**Loan prediction Project**

* **TA: Dr . Verena Nashaat**
* **Team members**

|  |  |  |
| --- | --- | --- |
| **Section no.** | **ID** | **Names** |
| **Section 16** | **20201700449** | **عبدالرحمن محمد نجيب محي الدين** |
| Section 4 | 20201700102 | **أدهم مصطفى محسن محمد** |
| Section 31 | 20201700909 | **نانيس عادل شحاته حسن** |
| Section 7 | 20201700176 | **باسم عاطف السيد محمد** |
| Section 4 | 20201700091 | **أحمد وائل صلاح الدين حسن** |
| Section 34 | 202017001019 | **يوسف عبد الرؤوف أمين محمود** |
| Section 6 | 20201701060 | **أندرو عصام جورجي عطا** |

**Introduction**

**Topic:**

Loans are the core business of banks. The main profit comes directly from

the loan’s interest. The loan companies grant a loan after an intensive

process of verification and validation. However, they still don’t have

assurance if the applicant can repay the loan with no difficulties.

**Goal:**

Predict if the user can take loan or not from the given features.

**Dataset description:**

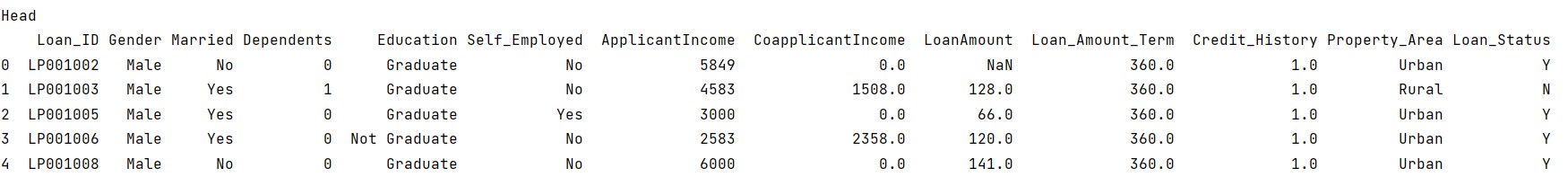
The dataset is composed of 614 persons with their Loan ID Gender , Married Dependents , Education, Self Employed ,Applicant Income Co-applicant Income ,Loan Amount ,Loan Amount Term ,Credit History Property Area and Loan Status.

**The project processes**

First:

We started by importing the libraries as (pandas and sklearn)

then the dataset loading and displaying the head of data



then the description of it (count, std and mean)

Text

Description automatically generated

Then we counted null values

Table

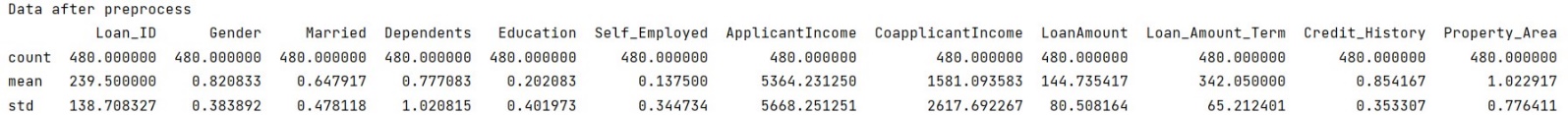
Description automatically generated

Second:

Pre-process:

Then we removed the null values from the dataset.

We made label encoding to the data by changing the data which has string values by giving it an integer value then We did the data scaling and displayed the new data description after the scaling.



Then we split data to x and y where x means the data which the prediction depends on it so it contained the data set except Loan id and Loan status and y means the predicted data which is loan status

Then we split data to 80% for train and 20% for test

So, data become as the following



Third:

Classification:

We made algorithms to calculate the accuracy of data like and print the accuracy before and after scaling:

support-vector machine (SVM).

SVM algorithm accuracy is equals to:

before scaling= 0.5833333333333334.

after scaling = 0.7291666666666666.

“Decision Tree Classifier “we made the decision tree to classify the data and calculate the accuracy of these data which is

before scaling=0.6666666666666666.

after scaling=0.7604166666666666.

“Logistic Regression” we made logistic regression to return the probability value by sigmoid function which gave us an accuracy equals to:

before scaling=0.6666666666666666.

after scaling=0.7604166666666666.

Fourth:

extracted a new feature principal component analysis (PCA).