Telecom Churn Prediction Using Machine Learning

line 1: 2nd Given Name Surname
line 2: *dept. name of organization*
(of Affiliation)
line 3: *name of organization*
(of Affiliation)
line 4: City, Country
line 5: email address or ORCID

***Abstract*— Customer Churn is one of the main concerns for telecommunications which directly affects the growth and revenues. It is important to identify the factors that increase customer churn so that we can reduce it and build trust toward the customers by providing better services to beat the other brands. Being able to predict the churn of the customer is the biggest success for the industry because based on facilities and services, customers switch the network easily. In this research paper, we are going to perform machine learning techniques such as logistic regression, XG Boost, and decision trees to predict the customer churn factors that help reduce it. Data visualization is important for data analysis to present the data knowledge in the form of charts and graphs. The model performance is calculated under the research paper and justifies that the XG Boost algorithm is better with an accuracy of 82%.**

I. Introduction

To make any country develop, the telecommunication sector plays a very important role. Increasing the number of operators and providing better technical services will also increase the competition. Organizations are working daily to beat the competition by using various strategies. To decrease the customer churn in the telecom sector, we have to follow 3 strategies such as achieve new and more customers, upsell the present customers, and improve and increase the retention period because if we retain the existing customer then it will be helpful to decrease churn as well as improve the value of return on investment. Customer churn is a highly competitive sector. The telecom churn dataset can be found on either Kaggle or Maven Analytics public platforms. It encompasses data of customers of telecommunications companies in California, including details like their location, subscribed services, account status, and demographic information. We've utilized pandas to ingest and analyze this dataset on telecom churn. Data cleaning and preprocessing are very important parts of the data analysis which helps here to clean the data. So many missing values are there so we removed it through the Python panda's library. Data visualization is also a very important part of plotting the charts and presenting the information in a better way.

To operate the services, customers remain a perfect entity in the telecom sector. So to improve the relationship with customers, it is important to use personalized marketing strategies and for this, the telecom sector has to provide better customer support and proper training to handle the clients. Machine learning plays an important I various sectors and telecommunication is one of the sectors to predict customer churn. Logistic regression, XG boost, and decision tree algorithm are selected here to predict the churn on the categorical dataset.

II. Implementation of Machine Learning Techniques

The implementation of machine learning techniques is a very important part of any dataset. Telecom industry is one of the main industries for all countries.

III. DATASET

The telecom churn dataset is available on the Kaggle or Maven analytics public source. This dataset contains the information of customers from telecommunication organizations in California. The dataset contains customer information such as location, services, status, and demographic details. We have read the telecom churn dataset using pandas.

```
In [53]: #read dataset
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

telecom = pd.read_csv('telecom_customer_churn.csv')
telecom.head()
```

Out[53]:

	Customer ID	Gender	Age	Married	Number of Dependents	City	Zip Code	Latitude	Longitude	Number of Referrals	...	Payment Method	Monthly Charge	Total Charges	Total Refunds	Total Extra Data Charges	Dist Cha
0	0002-ORFBO	Female	37	Yes	0	Frazier Park	93225	34.827662	-118.999073	2	...	Credit Card	65.6	593.30	0.00	0	3
1	0003-MKNFE	Male	46	No	0	Glendale	91206	34.162515	-118.203869	0	...	Credit Card	-4.0	542.40	38.33	10	[illegible]
2	0004-TLHLJ	Male	50	No	0	Costa Mesa	92627	33.645672	-117.922613	0	...	Bank Withdrawal	73.9	280.85	0.00	0	1
3	0011-IGKFF	Male	78	Yes	0	Martinez	94553	38.014457	-122.115432	1	...	Bank Withdrawal	98.0	1237.85	0.00	0	3
4	0013-EXCHZ	Female	75	Yes	0	Camarillo	93010	34.227846	-119.079903	3	...	Credit Card	83.9	267.40	0.00	0	[illegible]

5 rows × 38 columns

Figure 1: Telecom Data

We have checked the dimensions of the dataset and 7043 customer data is present.

```
In [54]: #check dimension
telecom.shape

Out[54]: (7043, 38)
```

Figure 2: Number of rows and columns

This dataset contains 38 features to define the telecom churn dataset. The attributes are['Gender', 'Age', 'Married', 'Number of Dependents', 'City', 'Number of Referrals', 'Tenure in Months', 'Offer', 'Phone Service', 'Avg Monthly Long Distance Charges', 'Multiple Lines', 'Internet Service', 'Internet Type', 'Avg Monthly GB Download', 'Online Security', 'Online Backup', 'Device Protection Plan', 'Premium Tech Support', 'Streaming TV', 'Streaming Movies', 'Streaming Music', 'Unlimited Data', 'Contract', 'Paperless Billing', 'Payment Method', 'Monthly Charge', 'Total Charges', 'Total Refunds', 'Total Extra Data Charges', 'Total Long Distance Charges', 'Total Revenue', 'Customer Status'].

```
[84]: #data columns
telecom.columns

t[84]: Index(['Gender', 'Age', 'Married', 'Number of Dependents', 'City',
       'Number of Referrals', 'Tenure in Months', 'Offer', 'Phone Service',
       'Avg Monthly Long Distance Charges', 'Multiple Lines',
       'Internet Service', 'Internet Type', 'Avg Monthly GB Download',
       'Online Security', 'Online Backup', 'Device Protection Plan',
       'Premium Tech Support', 'Streaming TV', 'Streaming Movies',
       'Streaming Music', 'Unlimited Data', 'Contract', 'Paperless Billing',
       'Payment Method', 'Monthly Charge', 'Total Charges', 'Total Refunds',
       'Total Extra Data Charges', 'Total Long Distance Charges',
       'Total Revenue', 'Customer Status'],
      dtype='object')
```

Figure 3: Column names in telecom data

Info Function is a very useful function of Python pandas that provides information on datasets like the count of null values , data type of columns, and total entries.

```
telecom.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 38 columns):
 #   Column                             Non-Null Count  Dtype
---  ------                             --------------  -----
 0   Customer ID                        7043 non-null   object
 1   Gender                             7043 non-null   object
 2   Age                                7043 non-null   int64
 3   Married                            7043 non-null   object
 4   Number of Dependents               7043 non-null   int64
 5   City                               7043 non-null   object
 6   Zip Code                           7043 non-null   int64
 7   Latitude                           7043 non-null   float64
 8   Longitude                          7043 non-null   float64
 9   Number of Referrals                7043 non-null   int64
 10  Tenure in Months                   7043 non-null   int64
 11  Offer                              7043 non-null   object
 12  Phone Service                      7043 non-null   object
 13  Avg Monthly Long Distance Charges  6361 non-null   float64
 14  Multiple Lines                     6361 non-null   object
 15  Internet Service                   7043 non-null   object
 16  Internet Type                      5517 non-null   object
 17  Avg Monthly GB Download            5517 non-null   float64
 18  Online Security                    5517 non-null   object
 19  Online Backup                      5517 non-null   object
 20  Device Protection Plan             5517 non-null   object
 21  Premium Tech Support               5517 non-null   object
 22  Streaming TV                       5517 non-null   object
 23  Streaming Movies                   5517 non-null   object
 24  Streaming Music                    5517 non-null   object
 25  Unlimited Data                     5517 non-null   object
 26  Contract                           7043 non-null   object
 27  Paperless Billing                  7043 non-null   object
 28  Payment Method                     7043 non-null   object
 29  Monthly Charge                     7043 non-null   float64
 30  Total Charges                      7043 non-null   float64
 31  Total Refunds                      7043 non-null   float64
 32  Total Extra Data Charges           7043 non-null   int64
 33  Total Long Distance Charges        7043 non-null   float64
```

Figure 4: Telecom Data Information

We have checked the unique values present in the dataset and telecom data contains multiple unique values.

```
[56]: #check all unique values
for i in telecom.columns:
    print(f"{i} ----------- {telecom[i].nunique()}")

Customer ID ----------- 7043
Gender ----------- 2
Age ----------- 62
Married ----------- 2
Number of Dependents ----------- 10
City ----------- 1106
Zip Code ----------- 1626
Latitude ----------- 1626
Longitude ----------- 1625
Number of Referrals ----------- 12
Tenure in Months ----------- 72
Offer ----------- 6
Phone Service ----------- 2
Avg Monthly Long Distance Charges ----------- 3583
Multiple Lines ----------- 2
Internet Service ----------- 2
Internet Type ----------- 3
Avg Monthly GB Download ----------- 49
Online Security ----------- 2
Online Backup ----------- 2
Device Protection Plan ----------- 2
Premium Tech Support ----------- 2
Streaming TV ----------- 2
Streaming Movies ----------- 2
Streaming Music ----------- 2
Unlimited Data ----------- 2
Contract ----------- 3
Paperless Billing ----------- 2
Payment Method ----------- 3
Monthly Charge ----------- 1591
Total Charges ----------- 6540
Total Refunds ----------- 500
Total Extra Data Charges ----------- 16
Total Long Distance Charges ----------- 6068
Total Revenue ----------- 6975
Customer Status ----------- 3
```

Figure 5: Unique Values

The describe function is a function that provides the summary of the dataset. It provides the minimum, maximum, 25%, 5

0%, 75%, count, mean, and standard deviation of the integer columns.

```
Churn Category ----------- 5
Churn Reason ----------- 20
```

```
]: #data description
telecom.describe()
```

]:

	Age	Number of Dependents	Zip Code	Latitude	Longitude	Number of Referrals	Tenure in Months	Avg Monthly Long Distance Charges	Avg Monthly GB Download	Monthly Charge	Total Charges
count	7043.000000	7043.000000	7043.000000	7043.000000	7043.000000	7043.000000	7043.000000	6361.000000	5517.000000	7043.000000	7043.000000
mean	46.509726	0.468692	93486.070567	36.197455	-119.756684	1.951867	32.386767	25.420517	26.189958	63.596131	2280.381264
std	16.750352	0.962802	1856.767505	2.468929	2.154425	3.001199	24.542061	14.200374	19.586585	31.204743	2266.220462
min	19.000000	0.000000	90001.000000	32.555828	-124.301372	0.000000	1.000000	1.010000	2.000000	-10.000000	18.800000
25%	32.000000	0.000000	92101.000000	33.990646	-121.788090	0.000000	9.000000	13.050000	13.000000	30.400000	400.150000
50%	46.000000	0.000000	93518.000000	36.205465	-119.595293	0.000000	29.000000	25.690000	21.000000	70.050000	1394.550000
75%	60.000000	0.000000	95329.000000	38.161321	-117.969795	3.000000	55.000000	37.680000	30.000000	89.750000	3786.600000
max	80.000000	9.000000	96150.000000	41.962127	-114.192901	11.000000	72.000000	49.990000	85.000000	118.750000	8684.800000

Figure 6: Data Summary

IV. DATA CLEANING

Data Cleaning is the step in which we clean the raw data and convert it into a useful format. It is necessary to clean the data because when we collect data from any source, initially it is in raw format and contains duplicate data, some columns contain no data, and data can be of any type. Here we check the null values and identify that the data contains null values.

Data Cleaning

```
In [68]: #check null values
telecom.isnull().sum()
```

```
Out[68]: Customer ID                          0
Gender                               0
Age                                  0
Married                              0
Number of Dependents                 0
City                                 0
Zip Code                             0
Latitude                             0
Longitude                            0
Number of Referrals                  0
Tenure in Months                     0
Offer                                0
Phone Service                        0
Avg Monthly Long Distance Charges  682
Multiple Lines                     682
Internet Service                     0
Internet Type                     1526
Avg Monthly GB Download           1526
Online Security                   1526
Online Backup                     1526
Device Protection Plan            1526
Premium Tech Support              1526
Streaming TV                      1526
Streaming Movies                  1526
Streaming Music                   1526
Unlimited Data                    1526
Contract                             0
Paperless Billing                    0
Payment Method                       0
Monthly Charge                       0
Total Charges                        0
Total Refunds                        0
Total Extra Data Charges             0
Total Long Distance Charges          0
```

Figure 7: Data Cleaning

We have dropped the columns that contain a high amount of missing values as well as not useful for the dataset because when we plot the heat map in zip code, latitude, and longitude showing very dark color coding and negative reflection that's why we dropped these columns from the data.

```
In [69]: #drop unwanted columns
telecom= telecom.drop(['Customer ID','Zip Code','Latitude','Longitude',
                       'Churn Category','Churn Reason'],axis=1)
telecom.head()
```

Out[69]:

	Gender	Age	Married	Number of Dependents	City	Number of Referrals	Tenure in Months	Offer	Phone Service	Avg Monthly Long Distance Charges	...	C
0	Female	37	Yes	0	Frazier Park	2	9	None	Yes	42.39	...	
1	Male	46	No	0	Glendale	0	9	None	Yes	10.69	...	t
2	Male	50	No	0	Costa Mesa	0	4	Offer E	Yes	33.65	...	t
3	Male	78	Yes	0	Martinez	1	13	Offer D	Yes	27.82	...	t
4	Female	75	Yes	0	Camarillo	3	3	None	Yes	7.38	...	t

5 rows × 32 columns

Figure 8: Drop Unwanted Columns

Now again we dropped the rows that contained missing values and used the drop function.

```
In [70]: #drop missing rows
telecom.dropna(inplace=True)
```

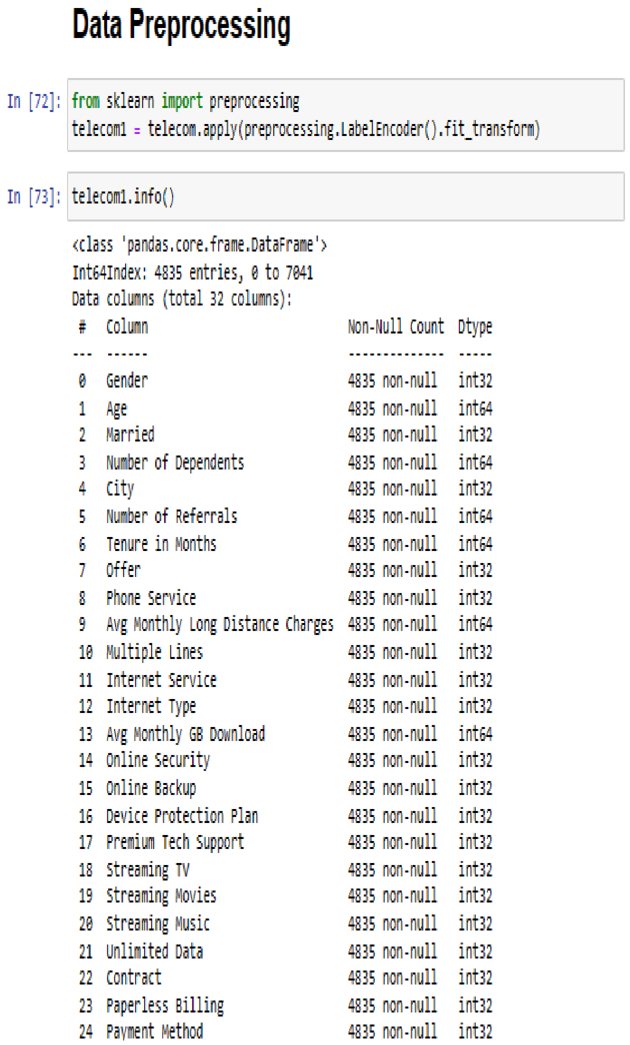
```
In [71]: telecom.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4835 entries, 0 to 7041
Data columns (total 32 columns):
 #   Column                             Non-Null Count  Dtype
---  ------                             --------------  -----
 0   Gender                             4835 non-null   object
 1   Age                                4835 non-null   int64
 2   Married                            4835 non-null   object
 3   Number of Dependents               4835 non-null   int64
 4   City                               4835 non-null   object
 5   Number of Referrals                4835 non-null   int64
 6   Tenure in Months                   4835 non-null   int64
 7   Offer                              4835 non-null   object
 8   Phone Service                      4835 non-null   object
 9   Avg Monthly Long Distance Charges  4835 non-null   float64
 10  Multiple Lines                     4835 non-null   object
 11  Internet Service                   4835 non-null   object
```

Figure 9: Drop Missing Value Rows

V. DATA PREPROCESSING

Data preprocessing is a crucial step in data analysis to transform the data. This step contains data encoding which converts the categorical data into a numerical representation. In this step, we apply the label encoding to transform the data.



Data Preprocessing

```
In [72]: from sklearn import preprocessing
         telecom1 = telecom.apply(preprocessing.LabelEncoder().fit_transform)

In [73]: telecom1.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 4835 entries, 0 to 7041
Data columns (total 32 columns):
 #   Column                             Non-Null Count  Dtype
---  ------                             --------------  -----
 0   Gender                             4835 non-null   int32
 1   Age                                4835 non-null   int64
 2   Married                            4835 non-null   int32
 3   Number of Dependents               4835 non-null   int64
 4   City                               4835 non-null   int32
 5   Number of Referrals                4835 non-null   int64
 6   Tenure in Months                   4835 non-null   int64
 7   Offer                              4835 non-null   int32
 8   Phone Service                      4835 non-null   int32
 9   Avg Monthly Long Distance Charges  4835 non-null   int64
 10  Multiple Lines                     4835 non-null   int32
 11  Internet Service                   4835 non-null   int32
 12  Internet Type                      4835 non-null   int32
 13  Avg Monthly GB Download            4835 non-null   int64
 14  Online Security                    4835 non-null   int32
 15  Online Backup                      4835 non-null   int32
 16  Device Protection Plan             4835 non-null   int32
 17  Premium Tech Support               4835 non-null   int32
 18  Streaming TV                       4835 non-null   int32
 19  Streaming Movies                   4835 non-null   int32
 20  Streaming Music                    4835 non-null   int32
 21  Unlimited Data                     4835 non-null   int32
 22  Contract                           4835 non-null   int32
 23  Paperless Billing                  4835 non-null   int32
 24  Payment Method                     4835 non-null   int32
```

Figure 10: Data Preprocessing.

VI. DATA VISUALIZATION

Data visualization is a representation of data in a graphical format. It plots the various elements such as charts, graphs, and maps to provide patterns and make decision-making. The main goal of implementing data visualization is to make difficult datasets understandable and accessible for makers to make decisions for their productivity. It plays an important role in various fields such as healthcare, education, telecom, e-commerce, and finance. In the telecom dataset, we perform the data visualization and analyze the data information in a graphical format.

This is the first graph that is histogram using the plotly library. This graph represents the information on the churn category and how many categories are present with how many count values. Competitors have a high amount of categories to perform the churn prediction.

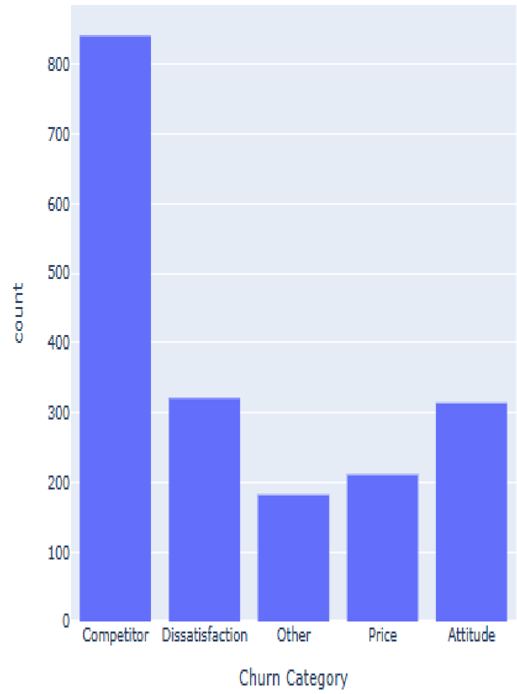


Figure 11: Histogram of Churn Category

This plot displays the frequency of churn reasons through a count representation. According to the analysis, competitors had better devices, more responsible for telecom churn.

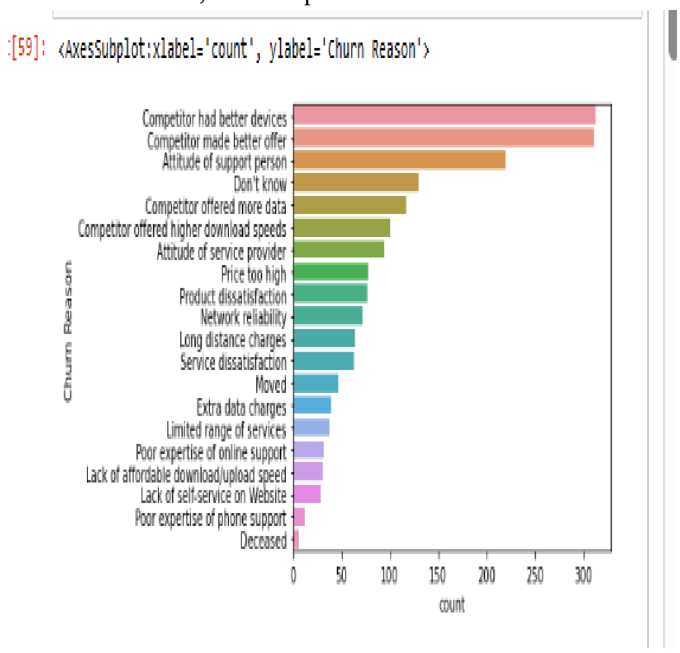


Figure 12: Count Plot of Column Churn Reason

This plot displays the frequency of gender through a count representation.

```
In [60]: #count plot of gender
sns.countplot(x ='Gender', data = telecom)
plt.show()
```

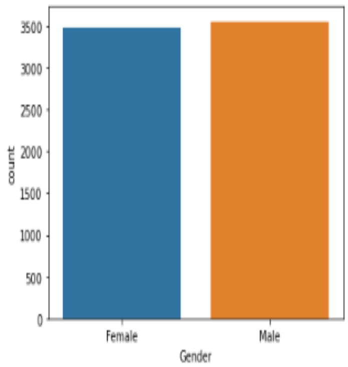


Figure 13: Count Plot of Gender

This plot displays the frequency of customer status through a count representation.

```
In [61]: #count plot of customer status
sns.countplot(x ='Customer Status', data = telecom)
plt.show()
```

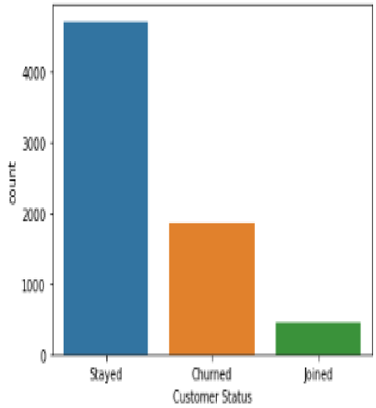


Figure 14: Count Plot of Column Customer Status

This plot displays the frequency of offers through a count plot.

```
In [62]: #count plot of offer
sns.countplot(x ='Offer', data = telecom)
plt.show()
```

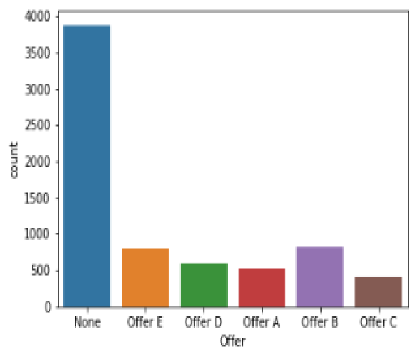


Figure 15: Count Plot of Column Offer

This plot displays the frequency of payment methods through a count representation.

```
53]: #count plot of payment method
sns.countplot(x ='Payment Method', data = telecom)
plt.show()
```

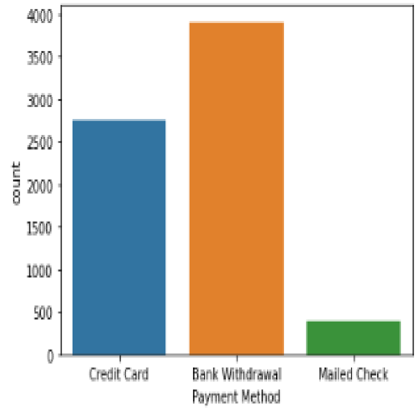


Figure 16: Count Plot of Column Payment Method

This plot displays the frequency of internet types through a count representation.

```
In [64]: #count plot of internet type
sns.countplot(x ='Internet Type', data = telecom)
plt.show()
```

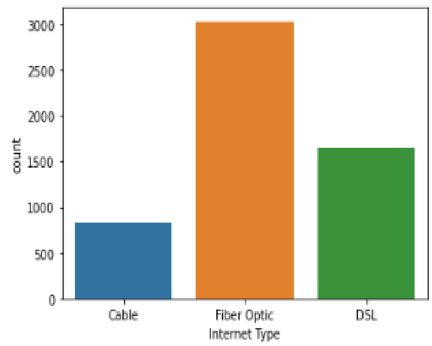


Figure 17: Count Plot of Column Internet Type

This plot displays the frequency of internet services through count representation.

```
In [65]: #count plot of internet services
sns.countplot(x ='Internet Service', data = telecom)
plt.show()
```

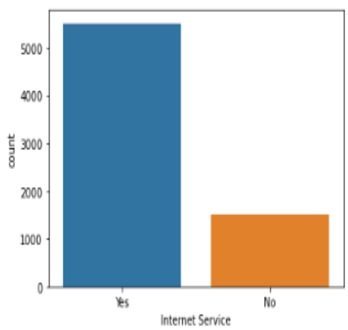


Figure 18: Count Plot of Column Internet Services

This plot displays the frequency of marital status through a count representation.

```
In [66]: sns.countplot(x ='Married', data = telecom)
         plt.show()
```

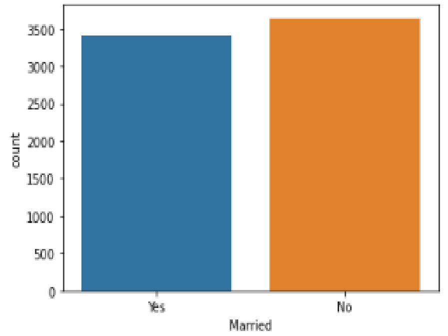


Figure 19: Count Plot of Column Married

We have plotted the heat map to show the correlation matrix of all features. According to the graph, we understand which features are highly relatable to each other and useful for further analysis and which attributes are less relatable and not useful for the machine learning models.

```
&lt;AxesSubplot:&gt;
```

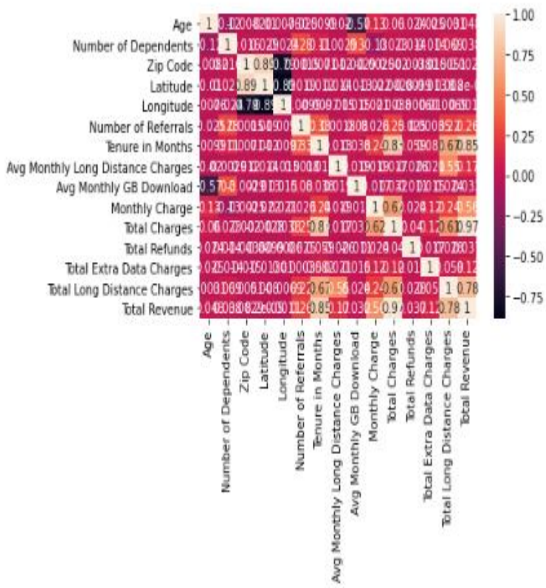


Figure 20: Heat Map

VII. MODELS

Model Evaluation is a part where we apply machine learning models such as logistic regression, XG boost, and decision tree and compare the performance score. This dataset is a type of multi-classification. We take the Customer status as a dependent variable and others as independent variables. We have split the data into training and test parts. We define the test size as 20%.

VIII. LOGISTIC REGRESSION

Logistic regression is a supervised model with is worked with binary as well as multi-classification. In this dataset, we applied this model because of multi-classification. This model is famous for providing better accurate results and being computationally efficient. Logistic regression helps to predict results which is helpful for telecom companies to identify the key drivers that are contributing to telecom churn. Using this also they can continuously monitor the customer churn risk and timely take actions against them.

Logistic Regression

```
In [77]: from sklearn.linear_model import LogisticRegression
         classifier=LogisticRegression()
         classifier.fit(x_train,y_train)

C:\Users\SR\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:762:
ConvergenceWarning:

lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regressi
on

Out[77]: LogisticRegression()

In [78]: pred1 = classifier.predict(x_test)

In [79]: from sklearn.metrics import accuracy_score,classification_report
         score=accuracy_score(pred1,y_test)
         print(score)

         0.734229576008273
```

Figure 21: Logistic Regression

IX. XG BOOST

XG Boost is a supervised learning algorithm that is a part of the gradient boosting algorithm. Its main aim is to combine multiple weak learners and provide a strong learner with more accurate results. Its objective function is loss and entropy here because of the classification factor. It uses a learning rate to scale each tree for better prediction.

XG Boost

```
n [80]: import xgboost as xgb
        model = xgb.XGBClassifier()
        model.fit(x_train, y_train)
        pred2 = model.predict(x_test)

C:\Users\SR\anaconda3\lib\site-packages\xgboost\sklearn.py:1224: UserWarning:

The use of label encoder in XGBClassifier is deprecated and will be removed in
a future release. To remove this warning, do the following: 1) Pass option use_
label_encoder=False when constructing XGBClassifier object; and 2) Encode your
labels (y) as integers starting with 0, i.e. 0, 1, 2, ..., [num_class - 1].

[20:58:12] WARNING: C:/Users/Administrator/workspace/xgboost-win64_release_1.5.
1/src/learner.cc:1115: Starting in XGBoost 1.3.0, the default evaluation metric
used with the objective 'multi:softprob' was changed from 'merror' to 'mloglos
s'. Explicitly set eval_metric if you'd like to restore the old behavior.

n [81]: from sklearn.metrics import accuracy_score,classification_report
        score1=accuracy_score(pred2,y_test)
        print(score1)

        0.8221302998965874
```

Figure 22: XG Boost

X. DECISION TREE

A decision tree is also a supervised learning algorithm. It took the whole dataset as a starting point and then selected the best features to divide the dataset. It judges the accuracy score by checking the Gini impurity, information gain, and entropy. This model is very easy and useful. It tends to handle numerical and categorical datasets. It tends to perform the feature selection as well. But if we make small changes in the tree then it will directly affect the complete tree.

Figure 23: Decision Tree

XI. CONCLUSION

To keep the customer satisfied is very important to make the telecom business successful because it depends on the services we are providing in telecommunication. In this research analysis, telecom churn data is collected from Kaggle which is a customer data of California. Python language is used to predict customer churn in the telecom industry. Data Visualization is also performed to plot the charts to understand the relationship in variables and understand its variation. Some machine learning techniques such as logistic regression, XG boost, and decision algorithms are implemented on the data. The accuracy score of logistic regression is 73%. The accuracy score of the xg boost is 82%. The accuracy score of the decision tree is 74%. According to the result analysis, XG boost performs better to predict better telecom churn results and helps to improve the growth of telecommunication.

XII. FUTURE SCOPE

The future scope of telecom churn prediction is to use social network analysis for better accurate results. We can also use the hyperparameters in the future at the time of model building and use the randomized search CV or optimization techniques. Continuing refinement of hyperparameters helps to improve the performance and increase the decision-making process. The telecom churn system may be complex in the future for precise recommendations. For this, we can use deep learning techniques to calculate the estimate of survival likelihood.

REFERENCES

1. Aldalan, A.M. and Almaleh, A., 2023. Customer Churn Prediction Using Four Machine Learning Algorithms Integrating Feature Selection and Normalization in the Telecom Sector. International Journal of Electronics and Communication Engineering, 17(3), pp.76-83.
2. Sam, G., Asuquo, P. and Stephen, B., 2024. Customer Churn Prediction using Machine Learning Models. Journal of Engineering Research and Reports, 26(2), pp.181-193.
3. Fujo, S.W., Subramanian, S. and Khder, M.A., 2022. Customer churn prediction in telecommunication industry using deep learning. Information Sciences Letters, 11(1), p.24.
4. Saheed, Y.K. and Hambali, M.A., 2021, October. Customer churn prediction in telecom sector with machine learning and information gain filter feature selection algorithms. In 2021 International Conference on Data Analytics for Business and Industry (ICDABI) (pp. 208-213). IEEE.
5. Kavitha, V., Kumar, G.H., Kumar, S.M. and Harish, M., 2020. Churn prediction of customer in telecom industry using machine learning algorithms. International Journal of Engineering Research & Technology (2278-0181), 9(05), pp.181-184.
6. Jain, H., Yadav, G. and Manoov, R., 2020. Churn prediction and retention in banking, telecom and IT sectors using machine learning techniques. In Advances in Machine Learning and Computational Intelligence: Proceedings of ICMLCI 2019 (pp. 137-156). Singapore: Springer Singapore.
7. Prabadevi, B., Shalini, R. and Kavitha, B.R., 2023. Customer churning analysis using machine learning algorithms. International Journal of Intelligent Networks, 4, pp.145-154.
8. Beeharry, Y. and Tsokizep Fokone, R., 2022. Hybrid approach using machine learning algorithms for customers' churn prediction in the telecommunications industry. Concurrency and Computation: Practice and Experience, 34(4), p.e6627.
9. Mahalekshmi, A. and Chellam, G.H., 2022. Analysis of customer churn prediction using machine learning and deep learning algorithms. International Journal of Health Sciences, (I), pp.11684-11693.
10. Gopal, P. and MohdNawi, N.B., 2021, December. A Survey on Customer Churn Prediction using Machine Learning and data mining Techniques in E-commerce. In 2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) (pp. 1-8). IEEE.
11. Dalli, A., 2022. Impact of hyperparameters on deep learning model for customer churn prediction in telecommunication sector. Mathematical Problems in Engineering, 2022, pp.1-11.