**Credit Card Approval Prediction using ML**

Name : Sara Samaninia

Student Number : 501090785

Supervisor’s Name : Tamer Abdou

Date of Submission : 2022-09-26

Credit score cards to control the risk are based on historical data. Due to large economic fluctuations, previous built models may lose their original predictive power. With the development of machine learning algorithms, more predictive methods such as Boosting, Random Forest, and Support Vector Machines have been introduced into credit card scoring. In this project we will work to build a new machine learning model to predict if an applicant is classified as 'good' or 'bad' client, while the definition of 'good' or 'bad' is not given. We should use some technique, such as vintage analysis to construct the label.

There are two tables, which are connected by ID; application\_record.csv contains appliers personal information and credit\_record.csv contains the payment records for a given client. First table has 18 and the second one has 3 columns(attributes) which we use as features for predicting. The dataset has a combination of numerical and non-numerical features, values from different ranges, also it contains missing entries. The number of records in application\_record.csv is 438,557, but it contains duplicate IDs and it only has 438,510 unique IDs. Credit\_record.csv contains 1,048,576 records with 45,985 unique IDs. There are only 36,457 IDs or clients that are common in both datasets and can be used for judgment.

So, we have to pre-process the dataset to ensure the machine learning model is good and accurate in prediction as possible. After our data is pre-processed, some EDA(exploratory data analysis) and vintage analysis will be done to build some intuitions.

Eventually we will build a machine learning model that can predict if an individual's application for a credit card will be or should be accepted or not. Different algorithms such as Logistic Regression, Decision Tree, Random Forest, KNN, SVM, LightGBM and Xgboost will be applied in **Python** to build the models and then the accuracy of the models can be compared.

Dataset References:

<https://www.kaggle.com/code/rikdifos/credit-card-approval-prediction-using-ml/data?select=application_record.csv>

<https://www.kaggle.com/code/rikdifos/credit-card-approval-prediction-using-ml/data?select=credit_record.csv>