

PROJECT TITLE:

Customer Churn Analysis and Prediction

DESCRIPTION:

A data-driven machine learning project that identifies key factors behind customer churn in a telecom company. The project aims to analyse patterns and predict whether a customer is likely to leave the company, helping the business make informed decisions to reduce churn.

FEATURES:

1. Data cleaning, handling missing values, and encoding.
2. Exploratory Data Analysis (EDA) to find insights using visualisations.
3. Feature selection and transformation.
4. Machine learning model building and evaluation.
5. Churn prediction using the best-performing model.

TECHNOLOGIES USED:

1. Python
2. Pandas, NumPy

TASK1:DATA CLEANING AND PREPROCESSING

DESCRIPTION:

Loaded the Telco Customer Churn dataset, handled missing values in TotalCharges, converted it to numeric ,and encoded categorical columns.

Code Snippet:

```
import warnings
warnings.simplefilter(action='ignore',category=FutureWarning)
```

```
import pandas as pd
import numpy as np
```

```
df = pd.read_csv("customer_churn_sample.csv")
print("Initial Data Preview:")
print(df.head())
```

```
#convert total charges into numeric
df['TotalCharges'] = pd.to_numeric(df['TotalCharges'],
errors='coerce')
```

```
#Checking for missing values
print("\nMissing values before cleaning:")
print(df.isnull().sum())
df['TotalCharges'].fillna(df['TotalCharges'].median(),
inplace=True)#filling total charges with median
df['OnlineSecurity'].fillna(df['OnlineSecurity'].mode()[0],
inplace=True)
```

```

df['DeviceProtection'].fillna(df['DeviceProtection'].mode()[0],
inplace=True)

#Recheck for missing values
print('\nMissing values after cleaning:')
print(df.isnull().sum())

#encode categorical variables using onehot encoding
df_encoded = pd.get_dummies(df, drop_first=True)
print("\nEncoded data sample:")
print(df_encoded.head())
df_encoded.to_csv("cleaned_customer_churn.csv", index=False)
print("\ncleaned data saved as 'cleaned_customer_churn.csv'")

```

OUTPUT:

```

/usr/local/bin/python3 /Users/vishnuvardhana/Documents/Intern/Task1.py
vishnuvardhana@Vishnuvardhana-MacBook-Air Intern % /usr/local/bin/python3 /Users/vishnuvardhana/Documents/Intern/Task1.py
Initial Data Preview:
customerID  gender  SeniorCitizen  Partner  Dependents  ...  PaperlessBilling  PaymentMethod  MonthlyCharges  TotalCharges  Churn
0           1   Female             0       Yes         No  ...           Yes  Electronic check         29.85         29.85      No
1           2     Male             1       No         No  ...           No    Mailed check         56.95        1889.50     No
2           3   Female             0       No         Yes  ...           Yes    Bank transfer         53.85         108.15     Yes
3           4     Male             0       Yes         No  ...           No    Credit card         42.30         1840.75     No
4           5   Female             1       No         Yes  ...           Yes  Electronic check         70.70         70.70      Yes

[5 rows x 21 columns]

Missing values before cleaning:
customerID      0
gender           0
SeniorCitizen   0
Partner         0
Dependents      0
tenure          0
PhoneService    0
MultipleLines   0
InternetService 0
OnlineSecurity  1
OnlineBackup    0
DeviceProtection 1
TechSupport     0
StreamingTV     0
StreamingMovies 0
Contract        0
PaperlessBilling 0
PaymentMethod   0
MonthlyCharges  0
TotalCharges    1
Churn           0
dtype: int64

Missing values after cleaning:
customerID      0
gender           0
SeniorCitizen   0
Partner         0
Dependents      0
tenure          0
PhoneService    0
MultipleLines   0
InternetService 0
OnlineSecurity   0
OnlineBackup    0
DeviceProtection 0
TechSupport     0
StreamingTV     0
StreamingMovies 0
Contract        0
PaperlessBilling 0
PaymentMethod   0
MonthlyCharges  0
TotalCharges    0
Churn           0
dtype: int64

```

📁

🔍

🔗

📦

🔧

👤

⚙️

EXPLORER

...

OPEN EDITORS

Task1.py

INTERN

cleaned_customer_churn.csv

customer_churn_sample.csv

Task1.py

Task2.py

Task3.py

Task4.py

Task5.py

Task6.py

PROBLEMS

OUTPUT

DEBUG CONSOLE

TERMINAL

PORTS

customerID 0

gender 0

SeniorCitizen 0

Partner 0

Dependents 0

tenure 0

PhoneService 0

MultipleLines 0

InternetService 0

OnlineSecurity 1

OnlineBackup 0

DeviceProtection 1

TechSupport 0

StreamingTV 0

StreamingMovies 0

Contract 0

PaperlessBilling 0

PaymentMethod 0

MonthlyCharges 0

TotalCharges 1

Churn 0

dtype: int64

Missing values after cleaning:

customerID 0

gender 0

SeniorCitizen 0

Partner 0

Dependents 0

tenure 0

PhoneService 0

MultipleLines 0

InternetService 0

OnlineSecurity 0

OnlineBackup 0

DeviceProtection 0

TechSupport 0

StreamingTV 0

StreamingMovies 0

Contract 0

PaperlessBilling 0

PaymentMethod 0

MonthlyCharges 0

TotalCharges 0

Churn 0

dtype: int64

Encoded data sample:

customerID SeniorCitizen tenure ... PaymentMethod_Electronic check PaymentMethod_Mailed check Churn_Yes

0 1 0 1 ... True False False

1 2 1 34 ... False True False

2 3 0 2 ... False True True

3 4 0 45 ... False False False

4 5 1 5 ... True False True

[5 rows x 25 columns]

cleaned data saved as 'cleaned_customer_churn.csv'

vishnuvardhana@Vishnuvardhans-MacBook-Air Intern %

zsh

Python

Ln 31, Col 62

Spaces: 4

UTF-8

LF

{ } Python

3.13.3 64-bit