Model selection and regularization on prediction of survival on the Titanic **(Group 6)**

• Summary of the report.

The report investigates how the model selection and regularization method works in classification problem. More specifically, it will focus on using logistics regression model combine with various model selection method, information criteria and regularization method. Raw datasets are preprocessed and model building with diagnosis followed by selection and exploratory analysis. Lasso is the best model with 0.770 score on Kaggle.

• **Strengths of the report.**

Strong supportive data visualization (line graph of ridge and lasso coefficient in regularization) for the model selection section
Couple of models used (AIC, BIC, Ridge regression and Lasso Regression)
Analytic skills are shown in model comparison comparing the line graph of coefficient and loss as support

• Describe the weaknesses of the report.

Overall, the report is well designed. Unfortunately, the model configurations, hyperparameter settings 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.
5
The report is written in a neat way with subtopic, objectives and methodology are stated clearly. Lots of figures for visualization and demonstration. However, there is a grammatical mistake exist for example "Several steps <u>has</u> been taken to" it should be "have" in Data Preprocessing part. 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
5 The presentation is clear and organized. The time allocation is balanced which divided presentation into 3 parts (data model and analysis). The presentation explains AIC, BIC with graphs and mathematical result. Also, Logistic approach with regularization is

mentioned.

It shows improvement needed which is good.

• 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.

5 The model part is good for reader have easy understanding and may even inspire the reader, speaker provided lots of information about the model used and the idea behind the selection. However, the configuration of details of model is not provided. The reader has idea how to re-construction the model but may need multiple of try for same result and manipulate the result. That would be a great presentation to reader and reader may cite this report.

• Confidence on your assessment

3, I have carefully read the paper and checked the results