Report of Project

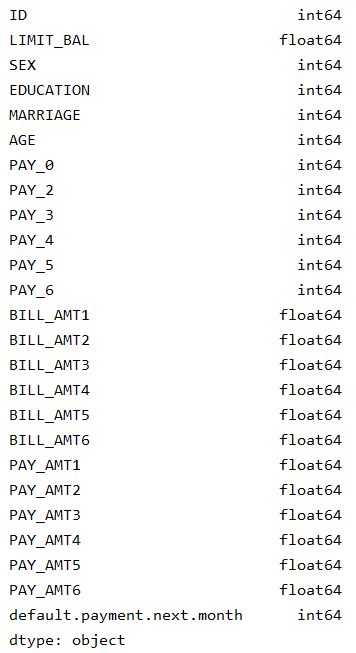
Elif Yılmaz

30716011

The source is **https://raw.githubusercontent.com/meauxt/credit-card-default/master/credit\_cards\_dataset.csv**

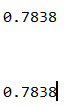
This research aimed at the case of customers and compares the predictive accuracy of probability of default among six data mining methods. From the perspective of risk management, the result of predictive accuracy of the estimated probability of default will be more valuable than the binary result of classification - credible or not credible clients.

Dataset has 25 features and 30000 inputs, so elements of number is 750000. This dataset also consists of 24 inputs and 1 output where the inputs is represents the attributes and the output is represent the default payment next month. All input types are numerical.

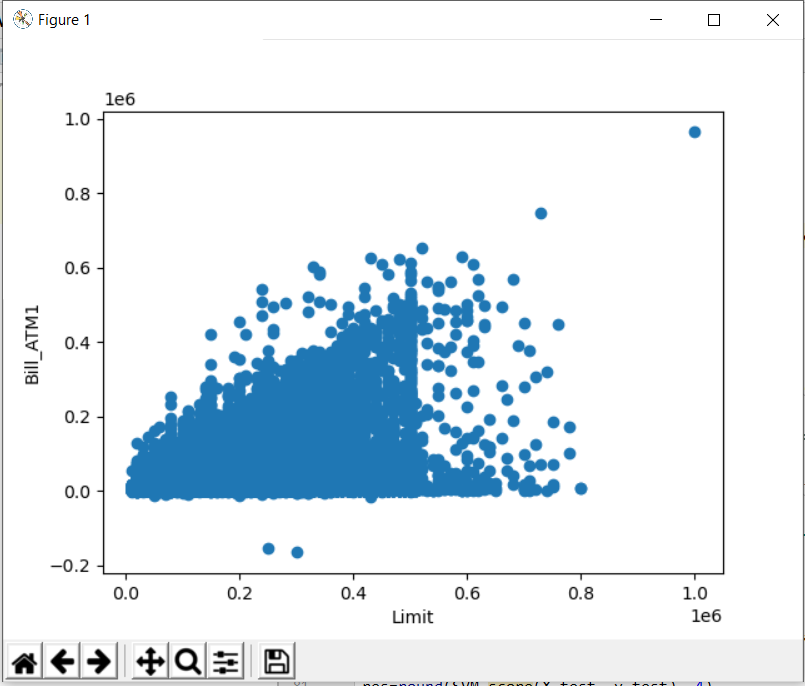


It was a Classification problem so I implement different algoritm to see which one is better for my dataset.

Linear Regression is not suitable for my dataset, because so much inputs I have. LR was insufficient for this. Support vector machine and Neural Networks gave same results. And Kernel(Linear),I wrote true codes but not running . So I choose SVM for apply on dataset.



Preprocessing means removing unefficient inputs so I look all datatypes, all of them are numerical. If there were any sitring, ı would delete that column.



Here I created a scatter using the inputs of X1(limit\_bal) and X12(bill\_atm1) columns.

For make predict, I needed to target. I should compare my inputs with this target. And decided to person is reliable for payment of credit. For that, I diveded dataset to X and Y. Y was target, I made prediction on that. I split data into training, testing.

Then used SVM for training, and make prediction to the processed data.

So I have 0 and 1 in predicitons. 1 means they paid, 0 means no paid.

