

Summary of the report

Group 14 uses 8 models to do the classification. Their work contains data visualization, data cleaning, model construction and model evaluations. Data visualization mostly plots the distributions of different predictors and the label 'target'. Dropping outliers and filling 0 in NaNs are what they do for data cleaning. The application training data is split into training part and testing part for model construction so that they can calculate ROC when doing model evaluations. XGBoost performs the best among all models with 0.739 score on Kaggle contest.

Strength and Weakness

Overall, the job is well-done. The models perform good classifications with fast program efficiency. The report is written pretty well. One improvement can be made in *Figure 3: The plot of how different predictors affecting the TARGET*, since its label is hard to read. The work for data cleaning is super detailed. That must be helpful for later model construction and analysis. And they use quite a lot of models. However, feature selection is not clarified well. I am little confused about the reason why they choose those specific features.

Evaluation on Clarity and quality of writing: 5

Evaluation on Technical Quality: 4

Overall rating: 4.5

Confidence on my assessment: 3