COSTUMER CHURN PREDICTION

PHASE 3:DEVELOPMENT PART 1



INTRODUCTION:

**$** Customer churn prediction is an ongoing process that requires constant monitoring and adaptation to changing customer behavior and market conditions. Regularly re-evaluate and update your churn prediction model to maintain its effectiveness in retaining customers.

Top of Form

**$** Customer churn prediction is a crucial task for businesses to identify and mitigate customer attrition.

**$** To perform customer churn prediction, follow these steps:

**$** We need to gather relevant data about the customers and their interactions with our business. This data can include customer demographics, transaction history, customer service interactions, website/app usage, and any other data that may be indicative of churn.

**$** Next, we need to clean and preprocess the data. This involves handling missing values, encoding categorical variables, and scaling numerical features.

**$** And then we have to Identify and create meaningful features that can help predict customer churn. Examples of relevant features include customer lifetime value, frequency of purchases, customer feedback scores, and recency of interactions.

**$** Divide the dataset into training, validation, and test sets. The training set is used to train the model and the validation set to tune hyperparameters.

**$** Next, we have to choose an appropriate machine learning algorithm for the churn prediction task. We have included logistic regression, decision trees, random forests. The selection depends on the nature of our data and our specific business needs.

**$** The next upcoming steps is to train the machine learning model using the training dataset.

**$** Assess the model's performance on the test set using relevant evaluation metrics, such as accuracy, precision, recall, F1 score, or area under the ROC curve (AUC).

*1.LOADING THE DATASET:*

We are importing the necessary Libraries and then we are loading the dataset.

Program:

import os

import pandas as pd

from pandas import DataFrame

pd.set\_option("display.max\_columns", 50)

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from scipy.stats.contingency import chi2\_contingency

from scipy.stats import normaltest, shapiro

import plotly.express as px

import warnings

warnings.filterwarnings ("ignore")

import matplotlib

matplotlib.rcParams["axes.labelsize"] = 9

matplotlib.rcParams["legend.fontsize"] = 9

matplotlib.rcParams["ytick.labelsize"] = 9

matplotlib.rcParams["xtick.labelsize"] = 9

Data= pd.read\_csv(“C:\Users\THANGAPANDI\Documents\costumer churn predicton.csv”)

pd.head(4)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  | |  | |  | |  | | |  | |  | |  | |  | |  | | OnlineBackup | | | DeviceProtection | | | TechSupport | | StreamingTV | | StreamingMovies | | Contract | | PaperlessBilling | | | PaymentMethod | | | MonthlyCharges | | | TotalCharges | | | Churn | | |
| customerID | gender | | SeniorCitizen | | Partner | | Dependents | | tenure | PhoneService | | MultipleLines | | InternetService | | OnlineSecurity | | OnlineBackup | | DeviceProtection | | TechSupport | StreamingTV | | StreamingMovies | Contract | | PaperlessBilling | | PaymentMethod | | MonthlyCharges | | TotalCharges | | Churn |  | |  |  | |  |  | |  |  | |  |  | |  |  |  |
| 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 | | 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 | | No | | 45 | No | | No phone service | | DSL | | Yes | | No | | Yes | | Yes | No | | No | One year | | No | | Bank transfer (automatic) | | 42.3 | | 1840.75 | | No |  | |  |  | |  |  | |  |  | |  |  | |  |  |  |

*2. DISPLAYING THE NUMBER OF ROWS AND COLUMNS:*

n\_rows, n\_columns = df.shape

print(f"Number of columns: **{**n\_columns**}** columns**\n**Number of rws: **{**n\_rows**}** rows")

Number of columns: 21 columns

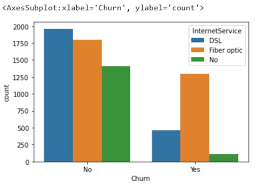
Number of rws: 7044 rows

*3.VISUAL REPRESENTATION:*

Now, let’s look into the relationship between cost and customer churn. In the real world, users tend to unsubscribe to their mobile service provider and switch to a different brand if they find the monthly subscription cost too high. Let’s check if that behavior is reflected in our dataset:

sns.boxplot(x='Churn', y='MonthlyCharges', data=df)

<Axes: xlabel='Churn', ylabel='MonthlyCharges'>



*4.RANDOM FOREST CLASSIFIER:*

clf = RandomForestClassifier(n\_estimators=100,random\_state=42)

clf.fit(X\_train,y\_train)

y\_pred= clf.predict(X\_test)

accuracy = accuracy\_score(y\_test,y\_pred)

conf\_matrix = confusion\_matrix(y\_  
test,y\_pred)

report = classification\_report(y\_test,y\_pred)

print(“Accuracy:”,accuracy)

print(“confusion matrix:”)

print(conf\_matrix)

print(“Classification Report:”)

print(report)

