TELECOM CHURN PREDICTION

Asmita Sharma

INTRODUCTION

Churn prediction is one of the most popular Big Data use cases in business. It consists of detecting customers who are likely to cancel a subscription to a service.

Churn is a problem for telecom companies because it is more expensive to acquire a new customer than to keep your existing one from leaving.

Wireless companies today measure voluntary churn by a monthly figure, such as 1.9 percent or 2.1 percent.

PROJECT OBJECTIVE

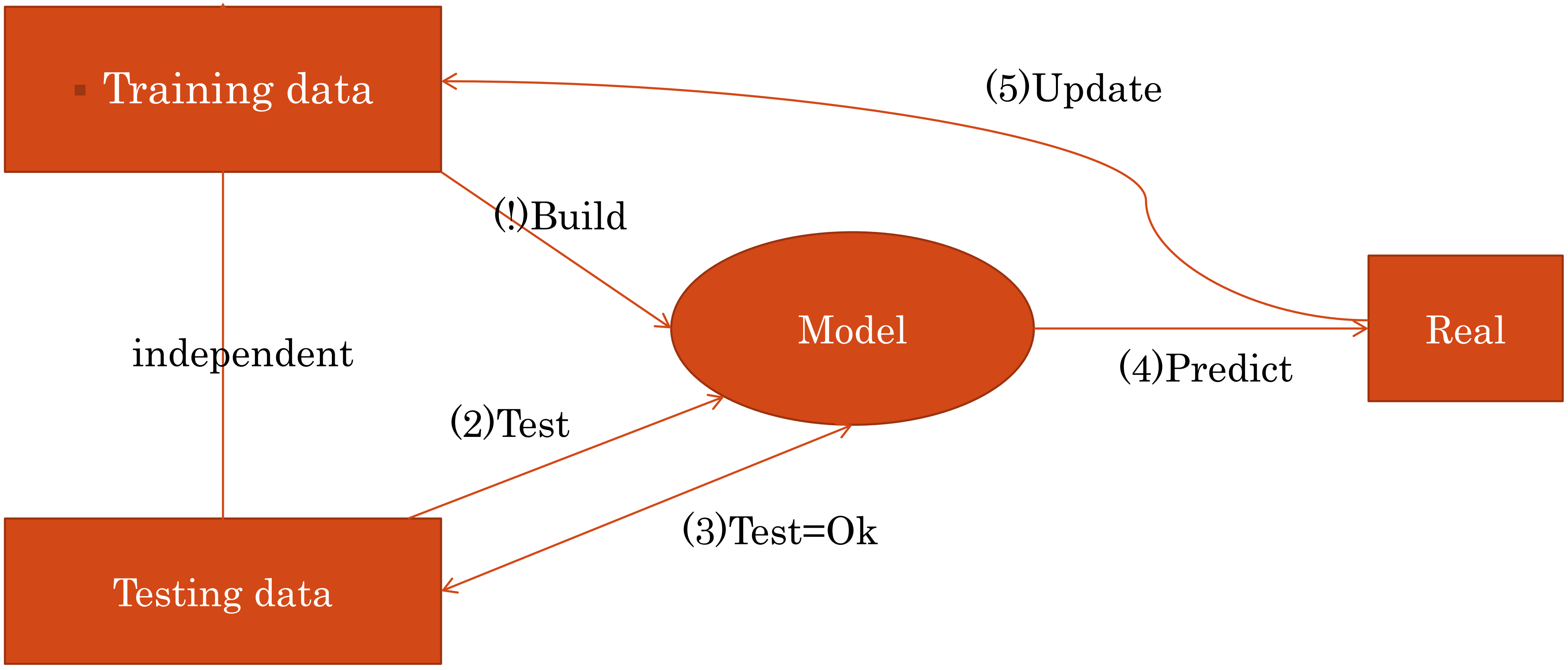
There are many ways: better products, better delivery methods, lower prices, building satisfactory customer relationships, better marketing and, above all, successful customer communications.

- To predict Customer Churn.
- Highlighting the main variables/factors influencing Customer Churn.
- Use various ML algorithms to build prediction models, evaluate the accuracy and performance of these models.
- Finding out the best model for our business case & providing executive summary.

DATASET DESCRIPTION

- Source dataset is in csv format.
- Dataset contains 7043 rows and 14 columns
- There is no missing values for the provided input dataset.
- Churn is the variable which notifies whether a particular customer is churned or not. And we will be developing our models to predict

CHURN PREDICTION MODEL



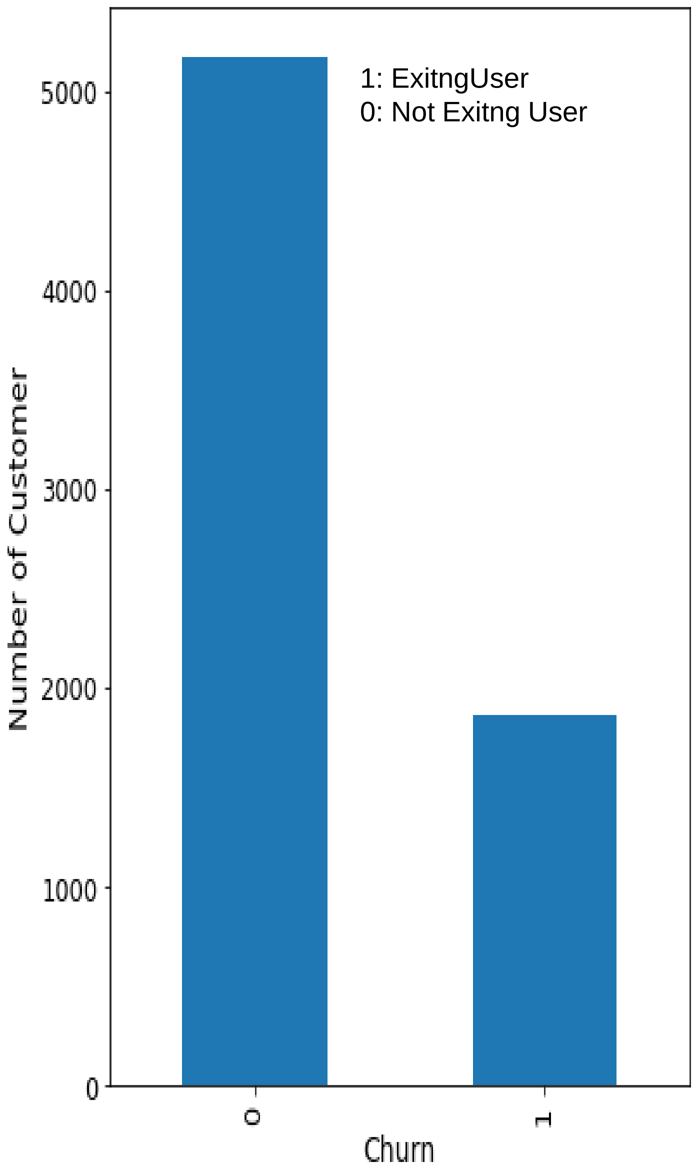
METHODOLOGIES

- EDA(Exploratory Data Analysis): The dataset consists of 12 variables in all. A few are continuous, rest are categorical. The control variable was customer.
- Model building which includes defining the purpose if model, determine the model boundary, build the model, create an interface and export the model.
- Evaluating machine learning algorithm is an essential part of project.

EXPLORATORY DATA ANALYSIS

- Data visualisation using seaborn and matplotlib
- Exploratory data analysis (EDA) is an approach to analyse data sets & to summarize their main characteristics, often with visual methods.
- A Statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modelling or hypothesis.

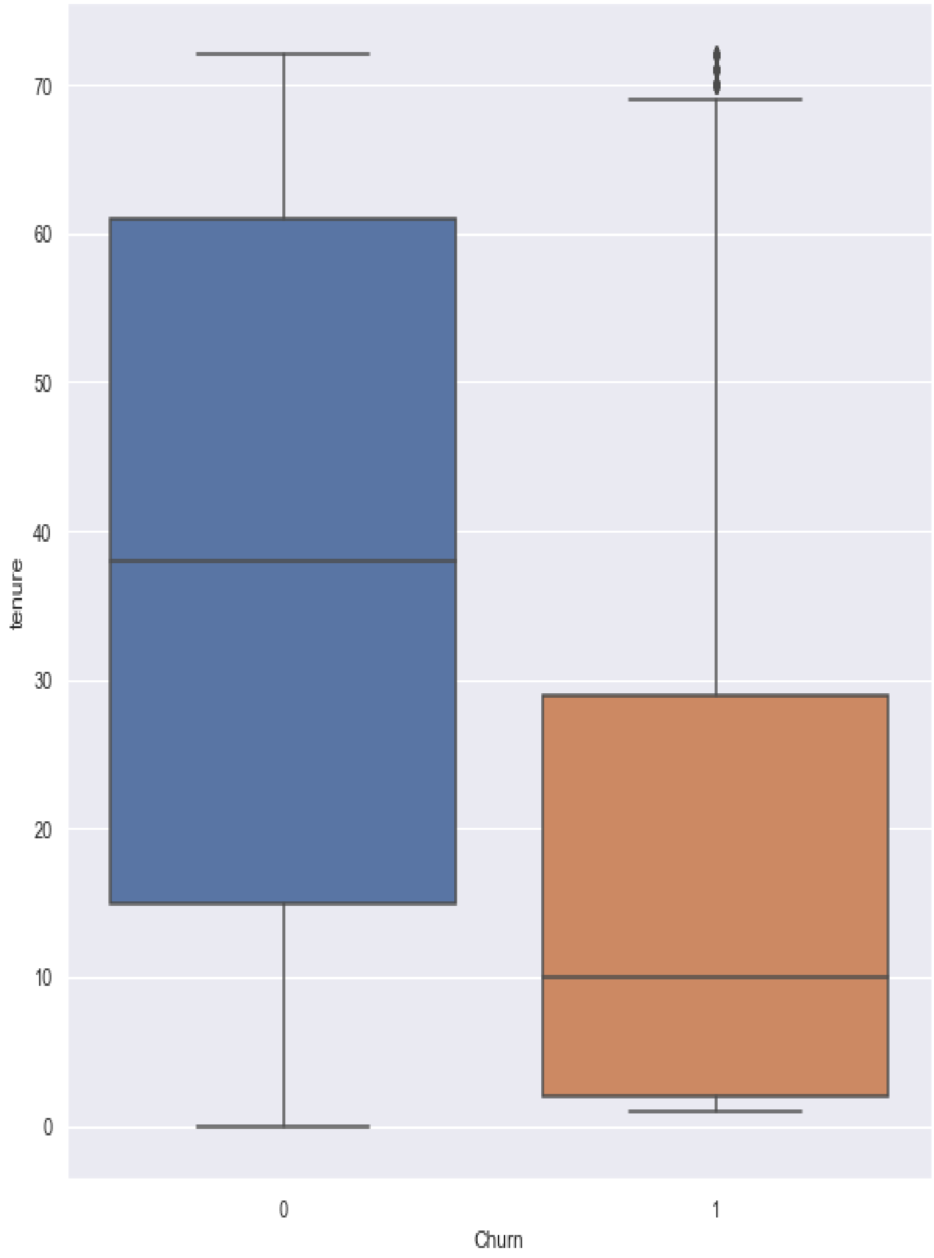
```
ax = dataset1["Churn"].value_counts().plot(kind='bar', figsize=(6, 8), fontsize=13)
ax.set_ylabel("Number of Customer", fontsize=14);
ax.set_xlabel("Churn", fontsize=14);
```



BAR GRAPH

Plot Shows that the Users from the Data are likely to be Continuing their Subscription plan(>70%)

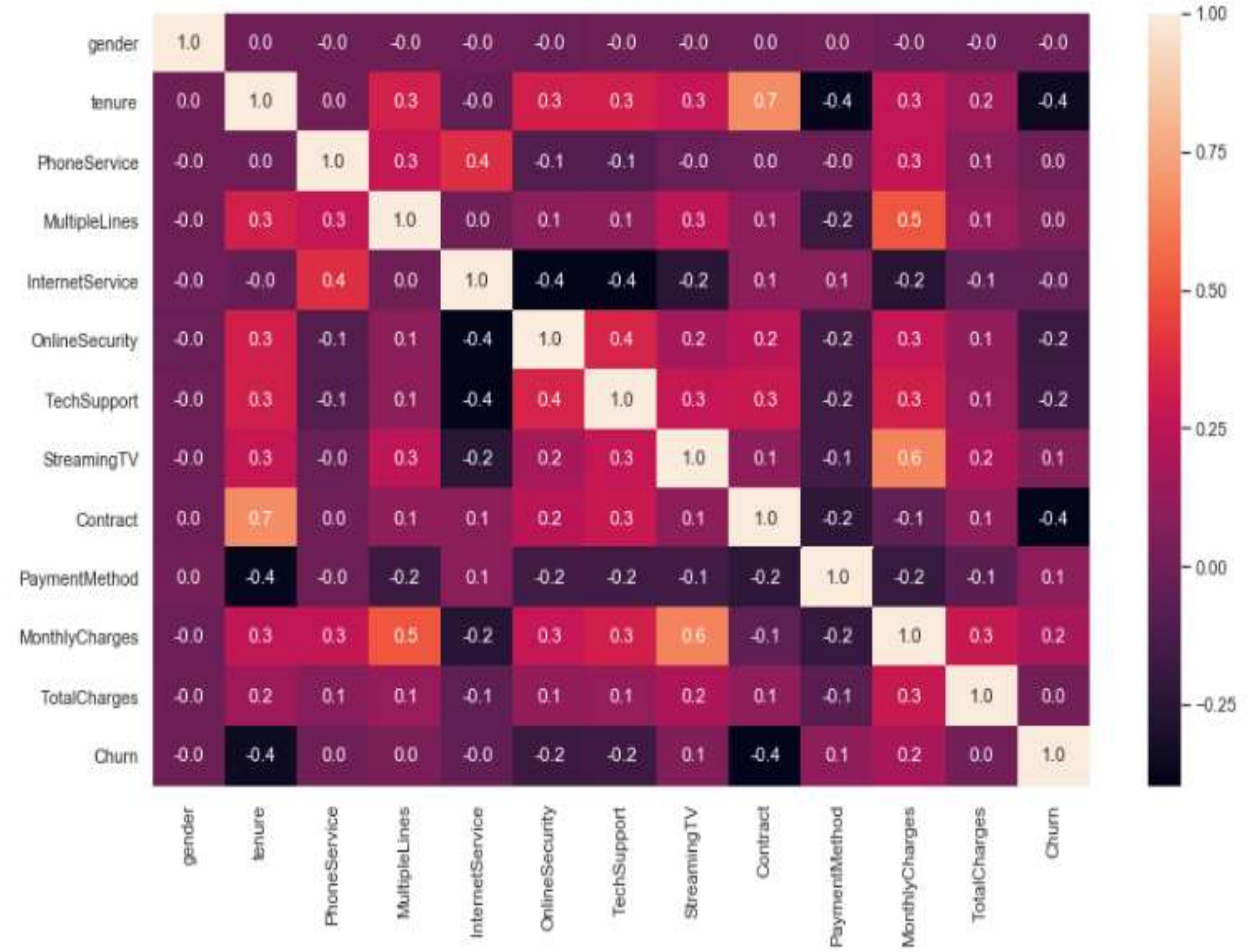
In [72]:
```
sns.boxplot(x='Churn', y='tenure', data=dataset1)
sns.set(rc={'figure.figsize':(10,8)})
```



BOX PLOT

- We found outliers in exiting customers which is out of whiskers. An Outlier is an observation that is numerically distant from the rest of the data.
- Using Skew() method we found that Churn data is inconsistent with tenure
- Customers who disconnecting their subscription plans are selecting short tenure Telecom Company need to offer better plans for those customers who choose short tenures.

```python
plt.figure(figsize = (14, 8))
sns.heatmap(dataset1.corr(), annot=True, fmt=".1f")
plt.show()
```



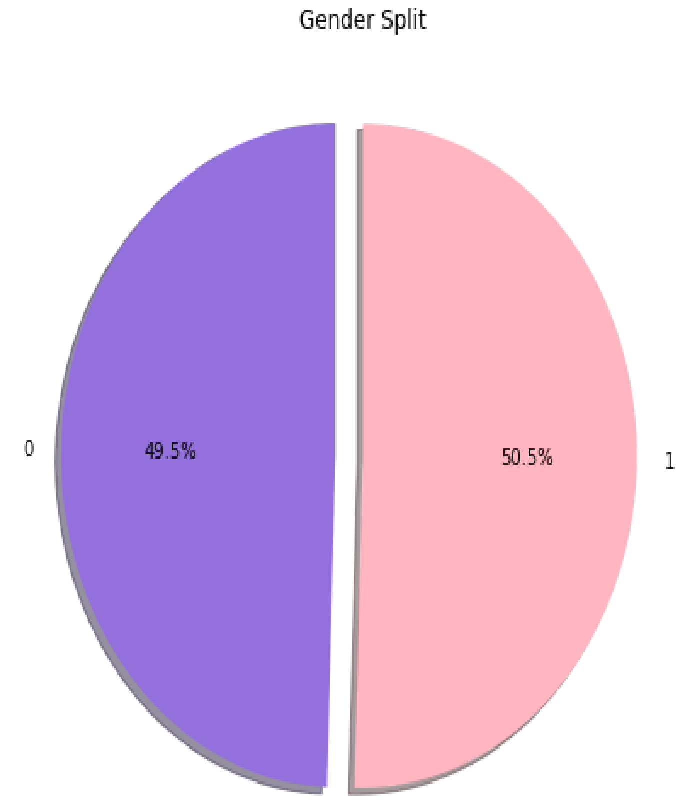
HEAT MAP

Correlation: Dependence or association is any statistical relationship, whether causal or not, between two random variables or bivariate data.
With the help of Correlation matrix, we can find interdependency between variables

1)Least dependency of variables for predicting churn are tenure and contract.
2)Churn variable is depending more on monthly charges.

```
labels, values = zip(*Counter(dataset1["gender"]).items())
colors = ['mediumpurple', 'lightpink']
piechart_df = (pd.DataFrame(list(values),list(labels)))
piechart_df = piechart_df.reset_index()
piechart_df
fig = plt.figure(figsize=[6, 6])

plt.pie(piechart_df[0],labels=piechart_df["index"],startangle=90,explode=(0.1,0),autopct="%1.1f%%", shadow = True, colors=colors
plt.tight_layout()
plt.title("Gender Split")
plt.show()
```



PIE PLOT

1)Female subscribers are 49.5% of the total

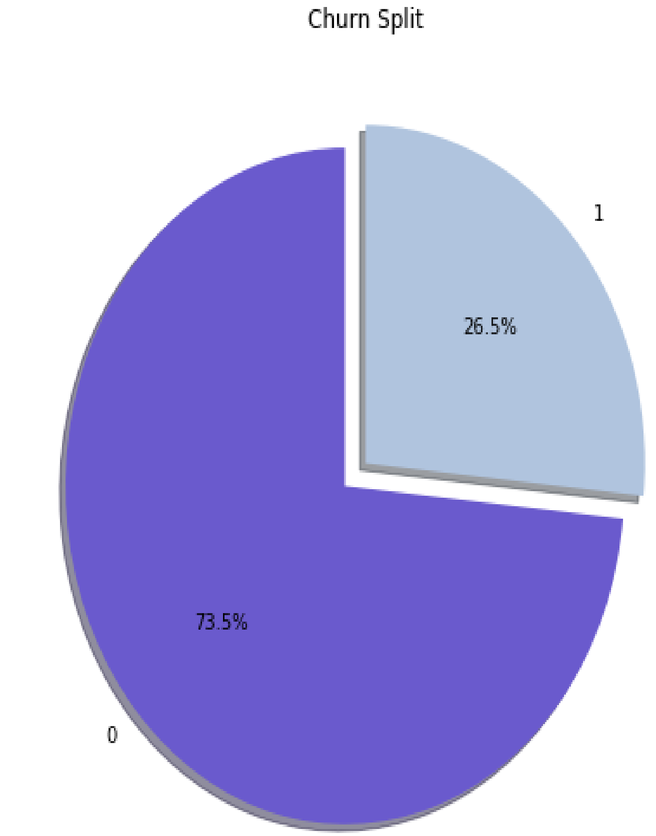
2)Male subscribers are 55.5% of the total

```python
from collections import Counter


labels, values = zip(*Counter(dataset1["Churn"]).items())
colors = ['slateblue', 'lightsteelblue']
piechart_df = (pd.DataFrame(list(values),list(labels)))
piechart_df = piechart_df.reset_index()

fig = plt.figure(figsize=[6, 6])

plt.pie(piechart_df[0],labels=piechart_df["index"],startangle=90,explode=(0.1,0),autopct="%1.1f%%", shadow=True, colors=colors)
plt.tight_layout()
plt.title("Churn Split")
plt.show()
```

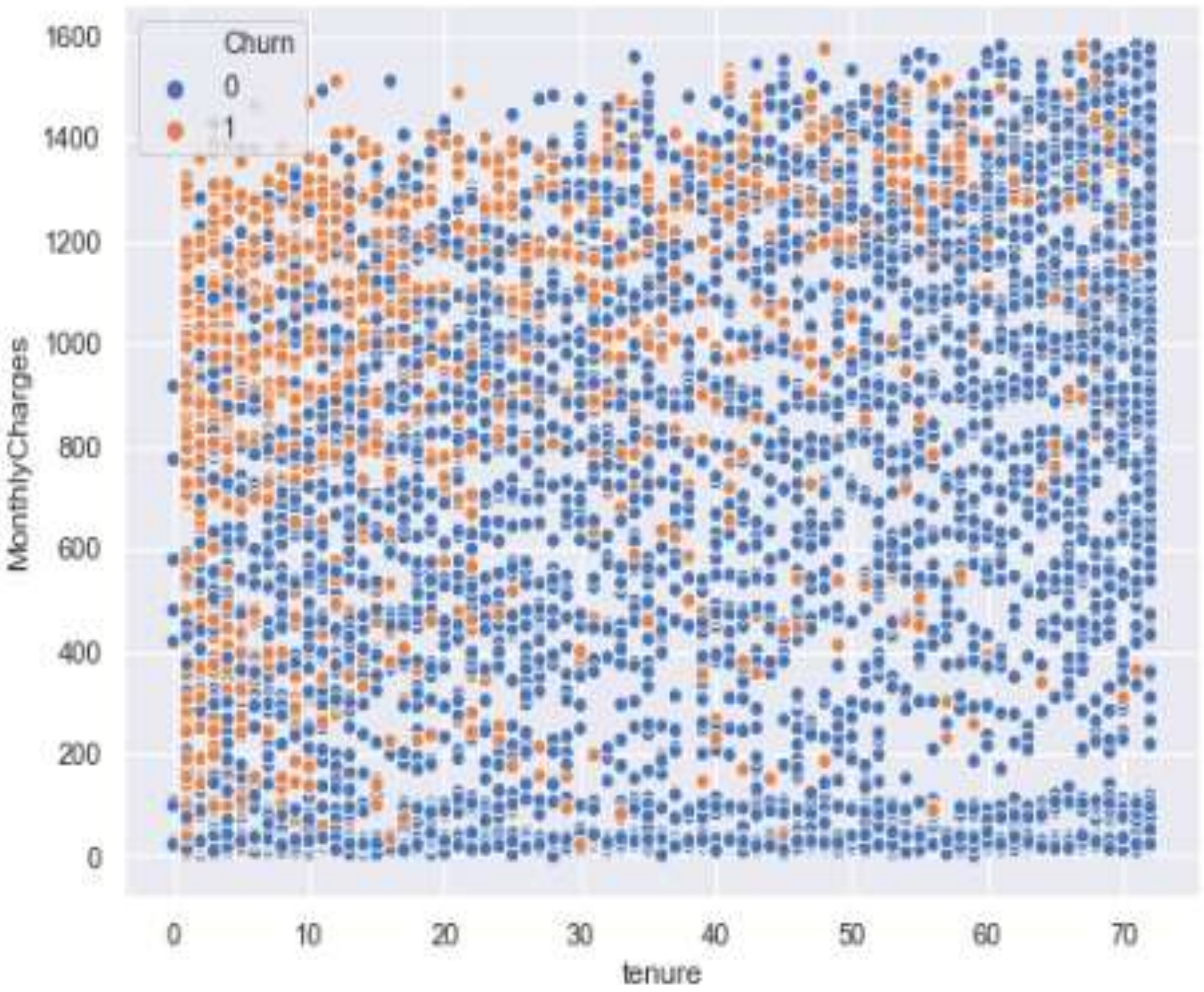


PIE PLOT

According to the collected data information 73.5% customers are continuing their subscription and 26.5% customers disconnected.

```
plt.figure(figsize=(8, 6))
sns.scatterplot(x = 'tenure', y = 'MonthlyCharges',hue="Churn" ,data = dataset1)
```

```
]: <matplotlib.axes._subplots.AxesSubplot at 0x26af16b8a90>
```



SCATTER PLOT

1)Customers paying high monthly charges for short tenures are disconnecting
2)Customers paying high monthly charges for long tenures continuing with their subscription plans, as it is reasonable cost

ACCURACY OF VARIOUS MODELS

MODELS	ACCURACY
KNN	62%
SVM	76%

SVM MODEL

```
#SVM
from sklearn.svm import SVC
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.3,random_state = 0)

classifier = SVC(kernel = 'rbf')
classifier.fit(x_train, y_train)
y_pred = classifier.predict(x_test)
```

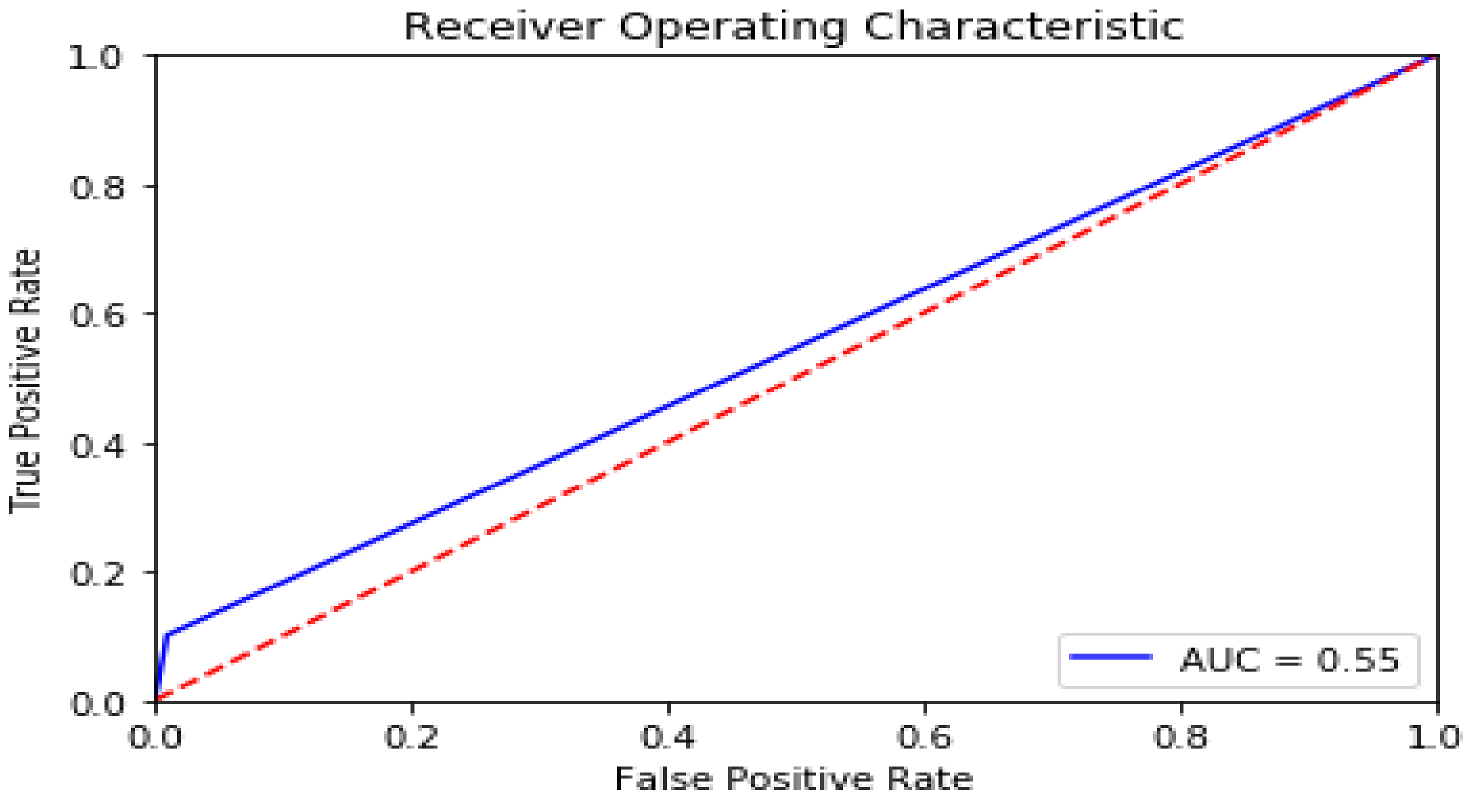
```python
accuracy1 = metrics.accuracy_score(y_test, y_pred)
print("Accuracy:",accuracy1)
```

```
Accuracy: 0.7597955706984668
```

```python
print("Precision:",metrics.precision_score(y_test, y_pred))
print("Recall:",metrics.recall_score(y_test, y_pred))
```

```
Precision: 0.7796610169491526
Recall: 0.10143329658213891
```

```
#plotting the roc scalar
plt.title('Receiver Operating Characteristic')
plt.plot(fpr, tpr, 'b', label = 'AUC = %0.2f' % roc_auc)
plt.legend(loc = 'lower right')
plt.plot([0, 1], [0, 1],'r--')
plt.xlim([0, 1])
plt.ylim([0, 1])
plt.ylabel('True Positive Rate')
plt.xlabel('False Positive Rate')
plt.show()
```



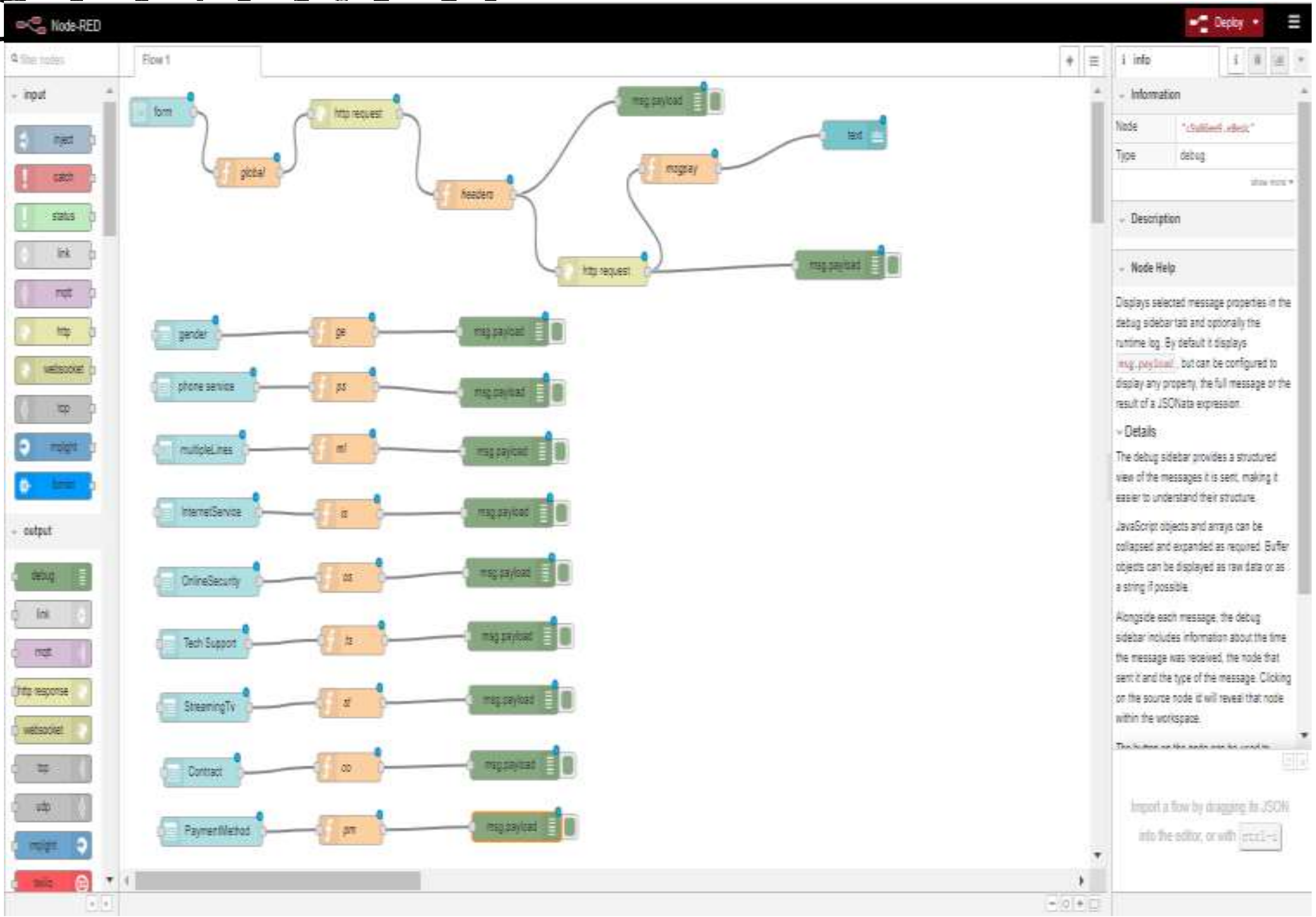
METRICS EVALUATION:

CONFUSION MATRIX:

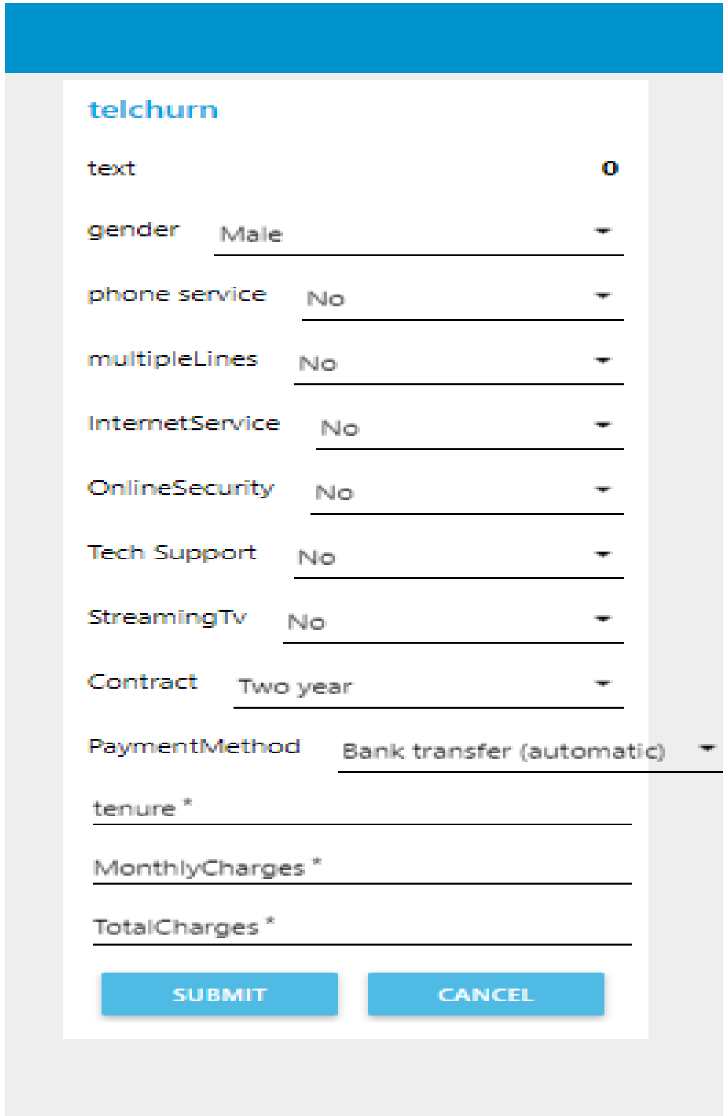
2589	26
815	92

ACCURACY: 76.12%
PRECISION : 77.96%

NODE RED(IBM WATSON STUDIO)



NODE RED FLOW WIRING



UI OF MODEL

FINDINGS AND SUGGESTIONS

- Try to offer the better service for the churn customers ,see how much this impact before and later .Some may use your service better move them to your active customers.
- Take the feedback and suggestions with in period of time and improve it ,strive for better communication.
- When your are taking the any change in plans of your business just predict the positive and negative share of that plan. If it is negative prepare the solution before so You can handy easily.

HOW TO REDUCE CUSTOMER CHURN

- Lean into your best customers.
- Be proactive with communication.
- Define a roadmap for your new customers.
- Offer incentives.
- Ask for feedback often.
- Analyze churn when it happens.
- Stay competitive.

CONCLUSION

- The importance of this type of research in the telecom market is to help companies make more profit.
- It has become known that predicting churn is one of the most important sources of income to Telecom companies.
- Hence, this research aimed to build a system that predicts the churn of customers i telecom company.
- These prediction models need to achieve high AUC values. To test and train the model, the sample data is divided into 70% for training and 30% for testing.