

customer churn prediction

phase3 project submission

name : S.apsara jasmine

naan mudalvan id : au920821104006

phase 3 project : developement part 1

ABSTRACT:

- Customer churn, the rate at which customers discontinue their association with a company, poses a significant challenge for businesses across industries. In an era marked by data abundance, this study leverages the power of data analytics to predict and mitigate customer churn. This research employs a comprehensive dataset of

customer interactions, including demographics, transaction history, and customer feedback, and applies various machine learning and statistical techniques to develop predictive models. The aim is to identify the key factors that influence customer attrition and provide businesses with actionable insights to proactively retain their customer base. The results show promising predictive accuracy, offering companies an opportunity to optimize their customer retention strategies and enhance customer satisfaction. This research contributes to the growing field of customer relationship management by showcasing the potential of data analytics in predicting

and preventing customer churn, ultimately fostering sustainable business growth.

DATA SOURCE:

- Churn prediction relies on data from various sources, including senior citizen, gender, techsupport, phoneservice, multiple lines, internet service and customer feedback

DATA LINK:

<https://www.kaggle.com/datasets/blastchar/telco-customer-churn>

DATASET:

```
WA_Fn-UseC_-Telco-Customer-Churn.csv - Notepad
File Edit View

customerID,gender,SeniorCitizen,Partner,Dependents,tenure,PhoneService,MultipleLines,InternetService,OnlineSecurity,OnlineBackup,DeviceProtection,TechSupport
7590-VHVEG,Female,0,Yes,No,1,No,No,phone service,DSL,No,Yes,No,No,No,No,Month-to-month,Yes,Electronic check,29.85,29.85,No
5575-GNVDE,Male,0,No,No,34,Yes,No,DSL,Yes,No,Yes,No,No,No,One year,No,Mailed check,56.95,1889.5,No
3668-QPYBK,Male,0,No,2,Yes,No,DSL,Yes,Yes,No,No,No,No,Month-to-month,Yes,Mailed check,53.85,108.15,Yes
7795-CFOCW,Male,0,No,45,No,No,phone service,DSL,Yes,No,Yes,Yes,No,No,One year,No,Bank transfer (automatic),42.3,1840.75,No
9237-HQITU,Female,0,No,2,Yes,No,Fiber optic,No,No,No,No,No,No,Month-to-month,Yes,Electronic check,70.7,151.65,Yes
9385-CDSKC,Female,0,No,8,Yes,Yes,Fiber optic,No,No,Yes,No,Yes,Yes,Month-to-month,Yes,Electronic check,99.65,820.5,Yes
1452-KIOVK,Male,0,No,22,Yes,Yes,Fiber optic,No,Yes,No,No,Yes,No,Month-to-month,Yes,Credit card (automatic),89.1,1949.4,No
6713-OKOMC,Female,0,No,10,No,No,phone service,DSL,Yes,No,No,No,No,No,Month-to-month,No,Mailed check,29.75,301.9,No
7892-PDOKP,Female,0,Yes,No,28,Yes,Yes,Fiber optic,No,No,Yes,Yes,Yes,Yes,Month-to-month,Yes,Electronic check,104.8,3046.05,Yes
6388-TABGU,Male,0,No,62,Yes,No,DSL,Yes,Yes,No,No,No,One year,No,Bank transfer (automatic),56.15,3487.95,No
9763-GRSKO,Male,0,Yes,13,Yes,No,DSL,Yes,No,No,No,No,Month-to-month,Yes,Mailed check,49.95,587.45,No
7469-LKBCT,Male,0,No,16,Yes,No,No,No,Internet service,No,Internet service,No,Internet service,No,Internet service,No,Internet service
8091-TTVAX,Male,0,Yes,No,58,Yes,Yes,Fiber optic,No,No,Yes,No,Yes,Yes,One year,No,Credit card (automatic),100.35,5681.1,No
0280-XJGEX,Male,0,No,49,Yes,Yes,Fiber optic,No,Yes,Yes,No,Yes,Yes,Month-to-month,Yes,Bank transfer (automatic),103.7,5036.3,Yes
5129-JLPII,Male,0,No,25,Yes,No,Fiber optic,Yes,No,Yes,Yes,Yes,Yes,Month-to-month,Yes,Electronic check,105.5,2686.05,No
3655-SNOVZ,Female,0,Yes,69,Yes,Yes,Fiber optic,Yes,Yes,Yes,Yes,Yes,Yes,Two year,No,Credit card (automatic),113.25,7895.15,No
8191-XMSZG,Female,0,No,52,Yes,No,No,No,Internet service,No,Internet service,No,Internet service,No,Internet service,No,Internet service
9959-WOFKT,Male,0,No,71,Yes,Yes,Fiber optic,Yes,No,Yes,No,Yes,Yes,Two year,No,Bank transfer (automatic),106.7,7382.25,No
4190-MFLUN,Female,0,Yes,10,Yes,No,DSL,No,No,Yes,Yes,No,No,Month-to-month,No,Credit card (automatic),55.2,528.35,Yes
4183-MYFRB,Female,0,No,21,Yes,No,Fiber optic,No,Yes,Yes,No,No,Yes,Month-to-month,Yes,Electronic check,90.05,1862.9,No
8779-QRDMW,Male,1,No,1,No,No,phone service,DSL,No,No,Yes,No,No,Yes,Month-to-month,Yes,Electronic check,39.65,39.65,Yes
1680-VDCMW,Male,0,Yes,12,Yes,No,No,No,Internet service,No,Internet service,No,Internet service,No,Internet service,No,Internet service
1066-JKSGK,Male,0,No,1,Yes,No,No,No,Internet service,No,Internet service,No,Internet service,No,Internet service,No,Internet service
3638-WEABW,Female,0,Yes,58,Yes,Yes,DSL,No,Yes,No,Yes,No,Two year,Yes,Credit card (automatic),59.9,3505.1,No
6322-HRPFA,Male,0,Yes,49,Yes,No,DSL,Yes,Yes,No,Yes,No,No,Month-to-month,No,Credit card (automatic),59.6,2970.3,No
6865-JZIKO,Female,0,No,30,Yes,No,DSL,Yes,Yes,No,No,No,No,Month-to-month,Yes,Bank transfer (automatic),55.3,1536.6,No
6467-CHFWZ,Male,0,Yes,47,Yes,Yes,Fiber optic,No,Yes,No,No,Yes,Yes,Month-to-month,Yes,Electronic check,99.35,4749.15,Yes
8665-UTDZT,Male,0,Yes,1,No,No,phone service,DSL,No,Yes,No,No,No,No,Month-to-month,No,Electronic check,30.2,30.2,Yes
5248-VGIJN,Male,0,Yes,72,Yes,Yes,DSL,Yes,Yes,Yes,Yes,Two year,Yes,Credit card (automatic),90.25,6369.45,No
8773-HHUOZ,Female,0,No,17,Yes,No,DSL,No,No,No,No,Yes,Month-to-month,Yes,Mailed check,64.7,1093.1,Yes
3841-NFECK,Female,1,Yes,No,71,Yes,Yes,Fiber optic,Yes,Yes,Yes,Yes,No,Two year,Yes,Credit card (automatic),96.35,6766.95,No
```

PROGRAM:

Spyder (Python 3.11)

File Edit Search Source Run Debug Consoles Projects Tools View Help

C:\churn\phase3.py

```

1 import pandas as pd
2 import numpy as np
3 import seaborn as sns
4 import matplotlib.ticker as mtick
5 import matplotlib.pyplot as plt
6 data="C:/churn/Telco-Customer-Churn.csv"
7 df=pd.read_csv(data)
8 print(df)
9
10
11 print(df.head())
12 print(df.info())
13
14 |
15 #step3:data preprocessing
16 df.columns.values
17 df.dtypes
18 df.isnull().sum()
19
20 # removing missings values
21
22 df.dropna(inplace=True)
23
24 #removing customer IDs from the dataset
25 df2=df.iloc[:,1:]
26
27 #converting the predictor variable to a binary numeric variable
28 df2.replace(to_replace='yes',value=1,inplace=True)
29 df2.replace(to_replace='no',value=0,inplace=True)
30
31 #let's convert all the categorical variables into dummy variables
32

```

phase3.py 24/10/2023 11:08 pm
Telco-Customer-Churn.csv 24/10/2023 12:37 am

Console 3/A X

```

...: df=pd.read_csv(data)
...: print(df)

```

	customerID	gender	SeniorCitizen	...	MonthlyCharges	TotalCharges	Churn
0	7590-VHVEG	Female	0	...	29.85	29.85	No
1	5575-GNVDE	Male	0	...	56.95	1889.5	No
2	3668-QPYBK	Male	0	...	53.85	108.15	Yes
3	7795-CFOGW	Male	0	...	42.30	1840.75	No
4	9237-HQITU	Female	0	...	70.70	151.65	Yes
...
7038	6840-RESVB	Male	0	...	84.80	1990.5	No
7039	2234-XADUH	Female	0	...	103.20	7362.9	No
7040	4801-JZAZL	Female	0	...	29.60	346.45	No
7041	8361-LTHKD	Male	1	...	74.40	306.6	Yes
7042	3186-AJIEK	Male	0	...	105.65	6844.5	No

conda (Python 3.11.5) LSP: Python Line 15, Col 1 ASCII CRLF RW Mem 44%

27°C Haze

Spyder (Python 3.11)

File Edit Search Source Run Debug Consoles Projects Tools View Help

C:\churn\phase3.py

```

17 df.dtypes
18 df.isnull().sum()
19
20 # removing missings values
21
22 df.dropna(inplace=True)
23
24 #removing customer IDs from the dataset
25 df2=df.iloc[:,1:]
26
27 #converting the predictor variable to a binary numeric variable
28 df2.replace(to_replace='yes',value=1,inplace=True)
29 df2.replace(to_replace='no',value=0,inplace=True)
30
31 #let's convert all the categorical variables into dummy variables
32
33 df_dummies=pd.get_dummies(df2)
34 df_dummies.head()
35
36 #get correlation of churn with other variables
37 plt.figure(figsize=(15,8))
38 df_dummies.corr().sort_values(ascending=False).plot(kind='bar')
39
40 # data exploration
41 colors=['#4D3425', '#E4512B']
42 ax=(df['gender'].value_counts()*100.0/len(df)).plot(kind='bar',stacked=True,
43 ax.yaxis.set_major_formatter(mtick.percentFormatter())
44 ax.set_xlabel('%customers')
45 ax.set_ylabel('gender')
46 ax.set_xlabel('%customer')
47 ax.set_title('Gender Distribution')
48

```

phase3.py 24/10/2023 11:08 pm
Telco-Customer-Churn.csv 24/10/2023 12:37 am

Console 3/A X

```

date columns (columns) Non-Null Count Dtype
# Column
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

```

conda (Python 3.11.5) LSP: Python Line 48, Col 1 ASCII CRLF RW Mem 44%

27°C Haze