MSBD 5013 Project 1 Peer Review

Group: 7 Reviewer: Yuen Zhikun (20505288)

1. **Summary**

   Their project mainly focus on handling the data imbalance problem with SMOTE. They ensemble random forest and gradient boosting tree to reduce the variance and bias. The whole pipeline is simple and easy to be understood. They showed how they pre-process the data e.g. missing data, data analysis and feature engineering work, e.g. using square features to augment the data. The result is pretty acceptable for using a few data only.

2. **Strengths**

- Good to using desampling and oversampling approach to tickle the imbalance data and show the significant improvement with figure.
- The figures can really show some important insights, e.g. number of estimators does affect the result.
- They know decision tree is easy to have overfitting problem so they ensemble gradient boosting tree and random forest together.
- They handle the missing data with some common techniques such as filling mean. Also, they decided not to delete the data with miss features because too many data contains missing features.

3. **Weaknesses**

- They use polynomial features but did not mention the reason or any insight about it.
- They selected the features having highest correlation with target but did not mentioned which type of correlation.
- They did not examine some other data tables, it is possible that some of the important features are in other dataframes.
- Not enough experiments on fine-tuning the classifier.

4. **Rating:**

- Clarity and quality of writing: 4
  The poster has a nice design and format. The figures can shows the main points with limited spaces.
- Technical Quality: 4
  The classifier is simple but robust and also the sampling approach can in deep boost the result a little bit. However, you can try to use more features and examine different kinds of classifier and compare their results.
- Overall rating: 4
- Confidence: 3