

# Review1 Group11

**HUANG Wenjin (Student ID: 20742991)**

## 1. Summary

Briefly speaking, I think the report of Group 11 is pretty a good article. Not only its academic expression but technical quality, they are both reach the requirement and principle of our course. Beyond this, they also did something more than the basic one, which really showed their creative enough through this work. However, little flaws still exist on the report. Fortunately, these would not obscure the goodness of their group mind.

## 2. Pons & Cons

This report seems to perform better on its technical side and expression content.

First, the article contains details information easy to understand by new-learner of machine learning, such as clarifying the meaning of Logistics Regression, in order to explain the contribution of features.

Second, the experiments of the paper have a rigorous logic and high trust worthy. In PCA analysis part, I could clearly find the logic to deduce dimensions so that they can dig out useful variables. But it fails to do so because it is unsupervised without label information. It is the same result of mine.

Third, they tried to use multi-evaluation proxies in the paper. Not only ROC curve, but also FPR, TPR and error rate are contained.

However, there is still a little point to be improved, as far as I suggest.

The typesetting of the paper could be irregular, at least, not from the latex model provided by our professor. Actually, I agree on your trial to use this more official typesetting way. But the "Abstract" word is looked a little bit weird. The polish work thus could be able to be improved by better picture setting and table formatting.

## 3. Clarity and quality of writing

**I think it should be 5 points in this part.** Here is my point of view with suggestion.

- Is the report clearly written? Its details information could contain enough understandable messages to a beginner.
- Is there a good use of examples and figures? There are many pictures to show any detail in

their experiment, but some tables are mixed inside the figure. Maybe they should format the table by latex instead.

- Is it well organized? All the experiment logic, from the failure to success, is written step by step. Its structure could also be accepted by the index at the beginning easily.
- Are there problems with style and grammar? Though the model of latex is different, there is not any grammar or style error of word.
- Are there issues with typos, formatting, references, etc.? Nothing wrong as far as I am considered.

## 4. Evaluation on Technical Quality

**I think it should be 4 points in this part.** Here is my point of view with suggestion.

- Are the results technically sound? Reasonably, by the experiment we could find the quadratic regression model could perform better than tree model (the same as us) because of the properties of dataset. So the result that uses logistics regression indeed looks good. But the variable selection seems not clear enough.
- Are there obvious flaws in the reasoning? No any. The logic of experiment seems to so clear and make others easily follow it. Each state about each classifier like XGB, has no obvious error on theoretical and logical area.
- Are claims well-supported by theoretical analysis or experimental results? Yes, especially the experimental results are quite trust worthy and make great contribution on the verification to the claims. But the article has seldom theoretical support.
- Are the experiments well thought out and convincing? Yes, the experimental results really achieve my expectation that quadratic regression model could perform better than tree model.
- Will it be possible for other researchers to replicate these results? I think it could be easily replicated by using 5 fold logistics regression. But it still has a difference on features selection, because in the article, they use a large number of features in final models. So it can be quite different from us or other groups.

## 5. Overall rating and confidence of mine on assessment

**I think it should be 4 points overall.** Here is my point of view with suggestion. I keep a level 3 confidence on my comment. I made this evaluation prudently by reading it many times.