Group that you review: Group 16

Reviewer's name: TANG Tsz Hong

Reviewer's ID: 20735194

Summary:

Group 16 finished Project 1 in poster style, with Python Notebook attached. Overall, I will rank this report 3.75/5, with good performance on clarity and quality of writing, and satisfactory performance on technical aspect.

They use random forest classification model as the main model. In Feature Engineering part, they added several average score of external sources scores as new feature. Also, making use of sin-cos transformation on the cyclic features. Finally, they choose the top 10 most correlated features, and uses SMOTE, under sampling to balance the uneven dataset.

For strength of the report, the main idea is clearly stated, figure and table make things clearer. Also, they have used good strategy to deal with the unbalance dataset.

For weakness part, I think one of the major problem of the report is there is no Kaggle score attached. One of the important criteria of this report or this assignment is to submit the result onto Kaggle and get the score.

Evaluation on Clarity and quality of writing (1-5): 4

The report is clearly written. Using table and bar chart to visualized data make things very clear. In terms of style, all spacing, format consistency are doing well. Small minor typo like in section 2, "stillhave" is missing a space, but overall is doing well.

For section 7, the title said "contribution" but the content seems not matching. For the requirement of the report, "with a clear remark on each person's contribution", this poster is missing this part. Maybe you can fill it back in the next report.

Evaluation on Technical Quality (1-5): 3.5

I think instead of balanced accuracy, AUC is a better choice to evaluate to model. One of the reason is AUC is more direct, we can use it to compare different models directly, while balanced accuracy is difficult to compare between methods when you use different method to analyse your data. Another reason is that Kaggle kind of using AUC to generate the score.

Also, in future reports, you can try different models to make the report more comprehensive.

Overall rating: 3.75

(5- My vote as the best-report. 4- A good report. 3- An average one. 2- below average. 1- a poorly written one).

Confidence on your assessment (1-3): 3

(3- I have carefully read the paper and checked the results, 2- I just browse the paper without checking the details, 1- My assessment can be wrong)