Telco Customer Churn Prediction Using Machine Learning:

- **Author : Harsh Priyam**
- **Submitted on : 15th February 2023**

1. Problem Statement :

Telecommunications companies face a number of challenges when it comes to managing customer churn, or the loss of subscribers. Some of the key problems in this area include :

Competition: Telecommunications companies face intense competition, with many different providers offering similar services. As a result, it can be difficult to retain customers who may be swayed by competitors' promotions or offers

Poor customer service: Poor customer service can be a major cause of churn, especially if customers are dissatisfied with the response they receive to their inquiries or complaints.

Technical issues: Technical problems, such as service outages or slow data speeds, can also contribute to churn, especially if they are not resolved quickly.

Inadequate customer insights: Without access to detailed customer data and insights, it can be difficult for telecommunications companies to understand the drivers of churn and develop effective retention strategies.

2. Market/Customer/Business Need Assessment:

The demand for solutions to manage customer churn in the telecommunications industry has been growing in recent years. This is due to a number of factors, including :

Increased competition: As the telecommunications market becomes increasingly competitive, companies are under pressure to retain their existing customers and prevent them from switching to competitors.

The need for cost-effective solutions: Telecommunications companies are facing pressure to reduce costs, and reducing customer churn can be a cost-effective way to achieve this. By retaining existing customers, companies can avoid the costs associated with acquiring new ones.

The importance of customer data: Companies are recognizing the importance of customer data in managing churn. By analyzing customer data, they can gain insights into customer behavior and preferences, and develop targeted strategies to retain customers.

Customer expectations: Customers are becoming more demanding and expect high-quality, personalized services from their telecommunications provider. Companies that fail to meet these expectations are at risk of losing customers to competitors.

So, this type of machine learning application can be a game changer and create a great market value for the telecommunication company.

3. Target Specifications and Characterization :

Target specifications and characterization are important for telecommunication companies, as they define the goals and objectives of the company and the characteristics of the customers they want to target. Some key target specifications and characterizations for telecommunication companies include:

Market segmentation: Telecommunication companies can use market segmentation to identify and target specific customer groups, such as business customers or consumers, who have specific needs and preferences.

Customer value: Telecommunication companies can target customers based on their value to the company, such as those who generate the highest revenue or have the longest tenure with the company

Customer demographics: Companies can target specific customer demographics, such as age, gender, income, or location, to better understand and meet the needs of their target market.

Technology and product offerings: Companies can target customers based on the technology and product offerings they are interested in, such as 5G or broadband services.

4. External search (information sources):

The dataset can be found on the Kaggle. The dataset was created in a project that aims to predict the behavior to retain customers. Here we can analyze all relevant customer data and develop focused customer retention programs.
Each row represents a customer, each column contains customer's attributes described on the column Metadata.

The data set includes information about:

- Customers who left within the last month – the column is called Churn
- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges

- Demographic info about customers – gender, age range, and if they have partners and dependents

Dataset Origin :

-[https://www.kaggle.com/datasets/blastchar/telco- customer-churn](https://www.kaggle.com/datasets/blastchar/telco-customer-churn)

[-https://data.gov.in/search?title=telecomm](https://data.gov.in/search?title=telecomm)

See some information about our dataset:

2/10/23, 4:45 PM Telco Customer Churn - Jupyter Notebook

In [451]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [452]:
```
pd.set_option('display.max_rows',None)
pd.set_option('display.max_columns',None)
```

In [453]:
```
df = pd.read_csv('Telco-Customer-Churn.csv')
```

In [454]:
```
df.head()
```

Out[454]:

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	Ir
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	
1	5575-GNVDE	Male	0	No	No	34	Yes	No	
2	3668-QPYBK	Male	0	No	No	2	Yes	No	
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	
4	9237-HQITU	Female	0	No	No	2	Yes	No	

In [455]:
```
df.shape
```

Out[455]: (7032, 21)

In [456]:
```
df.columns
```

Out[456]:
```
Index(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents',
       'tenure', 'PhoneService', 'MultipleLines', 'InternetService',
       'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport',
       'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling',
       'PaymentMethod', 'MonthlyCharges', 'TotalCharges', 'Churn'],
      dtype='object')
```

In [458]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7032 entries, 0 to 7031
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   customerID        7032 non-null   object 
 1   gender            7032 non-null   object 
 2   SeniorCitizen     7032 non-null   int64  
 3   Partner           7032 non-null   object 
 4   Dependents        7032 non-null   object 
 5   tenure            7032 non-null   int64  
 6   PhoneService      7032 non-null   object 
 7   MultipleLines     7032 non-null   object 
 8   InternetService   7032 non-null   object 
 9   OnlineSecurity    7032 non-null   object 
 10  OnlineBackup      7032 non-null   object 
 11  DeviceProtection  7032 non-null   object 
 12  TechSupport       7032 non-null   object 
 13  StreamingTV       7032 non-null   object 
 14  StreamingMovies   7032 non-null   object 
 15  Contract          7032 non-null   object 
 16  PaperlessBilling  7032 non-null   object 
 17  PaymentMethod     7032 non-null   object 
 18  MonthlyCharges    7032 non-null   float64
 19  TotalCharges      7032 non-null   float64
 20  Churn             7032 non-null   object 
dtypes: float64(2), int64(2), object(17)
memory usage: 1.1+ MB
```


In [459]:
```
df.describe()
```

Out[459]:

	SeniorCitizen	tenure	MonthlyCharges	TotalCharges
count	7032.000000	7032.000000	7032.000000	7032.000000
mean	0.162400	32.421786	64.798208	2283.300441
std	0.368844	24.545260	30.085974	2266.771362
min	0.000000	1.000000	18.250000	18.800000
25%	0.000000	9.000000	35.587500	401.450000
50%	0.000000	29.000000	70.350000	1397.475000
75%	0.000000	55.000000	89.862500	3794.737500
max	1.000000	72.000000	118.750000	8684.800000

5. Benchmarking :

Benchmarking in telecommunication companies involves comparing a company's performance metrics and practices against those of competitors or industry standards. This helps identify areas for improvement and provides insights into best practices. Common metrics to benchmark in telecommunications include network coverage, call quality, customer satisfaction, and financial performance.

Some of the specific key points which makes my product better are listed below :-

Service offerings: My product may offer unique or innovative services that competitors do not have.

Network quality: My product can evaluate the quality of your network and compare it to that of other similar products. A more robust and reliable network can give your company an edge over competitors.

Customer service: Assess the quality of customer service and support offered by my product compared to other products. A customer-centric approach with excellent service can set your company apart from competitors.

Pricing: Evaluate the pricing of my product compared to other products. Offering competitive pricing may make your company more attractive to customers.

Innovation: Assess the level of innovation in my product compared to other similar products. A more innovative approach can help your company stay ahead of the curve and offer new and exciting products and services to customers.

By evaluating these factors and identifying areas where my product excels, It can differentiate my product from other similar products and highlight my competitive advantages to potential customers.

6. Applicable Regulations :

There are several regulations applicable to telecommunication companies. These regulations may vary by country, but some common ones include:

Licensing and spectrum allocation: Telecommunication companies need to obtain licenses from regulatory authorities to operate in a specific region. These licenses may also specify the frequency spectrum that a company can use for its services.
Data protection and privacy: Telecommunication companies must comply with data protection and privacy laws to safeguard the personal information of their customers.
Network neutrality: Some countries have regulations that require telecommunication companies to treat all internet traffic equally, without discrimination or blocking of specific content.
Quality of service: Regulatory authorities may set minimum standards for the quality of service that telecommunication companies must provide to their customers.
Interconnection and access: Telecommunication companies must comply with regulations that ensure fair and open access to their networks and services by other companies.
Tariffs and pricing: Regulatory authorities may set limits on the tariffs and pricing that telecommunication companies can charge for their services to ensure they are affordable and fair.
Consumer protection: Regulations are in place to protect consumers from misleading advertising, unfair contract terms, and other harmful business practices.

It's important for telecommunication companies to be aware of and comply with applicable regulations to avoid legal penalties and ensure they are providing their services in a fair and transparent manner.

7. Applicable Constraints and Carefulness :

Telecommunications companies operate in a highly regulated industry and are subject to a variety of constraints and requirements. Here are some of the most important considerations for telecommunications companies:

Data Privacy and Security: Telecommunications companies are responsible for protecting the sensitive personal and financial information of their customers. This requires them to implement robust data privacy and security measures, and to comply with a variety of laws and regulations related to data protection.

Environmental Impact: Telecommunications companies have an impact on the environment through the use of energy, water, and other resources. They must be careful to minimize their environmental impact and to comply with applicable laws and regulations related to environmental protection

Network infrastructure: Building and maintaining a reliable network infrastructure requires a significant investment, and companies need to carefully manage costs and plan for future growth.

Rapid technological changes: The telecommunications industry is rapidly evolving, and companies need to keep up with new technologies and trends to remain competitive.

8. Business Opportunity :

The telecommunications industry continues to grow and evolve, presenting new business opportunities for companies that are willing to take advantage of them. Some of the key opportunities in the telecommunications industry include:

5G Deployment: The rollout of 5G networks is creating new business opportunities for companies that can help with network deployment and the development of 5G-enabled devices and applications.

Internet of Things (IoT): The growth of IoT is creating new business opportunities for companies that can help connect devices and manage the massive amounts of data generated by these devices.

Cloud Services: The increasing use of cloud services is creating new business opportunities for companies that can help organizations move their IT systems to the cloud and manage their cloud-based operations.

Virtual and augmented reality: The rise of virtual and augmented reality technologies presents an opportunity for telecommunication companies to offer high-bandwidth network services and develop new applications and services that leverage these technologies.

9. Concept Development :

Concept development of a telecommunication churn model involves brainstorming and identifying the key components and factors that are relevant to predicting churn. Here are some steps for concept development of a telecommunication churn model:

Identify Churn: The first step is to define churn and identify the customer behaviors that indicate churn. Churn can be defined as the rate at which customers switch to a competitor or cancel their service. Customer behaviors that may indicate churn include reduced usage, increased complaints, and decreased satisfaction.

Define Key Metrics: The next step is to define the key metrics that will be used to measure churn. These metrics could include factors such as the number of customer complaints, the frequency of customer service calls, or the frequency of payment issues.

Data Collection: The third step is to collect relevant data from various sources such as call detail records, customer demographics, network quality, customer feedback, and competitor analysis. The data should be cleaned and pre-processed to remove any inconsistencies and errors.

Feature Engineering: The next step is to create relevant features that can help to predict churn. This involves identifying important variables, calculating metrics, and selecting the appropriate algorithms to generate the features.

Model Selection: The next step is to select the appropriate machine learning model that can accurately predict churn. Popular models include logistic regression, decision trees, random forests, and neural networks.

Model Training and Validation: The model is trained on the data and validated using various techniques such as cross-validation, AUC, and F1 score. This helps to identify the best performing model and fine-tune its parameters.

Model Deployment: Once the model is trained and validated, it is deployed in a production environment. The model should be integrated with the company's infrastructure, including its data storage, APIs, and web services.

Model Monitoring and Maintenance: The final step is to monitor the model's performance and maintain it over time. This involves identifying any issues, retraining the model with new data, and updating it with the latest algorithms and techniques.

These are some key steps involved in concept development of a telecommunication churn model. The specific details of each step may vary depending on the company's needs, resources, and infrastructure. It's important to work with a team of experts who have experience in developing churn models and can help to tailor the approach to the specific needs of the company.

10. Concept Generation

Telecommunication churn model is a predictive model that helps telecommunication companies to identify the customers who are likely to churn or switch to a competitor. Here are some concept generation ideas for a telecommunication churn model:

Customer Behavior: The model could analyze customer behavior such as call duration, data usage, and frequency of text messages. This analysis can help to predict when a customer is likely to churn, based on changes in their behavior.

Customer Demographics: Another concept is to analyze customer demographics such as age, gender, income level, location, and occupation. The model could determine which customer demographics are more likely to churn and target retention efforts accordingly.

Service Quality: The model could analyze customer complaints and service quality metrics such as call drop rate, network coverage, and response time. This analysis can help to predict when a customer is likely to churn based on a decline in service quality.

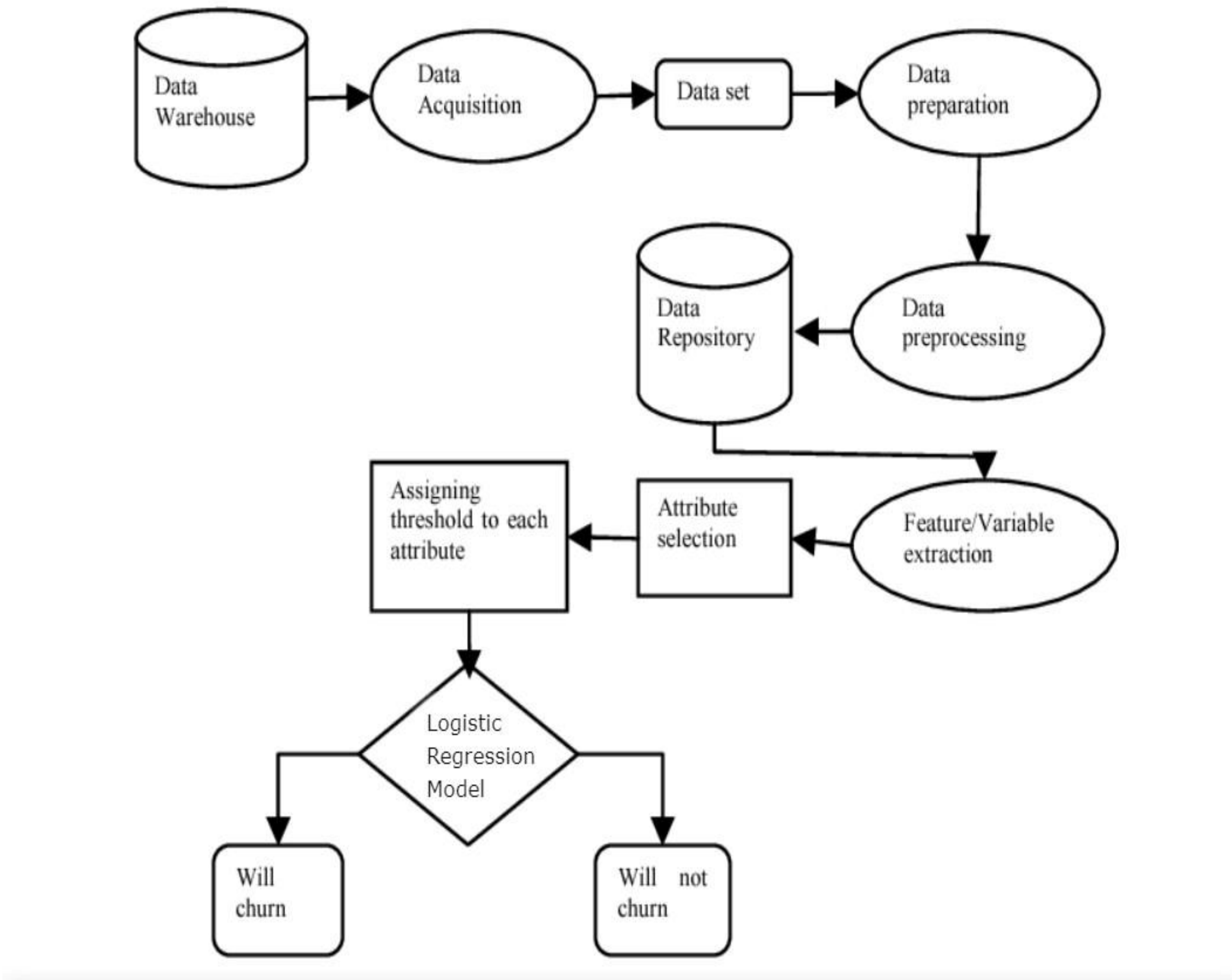
Customer Satisfaction: The model could also analyze customer satisfaction scores, feedback, and reviews. This analysis can help to determine which customers are most likely to churn based on low satisfaction scores.

Promotions and Offers: The model could analyze customer response to promotions and offers. This analysis can help to identify which promotions and offers are most effective in retaining customers and reducing churn.

Customer Loyalty: The model could also analyze customer loyalty programs and rewards. This analysis can help to determine which customers are most loyal and least likely to churn.

Competitor Analysis: The model could also analyze competitor offerings, prices, and promotions. This analysis can help to determine which customers are most likely to switch to a competitor and target retention efforts accordingly.

11. Final Product Prototype with Schematic Diagram :



Steps and components involved in creating a final prototype of a telecommunication churn model- :

Data Collection: The first step is to collect relevant data from various sources such as call detail records, customer demographics, network quality, customer feedback, and competitor analysis. The data should be cleaned and pre-processed to remove any inconsistencies and errors.

Feature Engineering: The next step is to create relevant features that can help to predict churn. This involves identifying important variables, calculating metrics, and selecting the appropriate algorithms to generate the features.

Model Selection: The next step is to select the appropriate machine learning model that can accurately predict churn. Popular models include logistic regression, decision trees, random forests, and neural networks.

Model Training and Validation: The model is trained on the data and validated using various techniques such as cross-validation, AUC, and F1 score. This helps to identify the best performing model and fine-tune its parameters.

Model Deployment: Once the model is trained and validated, it is deployed in a production environment. The model should be integrated with the company's infrastructure, including its data storage, APIs, and web services.

Model Monitoring and Maintenance: The final step is to monitor the model's performance and maintain it over time. This involves identifying any issues, retraining the model with new data, and updating it with the latest algorithms and techniques.

12. Product Details :

- *How does it work?*

The product will take inputs to build an effective telecommunications churn model, it is important to gather data from various sources to gain a comprehensive understanding of the customer's behavior and patterns.

The data set includes information about:

- Customers who left within the last month – the column is called Churn
- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
- Demographic info about customers – gender, age range, and if they have partners and dependents

- Algorithm Used :

Logistic Regression :-

Logistic regression is a statistical method used for binary classification problems, where the goal is to predict the probability of a binary outcome variable based on a set of predictor variables. It is a supervised learning algorithm that is commonly used in machine learning and data analysis.In logistic regression, the outcome variable is binary, meaning it can take one of two values, typically 0 or 1. The predictor variables can be continuous or categorical, and their values are used to calculate the probability of the outcome variable being 1. The logistic regression model uses a mathematical function called the logistic function or sigmoid function to model the relationship between the predictor variables and the outcome variable. The logistic function maps any real-valued input to a value between 0 and 1, which can be interpreted as the probability of the outcome variable being 1.

-Framework Used:

Sklearn :-

scikit-learn, commonly abbreviated as "sklearn", is a popular machine learning library for Python. It provides a wide range of tools for data preprocessing, feature selection, model selection, and performance evaluation. Scikit-learn is built on top of other popular libraries such as NumPy, SciPy, and Matplotlib, which makes it easy to integrate with other data science tools.

Scikit-learn provides a comprehensive set of machine learning algorithms for supervised and unsupervised learning, including regression, classification, clustering, and dimensionality reduction. The library also includes various preprocessing techniques such as scaling, normalization, and feature extraction to prepare the data for modeling.

Steps for proposed methodology :-

Data collection and preparation: The first step is to collect and clean customer data, including demographics, usage patterns, and billing information. This data is then pre-processed and transformed into a format suitable for analysis.

Feature engineering: The next step is to identify and select the features that are most predictive of customer churn. This can include call duration, data usage, customer complaints, billing history, and other relevant data points. These features are then transformed into numerical values that can be used by the machine learning algorithm.

Data splitting: The data is then split into two sets: a training set and a testing set. The training set is used to build the model, while the testing set is used to evaluate its accuracy.

Model selection and training: Next, a machine learning algorithm is selected to build the model. The algorithm is trained on the training set using the selected features. The goal is to find the best combination of features and algorithm that can accurately predict customer churn.

Model evaluation: The trained model is evaluated on the testing set to measure its accuracy and performance. Common evaluation metrics used in churn prediction models include accuracy, precision, recall, and F1 score.

Model deployment: Once the model is deemed accurate and effective, it can be deployed in the production environment. This involves integrating the model into the telecommunications company's existing systems to allow for real-time predictions and interventions.

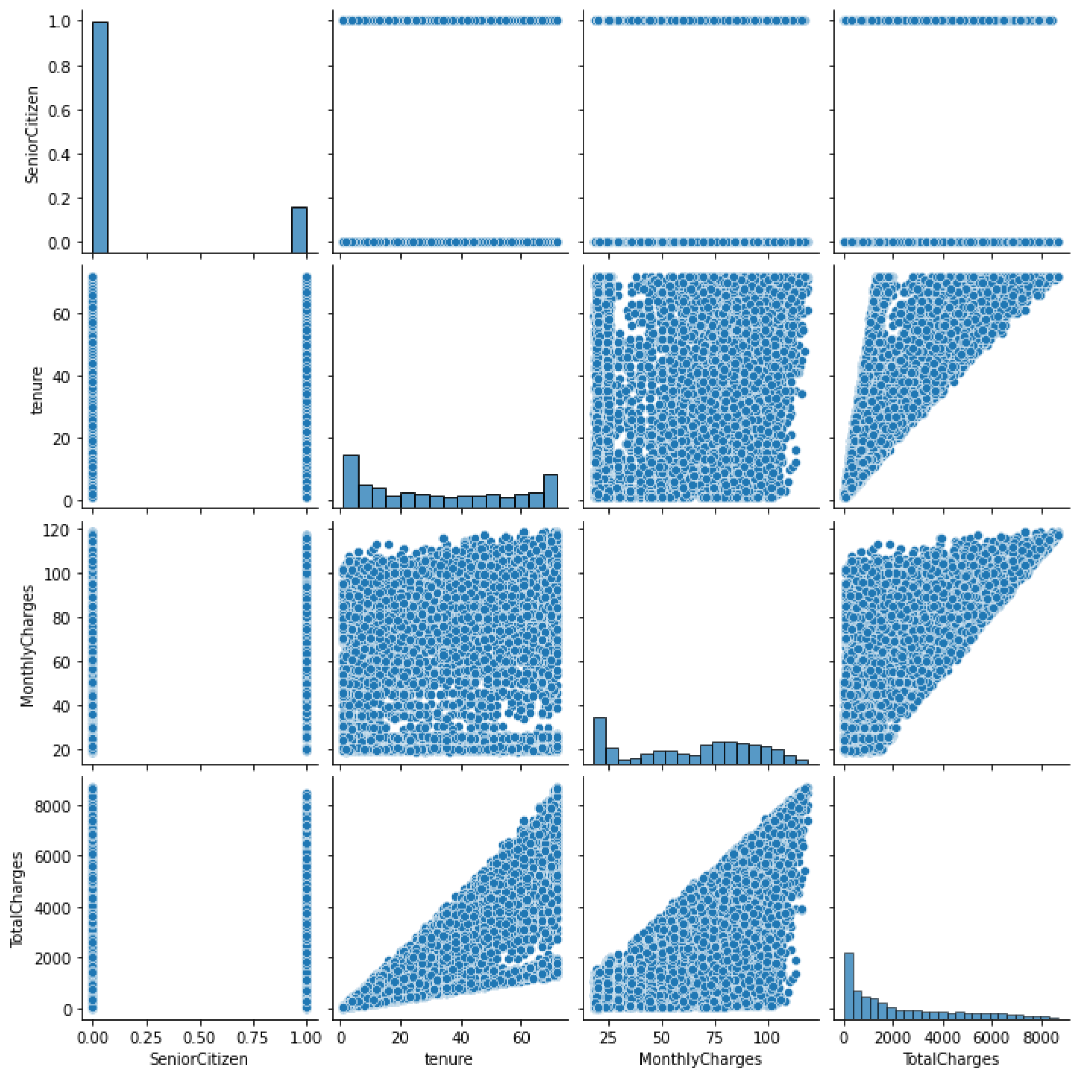
Prediction and intervention: With the model deployed, it can be used to predict which customers are most likely to churn in the future. The telecommunications company can then take proactive measures to retain these customers, such as offering targeted promotions or improving the quality of service. By doing so, the company can reduce churn rates, improve customer satisfaction, and ultimately increase revenue.

13. CODE Implementation :

Exploratory data analysis :

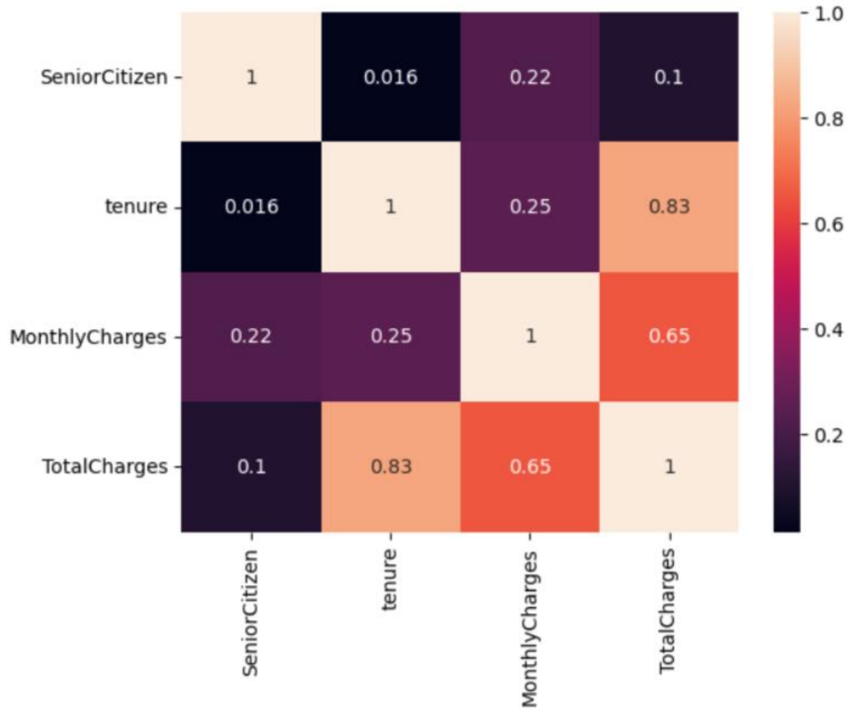
```
sns.pairplot(df)
```

```
<seaborn.axisgrid.PairGrid at 0x7f1f60527cd0>
```

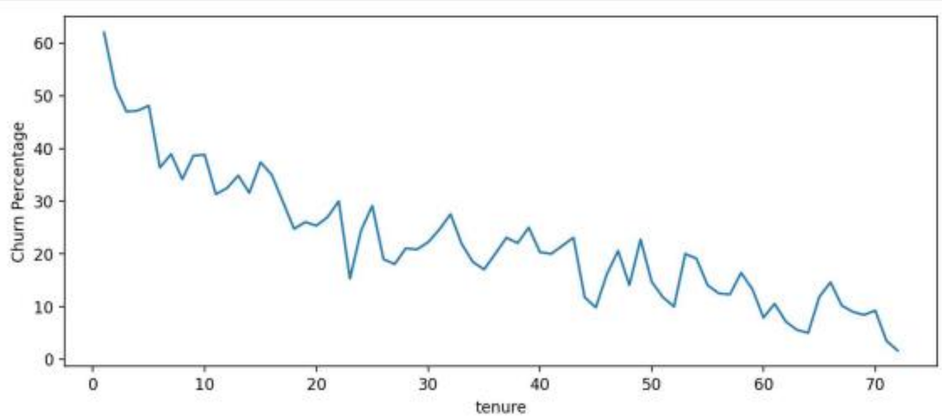


In [11]:
```
sns.heatmap(df.corr(),annot=True)
```

Out[11]: <AxesSubplot:>

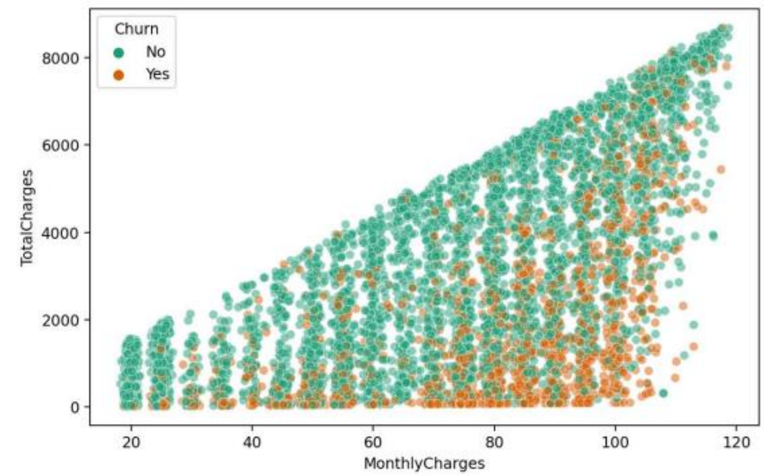


In [476]:
```
plt.figure(figsize=(10,4),dpi=200)
churn_rate.iloc[0].plot()
plt.ylabel('Churn Percentage');
```

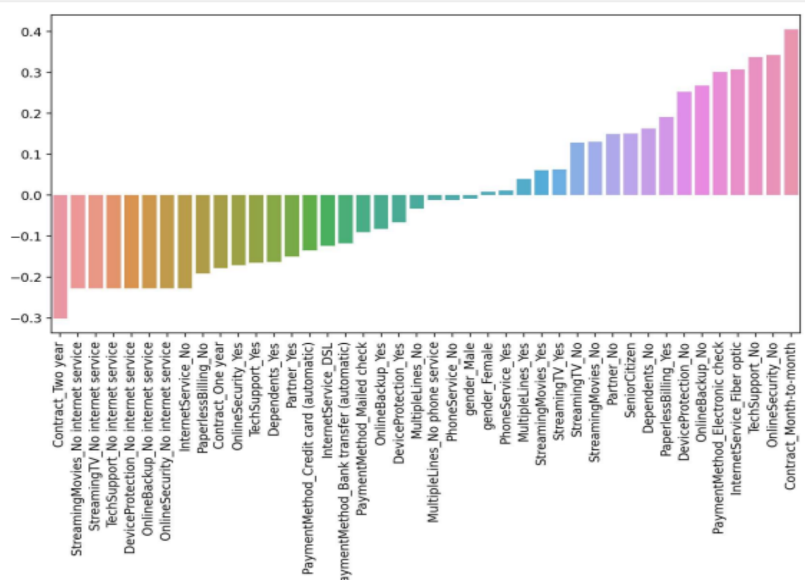


In [471]:
```
plt.figure(figsize=(8,5),dpi=175)
sns.scatterplot(data=df,x='MonthlyCharges',y='TotalCharges',hue='Churn',palette='
```

Out[471]: <AxesSubplot:xlabel='MonthlyCharges', ylabel='TotalCharges'>



In [466]:
```
plt.figure(figsize=(10,5),dpi=200)
sns.barplot(data=df,x= corr_df['Churn_Yes'].sort_values().iloc[1:-1].index,y= cor
plt.xticks(rotation=90);
```



Feature Engineering On the DataSet :

jupyter Telco Customer Churn Last Checkpoint: Yesterday at 4:44 PM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

In [486]:
```
df = df.drop('customerID',axis=1)
```

In [487]:
```
df['tenure_cohort'].unique()
```

Out[487]:
```
array(['0-12 months', 'Over 48 months', '12-24 months', None],
      dtype=object)
```

In [488]:
```
dummies = pd.get_dummies(df[['gender', 'InternetService','Contract','PaymentMethod','tenure_cohort']],drop_first=True)
df = df.drop(['gender', 'InternetService','Contract','PaymentMethod','tenure_cohort'],axis=1)
df = pd.concat([df,dummies],axis=1)
```

In [489]:
```
df.head(5)
```

Out[489]:

	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	Streami
0	0	Yes	No	1	No	No phone service	No	Yes	No	No	No	
1	0	No	No	34	Yes	No	Yes	No	Yes	No	No	
2	0	No	No	2	Yes	No	Yes	Yes	No	No	No	
3	0	No	No	45	No	No phone service	Yes	No	Yes	Yes	No	
4	0	No	No	2	Yes	No	No	No	No	No	No	

In [490]:
```
df['Partner'] = df['Partner'].map({'Yes' : 1 , 'No' : 0})
```

jupyter Telco Customer Churn Last Checkpoint: Yesterday at 4:44 PM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

In [490]:
```
df['Partner'] = df['Partner'].map({'Yes' : 1 , 'No' : 0})
```


In [491]:
```
df.head()
```

Out[491]:

	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	Streami
0	0	1	No	1	No	No phone service	No	Yes	No	No	No	
1	0	0	No	34	Yes	No	Yes	No	Yes	No	No	
2	0	0	No	2	Yes	No	Yes	Yes	No	No	No	
3	0	0	No	45	No	No phone service	Yes	No	Yes	Yes	No	
4	0	0	No	2	Yes	No	No	No	No	No	No	

In [ ]:

In [492]:
```
df['Dependents'] = df['Dependents'].map({'Yes' : 1 , 'No' : 0})
df['PhoneService'] = df['PhoneService'].map({'Yes':1,'No':0})
df['MultipleLines'] = df['MultipleLines'].map({'Yes':1,'No':0,'No phone service':-1})
df['OnlineSecurity'] = df['OnlineSecurity'].map({'Yes':1,'No':0,'No internet service':-1})
df['OnlineBackup'] = df['OnlineBackup'].map({'Yes':1,'No':0,'No internet service':-1})
df['DeviceProtection'] = df['DeviceProtection'].map({'Yes':1,'No':0,'No internet service':-1})
df['TechSupport'] = df['TechSupport'].map({'Yes':1,'No':0,'No internet service':-1})
df['StreamingTV'] = df['StreamingTV'].map({'Yes':1,'No':0,'No internet service':-1})
df['StreamingMovies'] = df['StreamingMovies'].map({'Yes':1,'No':0,'No internet service':-1})
df['PaperlessBilling'] = df['PaperlessBilling'].map({'Yes':1,'No':0})
df['Churn'] = df['Churn'].map({'Yes':1,'No':0})
```

Splitting the dataset in x,y variable :

jupyter Telco Customer Churn Last Checkpoint: Yesterday at 4:44 PM (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

```
In [494]: X = df.drop('Churn',axis=1)

In [495]: y = df['Churn']

In [496]: X
```

3	0	0	0	45	0	-1	1	0	1	1	0
4	0	0	0	2	1	0	0	0	0	0	0
5	0	0	0	8	1	1	0	0	1	0	1
6	0	0	1	22	1	1	0	1	0	0	1
7	0	0	0	10	0	-1	1	0	0	0	0
8	0	1	0	28	1	1	0	0	1	1	1
9	0	0	1	62	1	0	1	1	0	0	0
10	0	1	1	13	1	0	1	0	0	0	0
11	0	0	0	16	1	0	-1	-1	-1	-1	-1
12	0	1	0	58	1	1	0	0	1	0	1
13	0	0	0	49	1	1	0	1	1	0	1
14	0	0	0	25	1	0	1	0	1	1	1

```
In [497]: y

Out[497]: 0       0
          1       0
          2       1
          3       0
          4       1
          5       1
          6       0
```

Preparing the Model :

Here, we are splitting the dataset in 8:2 ratio (train to test) and then using standard scaler to scale the dataset.
We will create a Logistic Regression model in order to achieve the highest accuracy and will use GridSearchCV to find the best parameter for the model.

jupyter Telco Customer Churn Last Checkpoint: Yesterday at 4:44 PM (autosaved)

File Edit View Insert Cell Kernel Widgets Help   Not Trusted   Python 3 (ipykernel)

```
In [498]: from sklearn.model_selection import train_test_split

In [499]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=101)

In [500]: from sklearn.preprocessing import StandardScaler

In [501]: scaler = StandardScaler()

In [502]: scaled_X_train = scaler.fit_transform(X_train)

In [503]: scaled_X_test = scaler.transform(X_test)

In [519]: from sklearn.metrics import classification_report,confusion_matrix,accuracy_score

In [505]: # Creating Models

In [506]: from sklearn.linear_model import LogisticRegression

In [507]: from sklearn.model_selection import GridSearchCV

In [508]: from sklearn.pipeline import Pipeline

In [509]: log_model = LogisticRegression(max_iter=5000)

In [510]: penalty = ['l1','l2','elasticnet']
          C = np.logspace(0,4,10)

In [511]: grid_model = GridSearchCV(log_model,param_grid={'C':C,'penalty':penalty})

In [512]: grid_model.fit(scaled_X_train,y_train)
```

I have taken `l1`,`l2` & `elasticnet` as the penalty term and has taken 1.00000000e+00, 2.78255940e+00, 7.74263683e+00, 2.15443469e+01, 5.99484250e+01, 1.66810054e+02, 4.64158883e+02, 1.29154967e+03, 3.59381366e+03, 1.00000000e+04 for the value of C.

Finalising the model :

```
jupyter  Telco Customer Churn Last Checkpoint: Yesterday at 4:44 PM (autosaved)                Logout
File  Edit  View  Insert  Cell  Kernel  Widgets  Help                  Not Trusted | Python 3 (ipykernel)

Out[512]: GridSearchCV(estimator=LogisticRegression(max_iter=5000),
                       param_grid={'C': array([1.00000000e+00, 2.78255940e+00, 7.74263683e+00, 2.15443469e+01,
                 5.99484250e+01, 1.66810054e+02, 4.64158883e+02, 1.29154967e+03,
                 3.59381366e+03, 1.00000000e+04]),
                                   'penalty': ['l1', 'l2', 'elasticnet']})

In [513]: grid_model.best_params_

Out[513]: {'C': 7.742636826811269, 'penalty': 'l2'}

In [514]: predictions = grid_model.predict(scaled_X_test)

In [515]: print(classification_report(y_test,predictions))

              precision    recall  f1-score   support

           0       0.85      0.91      0.88      1052
           1       0.66      0.53      0.59       355

    accuracy                           0.81      1407
   macro avg       0.76      0.72      0.73      1407
weighted avg       0.80      0.81      0.81      1407

In [518]: confusion_matrix(y_test,predictions)

Out[518]: array([[955,  97],
                 [166, 189]], dtype=int64)

In [521]: accuracy_score(y_test,predictions)*100

Out[521]: 81.30774697938877
```

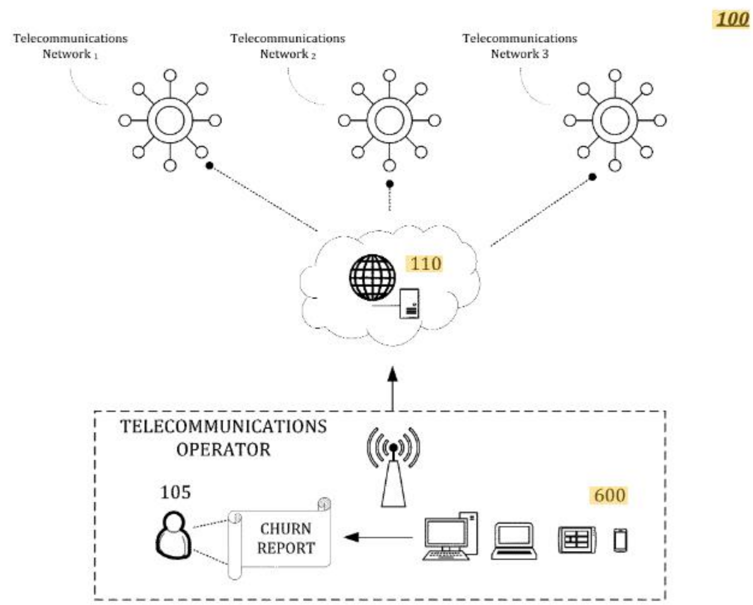
After training the model , we obtained the most suitable value for penalty as `l2` and 7.74 for `C` & after predicting , we obtained a good accuracy of 81.3077.

github *:- https://github.com/harsh-priyam/Telecommunication-Churn-Prediction*

14. Applicable Patents :

Embodiments of the present disclosure may provide a platform configured to forecast customer churn in a telecommunication network. The platform may be configured to receive customer activity data. The platform may then compute features associated with the customer activity data. These features are then inputted into a machine learning model used for predicting customer churn.

Finally, the platform may then provide a report indicating customer churn predictions. The platform may be trained in a training phase prior to entering a prediction phase.



The platform may employ an ensemble of statistical machine learning classifiers. An ensemble of classifiers may comprise a set of classifiers whose individual decisions are combined to generate a final decision. An ensemble consistent with embodiments of the present disclosure may be composed by several supervised classification algorithms, including, but not limited to: random forest, neural networks, support vector machines, and logistic regression.

Complete patent link -: https://patents.google.com/patent/US20150310336A1/en

15. Conclusion :

The telecommunications industry is highly competitive, with customers frequently switching providers for a variety of reasons. Therefore, it is crucial for telecommunications companies to have effective churn prediction models in place to retain their customers.

In conclusion, a telecommunications churn model is a predictive model that uses historical customer data to identify those customers who are most likely to leave the service provider. The model can be built using various machine learning algorithms such as logistic regression, decision trees, and random forests.

By accurately predicting which customers are likely to churn, telecommunications companies can take proactive measures to retain them, such as offering targeted promotions or improving the quality of service. This can ultimately lead to higher customer satisfaction, reduced churn rates, and increased revenue for the service provider.

However, it is important to note that a churn model is only as good as the data it is trained on. Therefore, it is essential to continually monitor and update the model with the latest customer data to ensure its accuracy and effectiveness.

References/Source of Information :

- https://www.kaggle.com/datasets/blastchar/telco-customer-churn
- https://chat.openai.com/chat
- https://en.wikipedia.org/wiki/Telecommunications_industry
- https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0191-6
- https://www.semanticscholar.org/paper/Customer-churn-analysis-in-telecom-industry-Dahiya-Bhatia/8417c9f074f2b9d06eed4e210a33c730bb28615e