

[/group_2/review_1.pdf]

1A Insufficient data visualisation (couldn't understand role of features)

All relevant features that I eventually chose to include in model experimentation have relevant exploratory plots (p. 2 distribution of ages), p. 4 (Sex, Age, Cabin class, number of parent/children, number of sibling/spouse, isMarried, Unique title) as well as histogram of gender vs survival.

Perhaps plots could be included for distribution of fare price and histogram of survival rate at each embarked point. However, Deck is a largely irrelevant feature, as is Name, so visualisation is unnecessary.

1B Number of training models inadequate

By number, do you mean model types, like LDA, Naive Bayes, and SVM? Certainly those could be used, but I believe experimentation and fine-tuning of those procedures in addition to the existing models would take too much time.

1C Training model is common, finetuning is rough, and result unsatisfactory

For a small dataset like Titanic, an overly complicated model (neural network, XGBoost) will overfit to the data (see Source [1]). I touched upon this point in p. 7 of the report and in the presentation by mentioning that using an overly complex AdaBoost estimator (high number of learners stacked on top of each other) causes overfitting and poor performance on the test set. Also, I tried using AdaBoost with random forest as opposed to simple decision tree as the base classifier, and the accuracy dropped even more. Occam's razor very much applies here.

What is meant by "finetuning is rough"? I have clearly outlined the procedures taken, which include backwards subset selection of relevant features, and grid search of relevant hyperparameters.

Performance: As the Titanic dataset is so commonly known, it is extremely easy to bitflip submissions, so 85%+ accuracy models on the leaderboard are unrealistic. Most Kaggle experts score in the 82%-85% range and 77.7% (my result) I would argue is fairly good.

Source [1]: <https://www.kaggle.com/carlmcbrideellis/overfitting-and-underfitting-the-titanic> (Links to an external site.)

[/group_2/review_2.pdf]

2A Organization: lack of conclusion

Agreed and thank you for the feedback.

2B Imbalanced discussion of different models

Section 5.2 on p.7 is attempting to refine logistic regression model with a subset of the existing predictors. AdaBoost, a new model architecture is also discussed in presentation and Section 5.1 of report and how it may not be the best choice for this project.

2C No reason given for model selection

Common model archetypes were chosen. However, SVM was not chosen due to lack of guarantee that data was linearly separable, and random forest offered same functionality but less variance than decision tree so only random forest was used.

2D Subset selection for AdaBoost

The same backwards subset selection as was used in random forest (all the predictors and their respective importances were the same.)

[/group_2/review_3.pdf]

3A Relating importance of features to real scenarios:

Some background knowledge would indeed be useful. However, this is very much a toy dataset and extrapolating which individuals are likely to survive to real disaster scenarios may not be suitable. Also, apparently the Titanic was the exception to the rule "every man for himself"; there is no legal basis for "women and children first" in maritime law. See:
<https://qz.com/321827/women-and-children-first-is-a-maritime-disaster-myth-its-really-every-man-for-himself/>

3B Weak argument for not using Precision+Recall:

Not sure what is meant here? We have no particular preference for predicting one class correctly over another (as you mentioned), so an accuracy metric should suffice.

3C Too wordy/lack of visualisation:

See section 1A. Also, perhaps a diagram to explain project workflow might be useful?

3D Organization of model selection and experiment:

Agreed and thank you for the feedback.

[/group_2/review_4.pdf]

no comment

[/group_2/review_5.pdf]

no comment