# **Project Meeting Minutes - III**

Date of meeting(s): **24-03-2023**

Group Number: **11**

Group members present (Name, ID):

* **Vaibhav Patel (0772934)**
* **Avikumar Patel (0790966)**
* **Yashkumar Patel (0797825)**
* **Smit Rana (0792056)**

**Specific Activities that were completed/worked on:**

* After handling outliers and cleaning up the data, our most recent data appears to have more potential for machine learning models.

**Table

Description automatically generated with low confidence**

*\*NaN implies that these columns were eliminated (makes no sense).*

**Chart

Description automatically generated**

* The preceding diagram depicts the distribution of Num\_of\_Delayed\_Payments before and after handling outliers. We did the same thing with other variables that had wrong values or were too far away from other values.

**Specific Output from work:**

* Several columns contained incorrect numbers, and we need to place them in a certain range that demands accurate information related with the financial business. We understand each of them and handled those false value and outlier numbers.
* Subsequently, we examined the correlation between each variable and discovered a few unwelcome associations between a few of them. This issue is resolved with the aid of MinMax Scaling. This estimator scales and translates each feature independently to the [0,1].

**On Target:**

* Indicate the current status of your project:

**✅ green:** everything on track for completion by due date

yellow: a small number of tasks are off track and completion by due date is at risk.

red: many tasks are off track, and the project will not be completed by due date.

**Challenges/Disagreements:**

* The most difficult aspect of this week was evaluating figures in a few of the columns in the data were inaccurate, and now we need to fit those numbers into a specific range that necessitates accurate information concerning the financial industry.

This required a considerable amount of effort to comprehend the meaning of each column; does this have an effect on the variable of interest (credit score)? , other relevant variables, true value range, and so on.

**Planned Activities for coming week:**

* Already, Smit and Yash have begun data modelling and still left with few other ML models. They are incorporating standardized data into various machine learning models. After training, the accuracy of the model is evaluated on test and training data using a confusion matrix. Using a tiny random sample of the x\_test data, we predict the value (y\_pred) and compare it to y\_test. This step is simply to determine the prediction behavior of each model.

Graphical user interface, application

Description automatically generated

* Values predicted by credit score (-1,0,1 represents bad, standard, good status of credit score respectively). The names of columns are numbers that denote the row number of data (randomly).
* This action will be executed for each model. After doing a prediction on a small data sample, Avi and Vaibhav will generate a feature importance bar plot to see which features contribute the most to credit score prediction.

A picture containing graphical user interface

Description automatically generated

* So, based on this information, the most accurate model will be utilized for model deployment.