Name Nikhil Singh

Reg No :12109514

Subject : SPEAKX Assignment (Predicting Customer Churn in a Telecommunications Company)

Predicting Customer Churn in a Telecommunication Company

We are having a dataset named WA\_Fn-UseC\_-Telco-Customer-Churn.csv (<https://www.kaggle.com/datasets/blastchar/telco-customer-churn/data>) .

* Firstly ,we will Upload the dataset in or jupyter Notebook

dataset has rows = 7043 and colums = 21

In which 3 are numerical and 18 are categorical.

Data.info()

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 7043 entries, 0 to 7042

Data columns (total 21 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 customerID 7043 non-null object

1 gender 7043 non-null object

2 SeniorCitizen 7043 non-null int64

3 Partner 7043 non-null object

4 Dependents 7043 non-null object

5 tenure 7043 non-null int64

6 PhoneService 7043 non-null object

7 MultipleLines 7043 non-null object

8 InternetService 7043 non-null object

9 OnlineSecurity 7043 non-null object

10 OnlineBackup 7043 non-null object

11 DeviceProtection 7043 non-null object

12 TechSupport 7043 non-null object

13 StreamingTV 7043 non-null object

14 StreamingMovies 7043 non-null object

15 Contract 7043 non-null object

16 PaperlessBilling 7043 non-null object

17 PaymentMethod 7043 non-null object

18 MonthlyCharges 7043 non-null float64

19 TotalCharges 7043 non-null object

20 Churn 7043 non-null object

dtypes: float64(1), int64(2), object(18)

memory usage: 1.1+ MB

dataFrame.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 |

There is no null values and no outliers in the three numeric data.

Univariate Analysis:

Discovers various trends in data that how many customers are having partner , about their gender etc.

Bivariate Analysis and Multivariate analysis and so on.

Then trained various machine learning models for predicting churn .

In which SVM serves = 81.9% accuracy

Logistic regression serves = 80% accuracy

Are highest .