# **Report for Guido Rossum on CreditOne customer data**

With the data provided and the business criteria evaluation we have tried to address the solution to CreditOneproblem in a few different ways .

* 1. Regression task – To Predict how much to credit to approve customer based on their demographic data
  2. Classification – To Predict if a customer is “At Risk “ Meaning would the customer be Default on payments or not.

**Regression Method :**

As Limit Balance of customer was a continuous variable, used

Linear Regression , RandomForestRegressor , SVR Algorithms to build a model. The model did not yield satisfactory results.

Random forest yielded performed better than others but still had a low score for R2, Hence changed our strategy to address the issue as a Classification.

A picture containing circle

Description automatically generated

**Classification Method :**

I have proposed 3 options to address the CreditOneproblem. Used DecisionTree classifier and RandomForest classifier for all of the below modeling.

## Classification model #1

DV - LIMIT\_BAL

1. Discretized LIMIT\_BAL, LIMIT\_BAL as category

2. Default has both default and non-default customers

Discretized limit balance into 4 categories with mostly equal number of data. Random forest accuracy score for predictions was 58%. Since this was low moved on to the next model to see if performance increased.

## Classification model #2

DV - LIMIT\_BAL

1. LIMIT\_BAL Discretized as categorical variable

2. Filtered Data to to have Default customers ONLY

This increased the accuracy score for DT – 54% and RF – 62%. But this may not be very effective since we only used one group if customers.

## Classification model #3

DV - DEFAULT

Finally built a model to see if we could classify customers if they are a

risk to CreditOne. The model performed as expected and using this

model we are confident we can predict if a customer would be default

or notTable

Description automatically generated.

* After preprocessing of data it was determined to be a regression modeling task.
* Used LinearRegression, RandomForest, SVR algorithms to build models.
* After multiple efforts tuning and normalizing data the model performance did not meet the standards.
* Hence it was deemed necessary to convert this to a classification problem to just classify if a customer was potentially Default or not default customer.
* Model performed at 80% on test set, which also satisfies the business criteria if we are unable to predict the loan amount our other option would be to flag customer as “at Risk”.

**Conclusion :**

Model #4 Performed satisfactorily and can predict if a customer would fall into the category ‘Default’ or ‘Not Default’ .