Group 1 Summary:

This group use the Titanic data set to work on a classification problem.

They use several technique in data cleaning and data processing stage. For instance, dropping variable, fill in missing data with the average, parameter estimation and feature transformation.

In the model diagnosis stage, they check for multicollinearity with heatmap. They apply 5 different models in this project, include 1. Logistics regression, 2. KNN, 3. SVM, 4. Decision Tree and 5.Random Forest. Other technique they use in the model selection stage are forward selection, backward selection, hyper parameter tuning and cross validation.

The result of this project is that Random Forest model perform the best among all models. With all other model perform very similarly with around 80%-82% accuracy.

Strength:

It has compared the accuracy with several model clearly in a table, which is great. Also, it has do multicollinearity check with a heatmap.

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.

Evaluation:

Criteria	**Points**
Quality of writing	4
Presentation	3
Creativity	2
Confidence on my assessment	3

Comment:

1. The report is well-written with concise language and clear structure. Also, the figure is chosen reasonably.
2. The presentation could be more concise, especially in the data cleaning and data processing part.
3. Most of the project is quiet standard, without much novel idea.