# Machine Learning for Survival Prediction of Passengers on the Titanic (Group 1)

Start from 15:20, data overview (sex/age, fare, embark, family, Name, cabin)
15:47 Data cleaning and processing
15:50Model selection (comparison, hyper parameter tuning, final performance(visual) )
15:53 Conclusion and discussion
End at 15:54

• Summary of the report.
The report investigates what kinds of passengers are more likely to survive on the Titanic. A dataset containing information about passengers' age, gender, and some other information is used. Several statistical models are discussed. Raw datasets are preprocessed with exploratory analysis, model building with diagnosis followed by selection. Random forest is the best model with 0.82 score on Kaggle.

• Strengths of the report.
Strong supportive data visualization (bar chart of importance) for the data analysis / processing section
Couple of models used (Logistic Regression, SVM, KNN)
Analytic skills are shown in model comparison taking the variable importance as support

• Describe the weaknesses of the report.
There are not enough details mention in the report especially in model section, only types of model is mention, however, the model configurations, hyperparameter settings, application methods did not mention in the report

• Evaluation on quality of writing (1-5): Is the report clearly written? Is there a good use of examples and figures? Is it well organized? Are there problems with style and grammar? Are there issues with typos, formatting, references, etc.? Please make suggestions to improve the clarity of the paper, and provide details of typos.
4
The report is written in a neat way with section number and subtopic, objectives and methodology are stated clearly. Lots of figures for visualization and demonstration. However, there is a typo exist for example "siginificance" in part 4.2. I would like to suggest that the details of model used and configurations should be explained instead of showing the model type and its result.

• Evaluation on presentation (1-5): Is the presentation clear and well organized? Are the
2 The presentation is not clear and not organized. As time stamp provided on the top, the time allocation is

extremely imbalanced which took near 30 minutes for describing raw data and data processing procedure while only 5 minutes left for model construction, model analysis, model selection and final result.
In the first 30 minutes, first speaker give explanation for different tag for example “S” mean “survive” and too much time on first page about the variable. In whole presentation data visualizes in variety of ways such as bar chart, table, heatmap. Showing some interesting fact(high fare) but not able to explain why exist, identifying special input( 7 members family share 1 fare) and providing treating method (divide fare by7)

• Evaluation on creativity (1-5): Does the work propose any genuinely new ideas? Is this a work that you are eager to read and cite? Does it contain some state-of-the-art results? As a reviewer you should try to assess whether the ideas are truly new and creative. Novel combinations, adaptations or extensions of existing ideas are also valuable.
3 The data analysis part is good for reader have easy understanding and may even inspire the reader, however, as mentioned above, the information of details of model is provided. The reader has no idea how to re-construction the model for same result and manipulate the result. That would be a great disadvantage for reader to cite this report if they do not have a clear understand of how the result come from.

• Confidence on your assessment
3, I have carefully read the paper and checked the results