest neighbors (KNN), Random Forest (RF), Exploratory Data Analysis (EDA), Earnings Before Interest, Taxes, Depreciation, and Amortization (EBITDA)*

I. INTRODUCTION

Customer churn indicates a process in which an existing customer who used the services of the organization stopped using it or shifted to using other services of the same organization which do not provide the organization much revenue. Thus, customer churn can directly impact the revenue made by an organization. The churn problem is not only limited to the telecommunication sector. The gaming sector and tourism sector also face a lot of customer churn problems. In the gaming sector, if a game is a too difficult customer is likely to churn or stop playing that game. On the other hand, if the game is too easy it would be too boring for the customer, hence, would ultimately become a reason for churning from that game. In the tourism sector, If the travel package released by the organization lack adequate rest during the journey customer will be uncomfortable and would churn from that package and will shift to another package. Apart from this, that customer will discourage his friends and relatives to not going with that package. Hence, in the worst- case scenario, the company could also witness an exponential rise in the churn rate. In our project, we would be primarily focussing on the customer churn rate in telecom organizations.

The telecommunications sector is one of the main industries accelerating the growth of multiple nations jointly with their services. So, the businesses (both new and established ones) operating within this sector face constant trouble of customer churn. The customer churn problem has been on a rise recently which is directly dependent on the establishment of other telecom organizations providing better services. Let's understand the reasons behind the origin of customer churn in today's telecommunication sector: (1) Tougher Telecom Environment: Dominance by few well-rounded market players making it saturated and leading to intense price wars. (2) Smarter and more demanding customers: Comparison shoppers increase their demands for better services at lower costs and reduce their loyalty.

The churn in the telecom environment can be of varied types some are: - (1) Tariff Plan Churn: - This type of churn usually occurs when a subscriber or customer changes from a Rs 60 to Rs 20 plan or vice versa. One probable reason could be that the company isn't providing good services in return for the subscriber investment. (2) Service Churn: This is the type of churn that occurs when a subscriber shifts from a weekly plan to a yearly plan, or a monthly plan to a weekly plan, or vice versa. (3) Product Churn: When a customer transitions from a post-paid plan to a prepaid plan or vice versa is known as product churn. There can be various reasons for product churn. One reason could be that the post-paid market is providing better subscription offers than its prepaid counterparts.

Companies must accurately estimate their customers' behavior to capture the aforementioned challenge. Managing customer turnover may be done in two ways: (1) reactive and
(2) proactive. After receiving a cancellation request from a client, the corporation offers the consumer enticing options to keep them as a customer. Proactively anticipating client attrition, strategies are made available to consumers in advance of any actual churn. In this issue, churners are sorted from non-churners using a binary classification technique [1]. Increased rivalry in the telecommunications business is a result of globalization and improvements in the industry. [2] Profit maximization is a must in today's competitive environment, thus many techniques, such as recruiting new consumers, up-selling existing customers, and extending the retention time for current customers, have been advocated. Retaining current clients is the most cost-effective strategy out of the bunch. Dissatisfaction with customer service and assistance is the most common cause of customer attrition. Predicting churning clients is the key to solving this issue.

Consumer churn is unavoidable, but that isn't always an unpleasant experience for the customer. Customer turnover represents a window of opportunity for businesses [8]. Customer turnover hurts revenue and marketing costs for most telecom companies. Rivals can't prevent consumer churn when it's introduced into the market. Dealing with client turnover is a fantastic opportunity for businesses to drastically alter their market position. On the one hand, customer churn may be a reflection of operational issues, aiding enterprises in a good perspective of their company and implementing focused actions to reduce customer churn while also improving operations and management. For their part, businesses may get a better knowledge of the goods and services that their consumers demand and implement changes in response to customer turnover, which would significantly improve the profitability and market position of the company in question.

II. RELATED WORK

In today's scenario there already exist several projects detecting customer churn in telecom organizations. Each project consists of some pros and cons.

Ullah et al. [3] propose a churn prediction model that uses classification, as well as clustering techniques to identify the churn customers and provides the factors behind the churning of customers in the telecom sector. Feature selection is performed by using information gain and correlation attribute ranking filter. The proposed model first classifies churn customers data using classification algorithms, in which the algorithm performed well with 88.63% correctly classified instances. Creating effective retention policies is an essential task of the CRM to prevent churners. After classification, the proposed model segments the churning customer's data by categorizing the churn customers in groups using cosine similarity to provide group-based retention offers. This paper also identified churn factors that are essential in determining the root causes of churn.

Li et al. [4] discuss the development of a model using big data analytics strategy for the dataset provided and using it predictions are made regarding the list of customers with their susceptibility listed in descending order. After getting the list through the techniques mentioned in the previous step, user segmentation and piecewise regression are used to find the highly relevant parameters followed by the division of customers into different categories based on the above-found parameters. Using regression analysis one can estimate the prediction rates for different groups of customers. High computing storage took some time in predicting results and accuracy rates of about 80% were achieved. Based on the results the organization was able to prioritize the customers who shall be given extra attention to influence their mindsets of continuing their relations with the company.

K. Sandhya Rani et al. [5] proposed a churn prediction model for telecommunication companies using machine learning techniques namely logistic regression. A comparison is done on the efficiency of the algorithm on the available dataset. They have used R programming language which is not as reliable as Python.

Essam Shaaban et al. [6] examined historical activity, and predictive modeling attempts to make predictions about future client behavior. Predictive modeling may be used to identify customers who are at risk of leaving [7]. Customer Relationship Management (CRM) data and DM are used to create client-level models that explain the chance that a certain customer would perform a specific action.

Sales, marketing, and client retention-related activities are common. Various models may be used to characterize the distinction between churners and non-churners in an organization in terms of their behavior.

While Lee et al. [8] looked into the relationship between customer happiness and switching cost in the context of French mobile communications, they found that the greater the switching cost the less probable it is for customers to churn when customer satisfaction remains constant. According to Madden et al. [9], the key variables driving customer turnover were monthly ISP use and family income. Win-back tactics were proposed by Amin et al. [10] after analyzing customer attrition from the viewpoints of businesses, rivals, and consumers and concluding. Han et al. [11] investigated the link between consumer attitude, switching obstacles, customer satisfaction, and customer retention, concluding that customer satisfaction was positively connected with customer retention. For this reason, Oghojafor et al. [12] came up with techniques to reduce the rate of client turnover. Effective customer win- back should be traced back to the root cause of churn. Customer churn is an important factor to consider when determining whether or not customers can be won back, according to Tokman and colleagues [13].

Companies may benefit from loyal consumers by minimizing the expense of public relations, negotiating, and attracting more new customers with a herd mentality, therefore reducing customer development costs and improving the chances and time for businesses to make basic profits. Additionally, they may aid in securing market position, reducing risk, and increasing entry hurdles for new businesses. Many businesses are preoccupied with finding new ways to bring in consumers while overlooking the importance of retaining current ones and maximizing their purchasing power. According to Reichheld et al. [14], the longer a company has a commercial connection with a client, the more money it makes from that consumer.

Almuqren et.al. [15] addressed this in their research, which studies customer churn, prediction models. Because they depend on previous customer data, the existing churn prediction algorithms have a limited shelf life. The data loses predictive value with time [16], which may not offer telecom firms the optimal churn prediction experience. Structural data framework and real-time analytics need to be integrated to target consumers in real-time [17]. The existing churn prediction algorithms do not take into account regional and linguistic characteristics, which results in geographical and cultural sampling mistakes [18].

III. PROPOSED WORK

Machine Learning in simple terms is a subset of Artificial Intelligence and Computer Science. It learns through a bunch of data and predicts a certain outcome [19]. The more the data the better will be the accuracy. Machine learning is mainly subdivided into three parts (1) Supervised Learning (2) Unsupervised learning (3) Semi-supervised. We will be using Supervised Learning in our project.

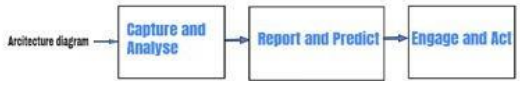


Fig. 1. Network Architecture of Proposed Model

The above-shown figure is the Block/Network Diagram of our project. The full version of this diagram will be shown and explained in the implementation section. As we have already discussed in our introduction slide, our project deals with the problem of customer churn mitigation in telecom organizations. To achieve a particular goal, we should first aim to establish certain milestones and should strive to achieve them one by one, we have also implemented the same strategy in our project and have grouped all tasks into various categories. The categories are as follows: -

A. Capture and Analyze

This section mainly aims at the dataset formation process. Under this section formation and

understanding of problem statement and business took place respectively. We have researched the customer churn impact on the organization and mapped the strategies to mitigate it. After the problem identification, we identified the type of data needed to build the data set. The exploration of data was carried out and after the data was found it was being extracted and requested from the source to build the data set. The data extraction took place from the Tariff and Usage Data, Customer level data, Recharge Data, etc. After the data extraction job, we aggregated cleaned and transformed the dataset. Now the dataset was all ready for Exploratory data analysis which is part of the Report and Predicts.

B. Report and Predict

When we dive into the Report and Predict section, we are equipped with a dataset that is ready to be explored further. In this section, our final objective is to finalize a model giving the best results on our dataset. Before this Exploratory data, the analysis part was carried out which involved operations like

(1) Data Sourcing (2) Data Cleaning (3) Univariate Analysis (4) Bivariate Analysis (5) Derived metrics. After this, we design a predictive model. Once the model has been created, we would be implementing it and the best one will be chosen for deployment which is the part of Engage and Act Section.

C. Engage and Act

Now, this is our final segment which is regarding the deployment of our project. As we have already discussed in our previous section that we will be finalizing a Machine Learning model giving the best results. So now we would be using the Flask tool for the deployment procedures. It's very important to have easy access to our project so that even a layman can understand what's going on and also when a particular subscriber is likely to churn. We would also be providing churn drivers and KPIs for tracking and monitoring purposes. Recommendations on monthly churn initiatives may also be provided.

D. Types of Subscribers in the Telecom Environment

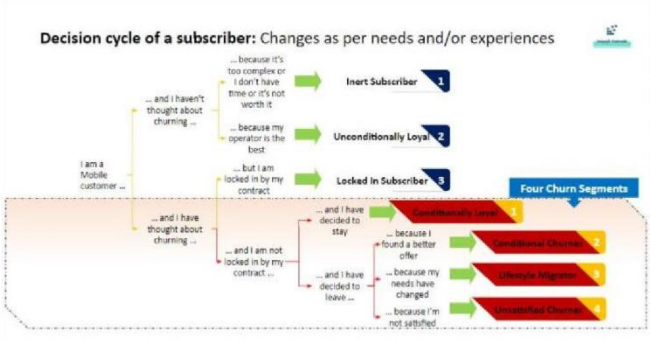


Fig. 2. Decision Cycle of a Subscriber

a) Inert Subscriber: The customers who fall into this category are non-churners. The reason is that they are too lazy to participate in the process of churning from one plan/company to other. A perfect example of such a scenario is as follows. Suppose there is a person who is currently a subscriber of Airtel. After 6 months of subscription, he realizes that airtel is not giving a good internet connection and also that its plan is also not as good as expected. Now he starts to think about churning but in the back of his mind is also aware of the formalities and documentation that he is required to do for churning. He is too lazy for the procedure of churning. Hence, he finally decides not to churn.

b) Unconditionally loyal customers: We can guess by the title that these customers are also non-churners. These are the type of customers are generally an old customer who is subscribed to a particular company for 10-30 years and is kind of attached to this company. Due to this emotional quotient and the level of comfort that develops over years of subscription the subscriber often forgets about the performance of the company and never churns from the organization.

c) Locked in Subscribers: These are the customers who are potential churners but are categorized as non- churner as they are bounded by a contract. The contract can be for a year or two.

d) Conditionally Loyal Customer: These are the types who are willing to churn and are not bounded by any contract but still have decided to stay back. The reason is that they are waiting for a better offer from some other organization so that they can leave this company at the earliest.

e) Conditional Churners: The customers falling into this category have thought about churning and are not locked under any contract and are churning as they have found a better package or plan in some other organization. These types of customers are in a majority.

f) Lifestyle Migrators: These customers are high- level churners. They have thought about churning and are not bounded by any contract. They usually churn due to their requirement changes or they want to experiment with something new. Customers falling into this segment are not at all loyal customers and have a high tendency of churning.

g) Unsatisfied Churners: When a customer is churning due to a lack of satisfaction from the company. We can call such a customer an Unsatisfied churner. There could be a ton of reasons leading to this type of churn. Firstly, the service of a particular company may not be good. The word "data" is plural, not singular.

IV. IMPLEMENTATION

Steps implemented for Data Preparation

The "Capture and Analyze" and "Report and Predict" components that handle the collection and aggregation of data into a single, easy to refer to as well as a compatible dataset making it easy to work with the data analysis and model building stages as well as using the dataset to find out about what factors influence the customer churn scenario concerning a particular organization and using the same to

build machine learning models that will predict whether a new customer will churn out or not. Before passing the dataset for the further stages, we also focus our efforts on cleaning the dataset (making sure that the data present within it has no abnormalities related to missing values, outdated matches, etc.) The data collection and assimilation process consist of collecting the data from the following sources:

label. We also used a correlation matrix and heat map for the depiction of the degree of relationship between features and labels.

After the conclusion of our EDA part we jumped on to our model building part in which we tried out the following four algorithms: - 1) K-Nearest Neighbors 2) Random Forest Classifier 3) Decision Tree Classifier 4) Principal Component Analysis.

Once we found KNN to be the best model giving the most accurate results we dumped it using pickle and deployed our project using Flask.

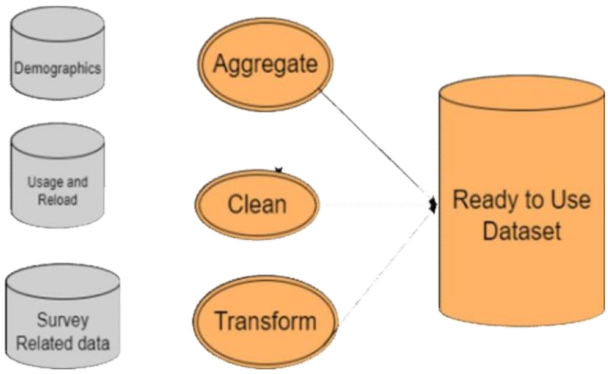


Fig. 3. Capture and Analyze

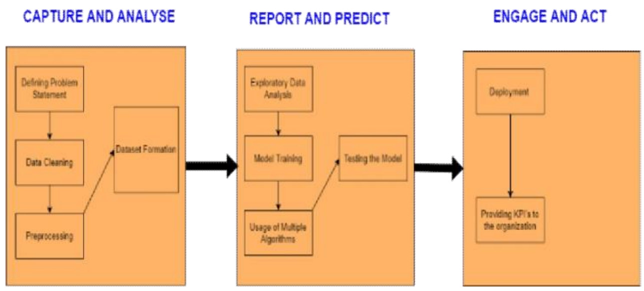


Fig. 5. Architecture Diagram

a) Customer Demographic Data:

A. Dataset

This data is concerned with the reports released by telecom organizations which provide us with important factors from the organization's perspective in a real-time environment. These factors are related to the crucial ones that impact a customer's decision related to utilizing the services of a particular telecom organization

b) Usage and Reload Data:

Several research studies have utilized datasets that contain information about the parameters related to subscribers' utilization of the network. Extensive work has been done on datasets through various other data scientists and data analysts that are shared on public sites such as Google Dataset research and Kaggle are also utilized here.

c) Survey related Data:

Lots of online surveys were conducted that collected customer feedback and their expectations in the present as well as future from telecommunication services. So, through these responses, we were able to map it to the real-time attributes that we collected in the usage and reload data module.

Fig. 3. Report and Predict

Under the report and predict section we carried out Exploratory data analysis in which we performed Data Sourcing, Data Cleaning, Univariate, and Bivariate Analysis. Under Data sourcing we imported our dataset and performed some basic data transformation. We also did the behavioral segmentation by segregating the churn ratio. Under Bivariate and Univariate analysis, we drew various insights and comparisons between various features and also the churn In our dataset, there are 21 columns and 7043 rows which were after the Exploratory Data Analysis part converted to 51 columns and 7043 rows. Our dataset contains the "Churn" column which is the label and the rest of the 50 columns are predictors. As we have imported our dataset from Kaggle it was imbalanced with some missing values. We have dealt with all the problems and have rectified them in the data cleaning and exploration section of our EDA.

B. Exploratory Data Analysis

Exploratory Data Analysis is the method of data summarization using various charts and graphs. It's also used to get familiar with the dataset. The objective of EDA is to increase confidence in our dataset to such a level so that we can build some model out of it.

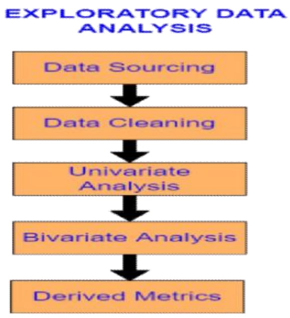


Fig. 6. EDA flow Diagram

In figure 7, we can see that through EDA we have known about the data types of all the features and labels smoothly. So from the above figure, we found out that there are only three numerical type features i.e. Senior Citizen, tenure, Monthly Chargers.

```
In [9]:  # Checking the data types of all the columns
         telco_base_data.dtypes

Out[9]:  customerID           object
         gender               object
         SeniorCitizen         int64
         Partner              object
         Dependents           object
         tenure                int64
         PhoneService         object
         MultipleLines        object
         InternetService      object
         OnlineSecurity       object
         OnlineBackup         object
         DeviceProtection     object
         TechSupport          object
         StreamingTV          object
         StreamingMovies      object
         Contract             object
         PaperlessBilling     object
         PaymentMethod        object
         MonthlyCharges      float64
         TotalCharges         object
         Churn                object
         dtype: object
```

Fig. 7. Datatypes of various feature

In figure 8, we have used describe function to get a good idea of the 3 numerical datasets. After studying and analyzing the above table we have found three insights. 1)Senior Citizen is a categorical feature hence the 25%-50%-75% distribution is not proper. 2) 75% of customers have a tenure of fewer than

55 months. 3) Average Monthly charges are USD 64.76 whereas 25% of customers pay more than USD 89.85 per month.

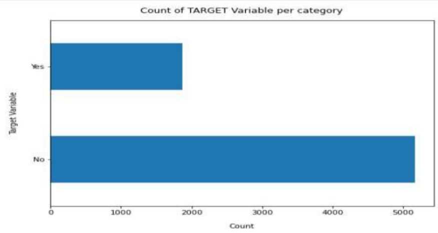
```
# Check the descriptive statistics of numeric variables
telco_base_data.describe()
```

	SeniorCitizen	tenure	MonthlyCharges
count	7043.000000	7043.000000	7043.000000
mean	0.162147	32.371149	64.761692
std	0.368612	24.559481	30.090047
min	0.000000	0.000000	18.250000
25%	0.000000	9.000000	35.500000
50%	0.000000	29.000000	70.350000
75%	0.000000	55.000000	89.850000
max	1.000000	72.000000	118.750000

Fig. 8. Description of numerical columns

As we already know that our target variable is churn. Hence, we have plotted a bar graph between a count of customers vs churn and no-churn. The above figure illustrates the graph and also the ratio between the churn and non-churn which accounts for 73:27. This is an imbalanced dataset. As our data set is not balanced, we analyzed our data with other features while taking the target values separately to get some insights.

```
telco_base_data['Churn'].value_counts().plot(kind='barh', figsize=(8, 6))
plt.xlabel("Count", labelpad=14)
plt.ylabel("Target Variable", labelpad=14)
plt.title("Count of TARGET Variable per category", y=1.02);
```




```
100*telco_base_data['Churn'].value_counts()/len(telco_base_data['Churn'])

No     73.463013
Yes    26.536987
Name: Churn, dtype: float64
```

Fig. 9. Churn Vs Non-Churn ratio

Data Cleaning: This is the 2nd step of our EDA after the data sourcing that we have already seen above. Data cleaning is the process of cleaning data and improving the quality of the dataset for further data analysis and model building. The sole benefit of data cleaning is that we get a good quality dataset that is without any outliers, missing data, and useless columns. This enhances the model's accuracy.

```
telco_data.TotalCharges = pd.to_numeric(telco_data.TotalCharges, errors='coerce')
telco_data.isnull().sum()

customerID           0
gender               0
SeniorCitizen        0
Partner              0
Dependents           0
tenure               0
PhoneService         0
MultipleLines        0
InternetService      0
OnlineSecurity       0
OnlineBackup         0
DeviceProtection     0
TechSupport          0
StreamingTV          0
StreamingMovies      0
Contract             0
PaperlessBilling     0
PaymentMethod        0
MonthlyCharges       0
TotalCharges        11
Churn                0
dtype: int64
```

Fig. 10. Illustrations of missing values

In the above figure, we can see that there are 11 missing values in the Total Charges column. After the discovery of a missing value in the Total charges column we did IsNull () operation to determine the location of missing values in the total charge column. We got a data frame containing only the missing values row. Then we performed dropna () to drop those 11 rows.

After dealing with the missing values, we introduced a new column named ‘Tenure Group’.

The figure 11, the shown range is in months (1-12 months). As we can see that the customers who have 1-12 months tenure are having higher churn counts. The logic behind this is quite obvious, less bonded the customer more will be the likelihood of churn and vice versa.

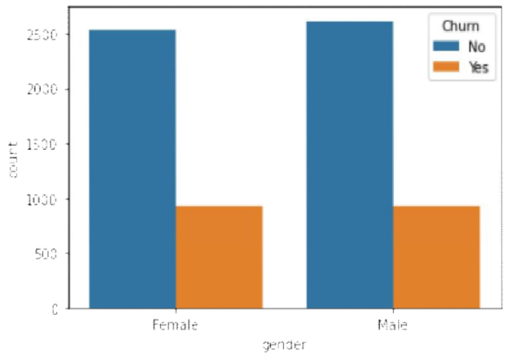
```
telco_data['tenure_group'].value_counts()

1 - 12     2175
61 - 72    1407
13 - 24    1024
25 - 36     832
49 - 60     832
37 - 48     762
Name: tenure_group, dtype: int64
```

Fig. 11. Formation of a new feature

Univariate analysis: Data exploration deals with the analysis of a good quality dataset which we obtain through the data cleaning step. This is a type of Data Exploration method and is the 3rd step of our EDA. It analyses a single feature along with the label at a time. The main objective of the univariate analysis is to interpret different variables easily. Histograms are mainly used in univariate analysis. Some of the graphs obtained are as follows: -

Fig. 12. Gender Vs Churn count



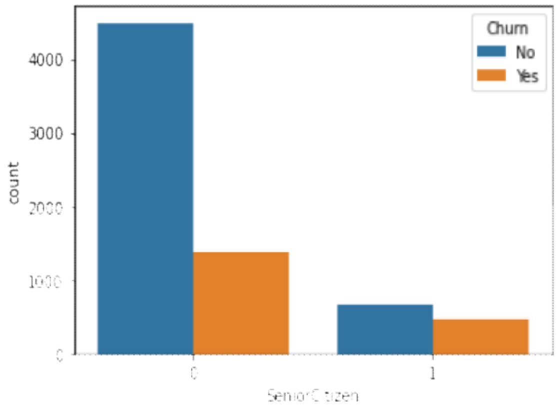


Fig. 13. Senior Citizen Vs Churn count

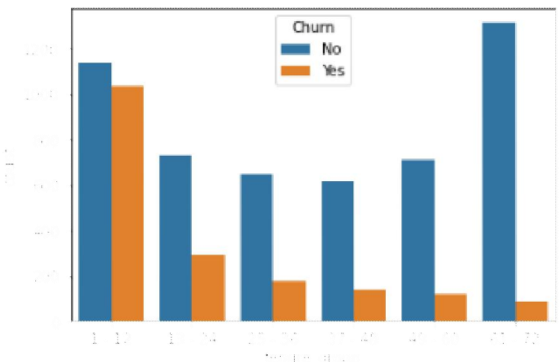


Fig. 14. Tenture_group Vs Churn count

There are plenty of histograms which is not possible to show in this paper. If we talk about the tenure group vs Churn (Figure 14) count the 1-12-month tenure group as having the highest churn ratio as the tenure is less. Whereas 61-72 months are having the lowest churn rate due to the longer duration of tenure.

Moreover, in the gender vs churn plot (Figure 12), the churn ratio of both male and female are almost the same. So, we can eliminate these columns as they are not giving us many insights about churn rate.

Likewise, In Figure 13 we can be confused for a moment and can think that non-senior citizens are the higher churners than senior citizens. But if we see the churn to a non-churn ratio in both senior citizens and non-senior citizens, we can see that ratio is higher in senior citizens. Hence, we can conclude that the senior citizens are high churners.

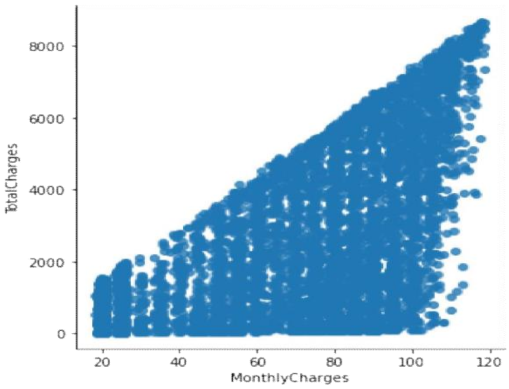


Fig. 15. Monthly Charges Vs Total Charges

As we can see in the above figure.15 monthly charges is having a linear relationship with the Total charges as expected.

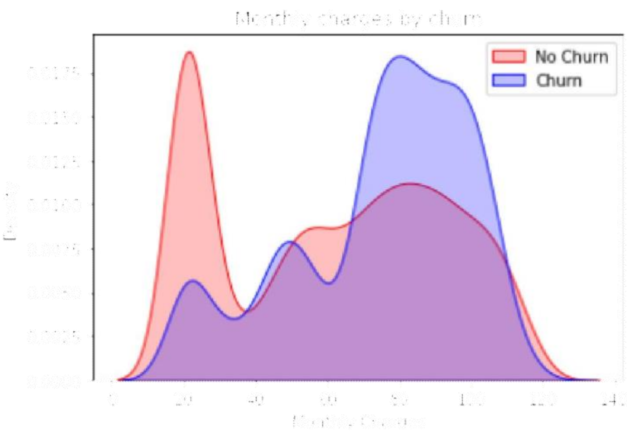
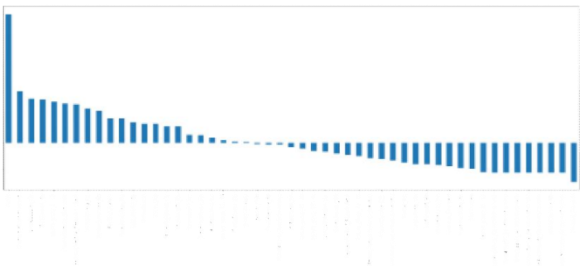


Fig. 16. Monthly Charges Vs Density Churn

In Figure [16] it is quite evident that churn density is higher when the monthly charges are high and lower when the monthly charges are low.

Bivariate Analysis: This is our 4^{th} and final step in our EDA. Unlike, univariate analysis, Bivariate analysis deals with two variables instead of one.

Fig. 17. Correlation Matrix



The above shown in Fig 17 depicts the correlation of different features concerning the churn label. In the above matrix, we can see some features having a positive, negative, and neutral correlation with the churn. The features which are neutral i.e. having a correlation score of 0, are not important features.

Figure 18 illustrates the correlation between different features along with the churn.

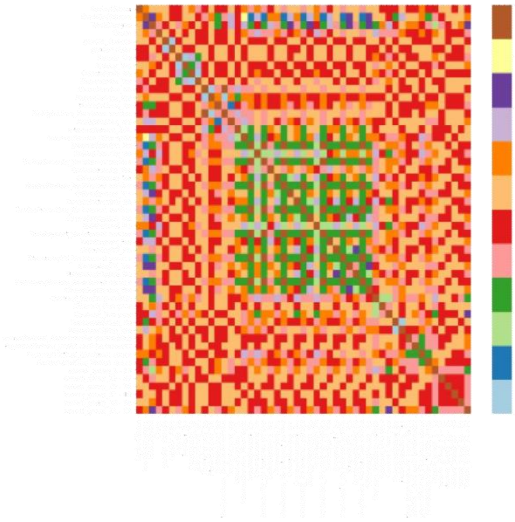


Fig. 18. Heat Map

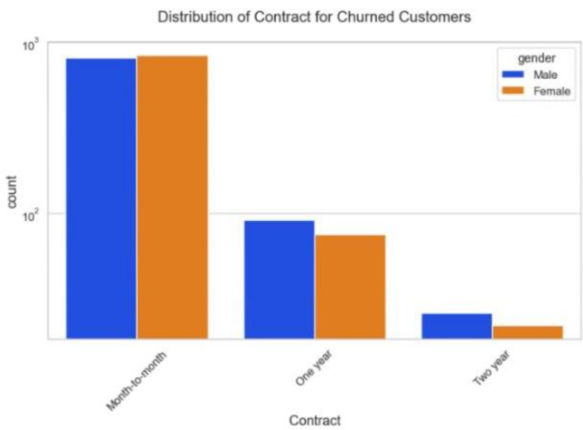


Fig. 19. Contract Type Vs Churn Count

From Figure 19 it's evident that the customers irrespective of their gender are high churners during the month-to-month contract type. Conversely, As the duration of the contract type increased to two years, the churn rate also decreased drastically for both male and female customers.

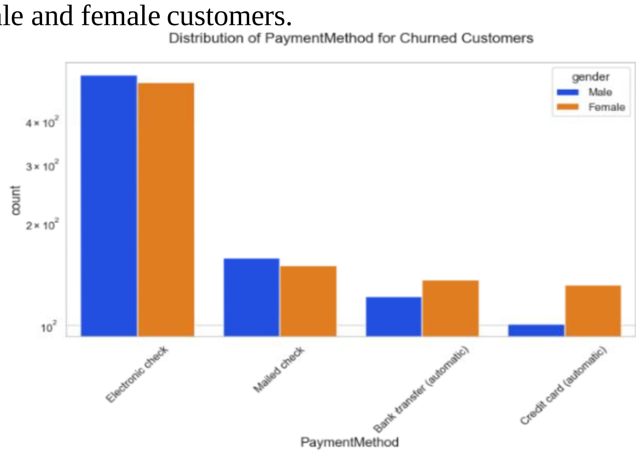


Fig.20. Payment Method Vs Churn Count

In Figure 20 we can visualize an interesting insight that the customers using an electronic check as their payment method are the highest churners. Moreover, male credit card users are higher churners than their female counterparts.

V. RESULTS AND DISCUSSIONS

A. Important EDA Insights:

After the completion of EDA, we found some extremely useful and interesting insights and results. Firstly, Senior Citizen is categorical hence the 25%-50%-75% distribution is not proper. Secondly, Higher Monthly charges, Lower tenure, and Lower Total charges are linked to High Churn. Thirdly, HIGH Churn has seen in case of Month-to-month contracts, No online security, No Tech support, First year of subscription, and Fiber Optics Internet. Fourthly, LOW Churn is seen in the case of Long-term contracts, Subscriptions without internet service, and the customers engaged for 5+ years. Fifth, Factors like Gender, Availability of phone service and the number of multiple lines have almost NO impact on Churn.

These are some more quick insights: -
1. Electronic check medium is the highest churner.
2. Contract Type - Monthly customers are more likely to churn because of no contract terms, as they are free-to-go customers.
3. No Online security, No Tech Support category are high churners.
4. Non-senior Citizens are high churners.

B. Model Accuracy:

A propensity score is used to train machine learning algorithms to manage client attrition. The propensity score is vital since it indicates the likelihood of a client leaving soon. CCP models' forecast accuracy relies on this. The next section will outline the different stages of the churn prediction process and the associated machine learning algorithms. We tried out many algorithms. Out of which three of the algorithms gave us the best accuracy.

Random Forest Classifier with SMOTEENN gave us a whopping 93-93.5% accuracy. 0 and 1 stand for non-churners and churners respectively. The precision determines the accuracy of the prediction. The recall is the ratio of actual positives and the total positive cases. F-1 score is the mean of precision and recall.

TABLE I. RANDOM FOREST CLASSIFICATION REPORT

	Precision	recall	F1-score	support
0	0.94	0.90	0.92	530
1	0.92	0.96	0.94	649
accuracy	-	-	0.93	1179
Macro avg	0.93	0.93	0.93	1179
Weighted avg	0.93	0.93	0.93	1179

TABLE II. RANDOM FOREST CONFUSION MATRIX

	Positive	Negative
Positive	TP = 479	FN = 51
Negative	FP = 28	TN = 621

Decision Tree Classifier with SMOTEEN gave us around the same 93.1% accuracy.

TABLE III. DECISION TREE CLASSIFICATION REPORT

	Precision	recall	F1-score	support
0	0.95	0.90	0.92	534
1	0.92	0.96	0.94	634
accuracy	-	-	0.93	1168
Macro avg	0.93	0.93	0.93	1168
Weighted avg	0.93	0.93	0.93	1168

TABLE IV. DECISION TREE CONFUSION MATRIX

	Positive	Negative
Positive	TP = 479	FN = 55
Negative	FP = 25	TN = 609

KNN with SMOTEEN yielded an impressive 94-95% accuracy. This is the model which gave us the best accuracy.

TABLE V. K-NEAREST NEIGHBOR CLASSIFICATION REPORT

	Precision	recall	F1-score	support
0	0.95	0.94	0.94	528
1	0.95	0.96	0.96	649
accuracy	-	-	0.95	1177
Macro avg	0.95	0.95	0.95	1177
Weighted avg	0.95	0.95	0.95	1177

The predict method is our POST method, which is called when we pass all the inputs from our front end and click SUBMIT.

```
@app.route("/", methods=['POST'])
def predict():
```

The run() method of the Flask class runs the application on the local development server.

Now after running the code we see localhost:5000 or http://127.0.0.1:5000/.

```
app.run()
```

```
* Running on http://127.0.0.1:5000 (Press CTRL+C to quit)
```

VI. CONCLUSION

The importance of this kind of project is very necessary for today's competitive environment. Especially in the telecom sector where customer quite often churns and results in big losses to the telecom company. From churning losses, we should learn a big lesson to be more aware and interactive with the customers to retain them in case they are willing to churn. One possible way could be through hiring and training a Customer Relationship Manager (CRM) to such an extent that customer never thinks of churning and become an Unconditionally loyal subscriber. The old and traditional ways of working in telecommunication organizations should be amended. Many companies are still not aware of the benefits of Machine Learning and Data Science in predicting the likelihood of churn. As a result, many companies tend to make mistakes by not hiring ML engineers and then losing customers. Our model has been prepared by using a variety of Algorithms like k-means, k-means++, Random Forest Classifier, Decision Tree Classifier, and K-Nearest- Neighbors. Out of many only three algorithms were selected and finally a single algorithm i.e. KNN was selected for our model building giving 94-95% accuracy. Apart from accuracy, we have also improved that our F1 score, Precision, and Recall are good.

```
app = Flask("_name_",template_folder='templates')
```

The load page method calls our home.html.

```
@app.route("/")
def loadPage():
    return render_template('home.html', query="")
```

REFERENCES

[1] Lalwani, P., Mishra, M.K., Chadha, J.S. et al. Customer churn prediction system: a machine learning approach. Computing 104, 271– 294 (2022). https://doi.org/10.1007/s00607-021-00908-y.

[2] Chen, H., Chiang, R.H., Storey, V.C.: Business intelligence and analytics: From big data to big impact. MIS quarterly pp. 1165–1188 (2012).

[3] Ullah, Irfan & Raza, Basit & Malik, Ahmad & Imran, Muhammad & Islam, Saif & Kim, Sung Won. (2019). A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2914999.

[4] L. Li, T. Chi, T. Hao, T. Yu Customer demand analysis of the electronic commerce supply chain using big data Ann. Oper. Res., 268 (2018), pp. 113-128.

[5] Rani, K. Sandhya and ., Shaik Thaslima and ., N.G.L. Prasanna and ., R.Vindhya and ., P. Srilakshmi, Analysis of Customer Churn Prediction in Telecom Industry Using Logistic Regression (JUNE 10, 2021). International Journal of Innovative Research in Computer Science & Technology (IJIRCST) ISSN: 2347-5552, Volume-9, Issue- 4, July 2021, https://doi.org/10.21276/ijircst.2021.9.4.6

[6] Shaaban, Essam & Helmy, Yehia & Khedr, Ayman & Nasr, Mona. (2012). A Proposed Churn Prediction Model. International Journal of Engineering Research and Applications (IJERA. 2. 693-697.

[7] M. Hassouna, Agent-Based Modelling, and simulation: An Examination of Customer Retention in the UK Mobile Market. Ph.D. thesis, Brunel University, UK, 2012.

[8] JOUR Farouk, Ahmed Zhao, Ming Zeng, Qingjun Chang, Ming Tong, Qian Su, Jiafu (2021) A Prediction Model of Customer Churn considering Customer Value: An Empirical Research of Telecom Industry in China Discrete Dynamics in Nature and Society1026-0226 https://doi.org/10.1155/2021/7160527.

[9] J. Lee, J. Lee, and L. Feick, ""The impact of switching costs on the customer satisfaction-loyalty link: mobile phone service in France," Journal of Services Marketing, vol. 15, no. 1, pp. 135–148, 2001.

[10] A. Amin, S. Anwar, A. Adnan et al., "Customer churn prediction in the telecommunication sector using a rough set approach," Neurocomputing, vol. 237, pp. 242–254, 2017.

[11] H. Han, K.-J. Back, and B. Barrett, "Influencing factors on restaurant customers' revisit intention: the roles of emotions and switching barriers," International Journal of Hospitality Management, vol. 28, no. 4, pp. 563–572, 2009.

[12] B. Oghojafor, G. Mesike, R. Bakarea, C. Omoera, and I. Adeleke, "Discriminant analysis of factors affecting telecoms customer churn," International Journal of Business Administration, vol. 3, no. 2, p. 59, 2012.

[13] M. Tokman, L. M. Davis, and K. N. Lemon, "The WOW factor: creating value through win-back offers to reacquire lost customers," Journal of Retailing, vol. 83, no. 1, pp. 47–64, 2007.

[14] F. F. Reichheld and W. E. Sasser, "Zero defections: quality comes to services," Harvard Business Review, vol. 68, no. 5, pp. 105–111, 1990.

[15] Almuqren, Latifah, Fatma S. Alrayes, and Alexandra I. Cristea. 2021. "An Empirical Study on Customer Churn Behaviours Prediction Using Arabic Twitter Mining Approach" Future Internet 13, no. 7: 175. https://doi.org/10.3390/fi13070175.

[16] Hassouna, M.; Tarhini, A.; Elyas, T.; Trab, M.S.A. Customer Churn in Mobile Markets: A Comparison of Techniques. Int. Bus. Res. 2015, 8, p224.

[17] Singh, I.; Singh, S. Framework for Targeting High-Value Customers and Potential Churn Customers in Telecom using Big Data Analytics. Int. J. Educ. Manag. Eng. 2017, 7, 36–45.

[18] Pagare, R.; Khare, A. Churn prediction by finding most influential nodes in a social network. In Proceedings of the 2016 International Conference on Computing, Analytics and Security Trends (CAST), Institute of Electrical and Electronics Engineers (IEEE), Pune, India, 19–21 December 2016; pp. 68–71.

[19] P. P. Shinde and S. Shah, "A Review of Machine Learning and Deep Learning Applications," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-6, doi: 10.1109/ICCUBEA.2018.8697857.

[20] S. De, P. P and J. Paulose, "Effective ML Techniques to Predict Customer Churn," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), 2021, pp. 895-902, doi: 10.1109/ICIRCA51532.2021.9544785.