Review4:

*• Describe the weaknesses of the report.*
*There are some issues about the report that I want to point out:*

- o *The report does not explain why they selected Logistics Regression, Decision tree and Random Forest as their model.*

Rebuttal:
The reason why selected those models is explained in the presentation. We select linear models because their simple structure reduce overfitting problems and tree models are more advantageous when dealing with high dimensional data. The aim is to compare different models' performance on the Titanic dataset.

Review5:

**Weakness:**

*It does not include analysis of why Random forest is the best performed model. It does not mention whether the accuracy in the table is test, validation or training accuracy.*

Rebuttal:
Our analysis of why random forest is best performed model is both mentioned at the *5. Conclusion and Discussion* in the report as well as in the presentation.
We are sorry that the second part is not clearly stated in the report. We divide the dataset into training and validation set. Then we use cross set validation on the training set to find a model with highest accuracy. After that the model's hyper parameter is tuned on validation set and then make prediction for submission on Kaggle.