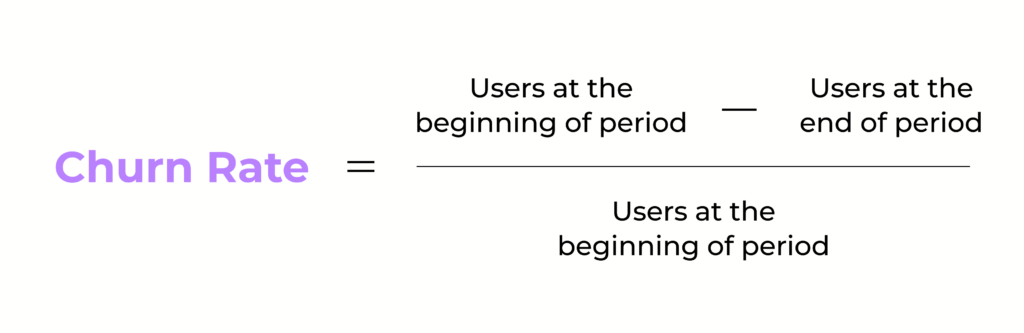
**Bank Customer Churn Prediction**

**Introduction About Project:**

Bank customer churn is when a customer closes their account or stops using a bank's products or services. It is a measure of how many customers a bank loses over a period of time. Churn can be measured as a percentage of the total customer base, or as the number of customers who leave.

Measuring customer churn:



This repository contains code and resources for predicting customer churn in a bank using machine learning techniques. The goal is to identify customers who are likely to churn (i.e., close their accounts) in order to take proactive measures and retain them.

**Project Flow**

The project follows the following flow:

***Problem Statement:***

The problem statement for bank customer churn is as follows:

**Problem:** The bank is losing customers at an unacceptable rate. In the past year, the churn rate has been 10%, which is significantly higher than the industry average.

**Goal:** The goal of this project is to develop a model that can predict which customers are most likely to churn. This information will be used to target retention efforts at the customers who are most at risk of leaving.

**Success Metrics:** The success of this project will be measured by the accuracy of the churn prediction model. The model should be able to predict which customers are most likely to churn with at least 80% accuracy.

**Constraints:** The model must be developed using publicly available data. The data must be anonymized to protect the privacy of the customers.

***Data Gathering:*** The necessary data is collected, which includes features such as credit score, region, estimated salary, gender, age, balance, number of bank products used, tenure, credit card usage, and active membership.

***Data Preprocessing and EDA:*** The collected data is preprocessed and explored through exploratory data analysis (EDA) techniques. This involves handling missing values, encoding categorical variables, and examining relationships between variables.

***Feature Engineering****:* Additional features are created or extracted from the existing dataset to enhance the predictive power of the model. This step involves transforming and combining existing features or creating new features based on domain knowledge.

***Feature Selection/Extraction:*** The most relevant features for predicting customer churn are selected or extracted. This helps reduce dimensionality and focuses on the most informative aspects of the data.

***Model Training:*** Machine learning models, such as logistic regression and K-nearest neighbors (KNN), are trained on the preprocessed data. These models learn patterns and relationships in the data to make predictions.

***Model Evaluation****:* The trained models are evaluated using various performance metrics, including accuracy, precision, recall, F1-score, confusion matrix, and ROC curve. This assessment helps gauge the effectiveness of the models in predicting customer churn.

***Web Development Framework****:* An API is developed using Flask, a web development framework in Python. This allows the trained model to be deployed and accessed through a web interface.

***Deployment to Cloud Platforms:*** The Flask application, along with the trained model, is deployed to a cloud platform, such as AWS EC2 instance. This enables the application to be accessible online.

**Requirements:**

To run the code in this repository, you need the following: To run the code in this repository, the following requirements are needed:

Flask==2.3.2 numpy==1.23.5 pandas==1.5.3 scikit-learn==1.2.2 plotly==5.14. Python 3.10.3 Jupyter Notebook (for running the code)

To install the required Python packages, run: pip install -r requirements.txt

**Usage**

Clone this repository: git clone https://github.com/SagarWadikar/churn\_prediction.git

Navigate to the project directory: cd churn\_prediction

Run the Jupyter Notebook: jupyter notebook

Open the Churn\_Prediction.ipynb notebook in your browser.

Follow the instructions in the notebook to preprocess the data, train the machine learning model, and evaluate its performance.

**Model**

We use a supervised learning approach to predict customer churn. The notebook contains code for data preprocessing, feature engineering, model training, and evaluation.

The model used for churn prediction is a Logistic Regression, KNN algorithm which has shown.

**Evaluation**

The notebook includes code to evaluate the trained model using various performance metrics such as accuracy, precision, recall, and F1-score. Additionally, it provides a confusion matrix and an ROC curve for further analysis. These evaluation metrics help assess the effectiveness of the model in predicting customer churn.

Keep in mind that the performance of the model may vary depending on the characteristics of the dataset and the chosen algorithm. It is recommended to assess the model's performance on multiple evaluation metrics and consider the business context when interpreting the results.

**Web Development and Deployment**

The Flask application is created with the necessary routes and functions to serve the churn prediction model as an API. Configuration for web development is handled using Flask. The application can be deployed to a cloud platform, such as an AWS EC2 instance, to make it accessible online.

References:

https://www.kaggle.com/code/nasirislamsujan/bank-customer-churn-prediction