

Machine Learning I (Final Project)

Objective:

- We can use the loan prediction dataset to detect the possible defaulters for Consumer Loans.
- In addition, you will use the regression analysis to predict the income of a person using various features provided in the dataset.

Dataset: Loan Prediction Based on Customer Behavior dataset

- Available at: <https://www.kaggle.com/subhamjain/loan-prediction-based-on-customer-behavior>
- The dataset consists of 13 different attributes for the customer such as age, marital status, income, experience ...etc and one target variable indicating the risk flag of being a loan defaulter.

You are requested to do the following **for both the regression and classification problems:**

Task 1: Classification

- 1- Load the data and perform all necessary data cleaning and scaling.
- 2- Data inspection. Use any relevant functions that can help you to understand the data. Use any necessary visualization techniques to inspect your data
- 3- Explore the selection of various feature variables for classification. You should include at least one categorical feature.
- 4- Classify the data using various classification methods explored in ML1 (logistic regression, SVC, Decision trees, KNN classifier). Explore using different model parameters in the built-in sklearn libraries.
- 5- Explore the use of your own implementations of each Model. Comment on your results.
- 6- For each model provide suitable quantitative metrics for assessing the performance of your model based on the required application.

Task 2: Regression Analysis

- 1- Load the data and perform all necessary data cleaning and scaling.
- 2- Data inspection. Use any relevant functions that can help you to understand the data. Use any necessary visualization techniques to inspect your data
- 3- Explore the selection of various feature variables for regression to estimate the income of each customer. You should include at least one categorical feature.

4- Perform various regression analysis using various methods explored in ML1 (Linear regression, Multiple regression, SVR regression, Polynomial regression). Explore using different model parameters in the built-in sklearn libraries.

5- Explore the use of your own implementations of each Model. Comment on your results.

6- For each model provide suitable quantitative metrics for assessing the performance of your model based on the required application.

.