### **Documentation of the project**

### **Customer churn prediction**

* **Problem statement :**
* To build a predictive model to estimate the likelihood of customer churn for a telecom company .
* Considering factors such as customer demographics, usage patterns ,billing history, and customer service interaction .
* **Dataset used:**
* <https://www.kaggle.com/datasets/blastchar/telco-customer-churn>
* **Model selected:** machine learning classification model

### Introduction

* **Background:** Customers who left within the last month – the column is called Churn
* **Problem Statement:** predicting customer churn using factors such as demographics, usage patterns, billing history, and customer service interactions.

### Data Collection and Preparation

* By using pandas in python read the csv file
* Checking shape and size of the data

Data Preprocessing:

* + Handling missing values(found no null and duplicate values)
  + Encoding categorical variables using label encoder by importing feature engine
  + Normalizing or scaling numerical features using minmax scalar from sklearn.preprocessing
  + Removing duplicates and outliers(no outliers )

### Exploratory Data Analysis (EDA)

* **Visualization:** Included visualizations (e.g., histograms, box plots, displots) to illustrate data distributions and relationships.

### Model Selection and Training

* **Algorithms Considered:** List of machine learning algorithms evaluated (e.g., logistic regression, random forest, extreme gradient boosting, adaboost, decision tree,SVC).
* **Model Training:**
  + Train-test split
  + Cross-validation approach
  + Serialized the model and saved it
* **Hyperparameter Tuning:** the hyperparameter tuning methods that are used such as grid search.

### Model Evaluation

* **Evaluation Metrics:** The metrics used to evaluate model performance (e.g., accuracy, precision, recall, F1-score).
* **Results:** the accuracies of the following models are

Logistic Regression training accuracy is 80

Logistic Regression testing accuracy is 79

decision tree training accuracy is 82

decision tree testing accuracy is 77

random forest train accuracy is 84

random forest testing accuracy is 80

svc training accuracy is 81

svc testing accuracy is 79

XGB CLASSIFIER training accuracy is 99

xgb classifier testing accuracy is 76

adaboost training accuracy is 73

adaboost testing accuracy is 73

* **Model Comparison:** By Compare the performance of different models I would justify the selected model is XGB Classifier.

**Deployment**

First Step: Preparing and Preserving Our Model

1. Using our favorite machine learning library (such as scikit-learn or TensorFlow) to train our model.

2. Using a serialization format like pickle or `joblib{ to save the trained model

Step 2: Setting Up the Scene

1. Setting up the Azure CLI

2. Enter our Azure credentials.

Step 3: Docker Image Creation

1. Setting up our Flask application's `Dockerfile}

2. Making a file called requirements.txt

3. Developing a prediction-serving Flask application

4. Constructing the image of Docker

5. Local Docker container testing

The Docker image is pushed to the Azure Container Registry (ACR) in step four.

1. Establishing a registry for Azure containers

2. Enter ACR and log in.

3. Adding a Docker image tag

4. Sending the ACR Docker image

Step 5: Setting up Azure App Service deployment

1. Setting up a plan for Azure App Service

2. Making a web application

3. Setting up the web application to utilize the container image

Step 6: Implementation Testing

1. Locate our online application's URL.

2. To obtain predictions, send a POST request to the `/predict} endpoint of the Flask app

**Summary and conculsions:**

In conclusion, the project successfully developed a robust predictive model for estimating customer churn, providing valuable insights into the factors driving churn and enabling the telecom company to implement effective retention strategies. This initiative not only helps in retaining customers but also enhances overall