**READ ME DESCRIPTION FILE FOR –**

***CUSTOMER-CHURN-PREDICTION-MODEL***

* **What is Customer Churn?**

Customer churn is the percentage of customers that stopped using your company's product or service during a certain time frame.

* **Why is Customer Churn Important?**

A natural process of businesses is that customers will come and go but not all customers are here to stay. Managing and reducing churn rate is important because it costs much more to acquire new customers than it does to retain existing customers.

* **OVERVIEW**

To effectively retain customers, telecommunications businesses need to be able to analyze customer churn. However, retaining existing customers is more costly than obtaining fresh customers. Thus we try to predict the retention rate of a customer using a customer churn prediction project.

* **What are We Building?**

In the Customer Churn Prediction Project, we are building a predictive model to determine approximately when the customers are about to leave the organization.

* **Pre-requisites**

Below are some pre-requisites -

->Knowledge of Python

->Familiarity with Utility Libraries like Pandas, Numpy

->Basic understanding of visualization libraries like Seaborn, Matplotlib

->Implementable knowledge of SciKit Learn

->Learning concepts like - Data Transformation, Feature Engineering, Model Training, Model Evaluation, and Model Testing

* **How Are We Going to Build This**?

1.Environment - Selecting an IDE or an environment to operate on, either online or offline

2.Importing - Loading the datasets and libraries needed

3.Analysing - Comprehending the structure of the data like the columns, values, and data types

4.Visualisation - Visualising the dataset to understand underlying designs

5.Filling out values - Cleaning the data by completing the dataset and filling out the missing values

6.Feature Engineering - Understanding the features and developing new ones based on essentials

7.Traning the model - Employing algorithms like logistic regression and random forest to train the Model

8.Evaluation - Assessing the predictive model

9.Testing - Testing the Model and computing the accuracy

* **Final Output**

The resultant output will be a prediction based on SVM, Ridge Classification, Random Forest, and XG Boost Algorithm.

The accuracy will be approximate as follows -

SVM - 76% Ridge Classification - 74% Random Forest - 90% XG Boost Algorithm - 80%

* **Requirements**

IDE - A compiler to build the project on,jupyter notebook etc

Libraries - Pandas, Numpy, Matplotlib, Seaborn, SciKit Learn

Algorithms - SVM, Ridge Classification, Random Forest, and XG Boost Algorithm

* **About the Data**

$ CustomerID - Distinctive ID is given to every consumer

$ gender - If the customer is a male or a female

$ SeniorCitizen - If the customer is a senior citizen

$ Partner - If the customer has a spouse

$ Dependents - If the customer has dependents

$ Tenure - Number of months the customer has remained with the business

$ Phone Service - If the customer has a phone service

$ Multiple Lines - If the customer has numerous lines

$ Internet Service - Consumer’s internet service provider

$ Online Security - If the consumer has online security

$ Online Backup - If the customer has an online tie-up

$ Device Protection - If the customer has gadget security

$ Tech Support - If the customer has tech support

$ Streaming TV - If the customer has streaming TV

$ Streaming Movies - If the customer has streaming movies

$ Contract - Contract duration of the customer

$ Paperless Billing - If the customer has paperless billing

$ Payment Method - Customer’s expenditure method

$ Monthly Charges - Amount levied to the customer per month

$ Total Charges - Total amount levied to the customer

$ Churn - If the customer churned

* **Understanding the Problem Statement**

This project aims to use the available information to build a machine learning model that will accurately predict which consumers will likely quit a company, enabling the business owner to make better-informed decisions.

* **About the Algorithms Used**

1. SVM - SVM or Support Vector Machine is a supervised machine learning technique used for classification and regression. Finding a hyperplane in an N-dimensional space that classifies the data points is the goal of the SVM method. The number of features determines the hyperplane's size.

2. Ridge Classifier - Ridge classification is a method used in machine learning to assess linear discriminant models. In order to prevent overfitting, this type of normalization limits model coefficients.

3.Random Forest - Random Forest is a classification algorithm that uses multiple decision trees on smaller sets of the input dataset and averages the results to enhance the dataset's prediction accuracy.

4.XG Boost - Formally speaking, XGBoost may be described as a decision tree-based ensemble learning framework that uses Gradient Descent as the underlying objective function. It offers excellent flexibility and efficiently uses computation to produce the mandated results.

* ***BUILDING THE MODEL IN JUPYTER NOTEBOOK***
* ***(CODE FOR THAT IS UPLOADED)***
* ***(OUTPUT SCREENSHOTS FILE IS ALSO ATTACHED)***
* **What’s Next**

To enhance the project further -

Flask - Interface implementation can be done

Model - Other models can be trained and tested for a better accuracy

Data - In this case, we have dropped some columns; they can be further converted to categorical features and used

* **Conclusion**

Customer Churn Prediction Project determines when a customer is about to leave the organization.

Random Forest and XGBoost give us the best accuracy in this situation.