

## Customer Churn Prediction

The main goal of this project is to predict the probability of customer churn using the provided bank database, by applying machine learning models and comparing their performances.

General Steps of the Project:

1. Importing and Initial Exploration of the Dataset
  - a. Importing the initial data from the CSV file
  - b. Exploring the general information of the DataFrame and identifying missing values using tables and Heatmap visualization
2. Outlier Detection and Removal
  - a. Identifying numerical and non-numerical columns
  - b. Calculating the Interquartile Range (IQR) to detect outliers
  - c. Removing outliers using the Z-score method
  - d. Displaying and comparing statistical information of the original and cleaned data
3. Demographic Feature Analysis: Gender and Geographical Location
4. Statistical Analysis
  - a. Investigating the relationship between features such as gender and geographical location with churn rate
  - b. Plotting frequency charts for categorical columns
5. Data Preprocessing and Preparation
  - a. Converting non-numerical variables into model-usable data (One-Hot Encoding)
  - b. Normalizing data (using Min-Max and Standardization/Z-score) for better modeling

## 6. Comparing Feature Distributions Before and After Normalization

- a. Plotting histograms to show the distribution of numerical features before and after normalization

## 7. Splitting the Data for Training and Testing

- a. Dividing the data into training (80%) and testing (20%) sets

## 8. Training and Evaluating Classic Machine Learning Models

- a. Logistic Regression
- b. Decision Tree
- c. XGBoost Classifier
- d. Gradient Boosting Classifier

Each of these models will be trained on the training set and their results will be evaluated on the test set using metrics such as Accuracy, Confusion Matrix, and Classification Report.

## 9. Comparing Model Performances Based on Accuracy and Evaluation Metrics

## 10. Design and Training of a Multi-Layer Neural Network (MLP)

- a. With the aim of increasing prediction accuracy compared to traditional models

## 11. Analysis of Results and Feature Interpretation

- a. Investigating which features have the most impact on customer churn
- b. Comparing the accuracy of the models and selecting the optimal model

## 12. Complete Documentation of the Project

## پیش‌بینی ریزش مشتری

هدف کلی پروژه پیش‌بینی احتمالی ریزش مشتریان بانک اطلاعاتی پیوست شده در پروژه با استفاده از مدل‌های یادگیری ماشین و مقایسه عملکرد مدل‌های مختلف می‌باشد.

مراحل کلی پروژه :

۱- خواندن و بررسی اولیه اطلاعات دیتاست

a. وارد کردن داده‌های اولیه فایل CSV

b. بررسی اطلاعات کلی Data Frame و شناسایی مقادیر گم‌شده (Missing Values) با

استفاده از جدول و نمودار Heatmap

۲- بررسی داده‌های پرت (Outliers) و حذف این مقادیر

a. شناسایی ستون‌های عددی و غیراعدادی

b. محاسبه IQR (فاصله بین چارکها) برای تشخیص داده‌های پرت

c. حذف داده‌های پرت با استفاده از Z-score

d. نمایش و مقایسه آماری داده‌های اصلی و تمیز شده

۳- تحلیل ویژگی‌های جمعیت شناختی جنسیت و موقعیت جغرافیایی

۴- تحلیل آماری

a. بررسی ارتباط ویژگی‌های جنسیت و موقعیت جغرافیایی با نرخ ریزش

b. ترسیم نمودارهای فراوانی برای ستون‌های Categorical

۵- پردازش و آماده‌سازی داده‌ها

a. تبدیل متغیرهای غیر عددی به داده‌های قابل استفاده برای مدل‌ها (One-Hot Encoding)

b. نرمال‌سازی داده‌ها (Min-Max و Standardization/Z-score) برای مدل‌سازی بهتر

۶- مقایسه توزیع ویژگی‌ها قبل و بعد از نرمال‌سازی

a. رسم هیستوگرام برای نمایش توزیع ویژگی‌های عددی قبل و بعد از نرمال‌سازی

۷- تقسیم‌بندی داده‌ها برای آموزش و تست

a. تقسیم داده‌ها به مجموعه آموزش (۸۰٪) و تست (۲۰٪)

۸- آموزش و ارزیابی مدل‌های کلاسیک یادگیری ماشین

a. Logistic Regression

b. Decision Tree

c. XGBoost Classifier

d. Gradient Boosting Classifier

هر کدام از این مدل‌ها روی دیتاست آموزش می‌بینند و نتیجه‌شان روی داده‌های تست با پارامترهایی

مانند دقت (Accuracy)، ماتریس آشفتگی (Confusion Matrix)، و گزارش معیارهای طبقه‌بندی

(Classification Report) ارزیابی می‌شود.

۹- مقایسه عملکرد مدل‌ها بر اساس دقت و معیارهای ارزیابی

۱۰- طراحی و آموزش شبکه عصبی چندلایه (MLP)

a. با هدف افزایش دقت پیش‌بینی نسبت به مدل‌های سنتی

۱۱- تحلیل نتایج و تفسیر ویژگی‌ها

a. بررسی اینکه کدام ویژگی‌ها بیشترین تاثیر را در خروج مشتری دارد

b. مقایسه دقت مدل‌ها و انتخاب مدل بهینه

۱۲- مستند کامل پروژه

