Rebuttal of Group 5

a) There wasn't a data processing procedure like eliminating errors and missing data before calculating the Pearson correlation.

We think we just arrange the orders of the code in a wrong way when we combine all the things together and use functions to simplify the code. The calculation of correlation is done after data cleaning.

b) They didn't use ROC curve to evaluate their models.

We use the AUC value, which is exactly Area under the Curve of ROC, so we believe using AUC score independently is good enough to express the accuracy of our models.

(1) Feature Engineering and feature selection can be considered at the beginning of the report, that can make the report more logical, rather than simply as a means to improve the effect of the model.

We aim at doing the project in a simple way in the beginning, so we first choose to do a little thing on the data and see what the result is. After that we then try more complex methods and see if any improvement happens. We think the logic is quite clear.

As they mentioned, most related or most important features are listed, without reasonable explanation on how these features affect credit default.

The important features are given by the classification model, and we mention at the last part that we lack the knowledge in real-world business so we cannot explain them quite clearly. Therefore, we tend not to make up the explanations.

there are many figures but not enough examples

We don't quite understand the meaning of "example" here. We have already used the information in the data to illustrate our steps in our report like the 2 best reports.

But the authors did not assess the strengths, and the analysis of weakness is not clearly enough.

We use nearly a whole page to explain the strengths of each model and different ways of dealing with the missing data. The weaknesses are discussed at the end of our report. Again, we don't know the exact interpretation of "strength" and "weakness" here even we have a look at the 2 best reports.