Individual Learning Journal

Data Mining and Machine Learning
Heart Disease Classification Project

Junaidh Haneefa Muhammedhaneefa
MSc Computing Science
Griffith College

December 2025

My Contribution to the Project

My main task in this group project involved the background and analytical aspects of our heart diseases classification system. My personal actions included creating the Business Understanding section that would entail defining the problem statement and establishing the linkage with the UN Sustainable Development Goal 3. I performed all the data loading with the help of the UCI Machine Learning Repository and performed thorough exploratory data analysis, creating more than five visualisations such as target distribution charts, age analysis, correlation heatmaps, and chest pain type breakdowns.

Two of our four classification algorithms: Logistic Regression and Random Forest were implemented and tuned on the modelling side. In the case of Random Forest, I conducted hyperparameter optimisation with 5-fold cross-validation on GridSearchCV. I also designed the ROC curves comparison visualisation that enabled us to analyse the model discrimination ability of all classifiers.

In the written report, I authored the Introduction, Business Understanding, and Data Understanding sections, accounting for about 1,500 words. I also made slides 1–5 for our video presentation.

What I Learned

This project informed me a lot more than I had expected. Technically, I received experience in the full machine learning process- loading and evaluation of the model. The realisation of why feature scaling is relevant to particular algorithms came as important. I also picked up on how random_state is necessary to be reproducible.

The phase of exploratory data analysis was an eye-opener. The revelation that 77 percent of asymptomatic patients had heart disease questioned my assumptions about the association of symptoms with disease presence. This observation strengthened the idea that data derived solutions may uncover trends that are not apparent in intuition.

I also acquired enhanced project management skills. The ability to divide a big task into small ones and

work towards small deadlines was a good experience.

Challenges Faced

The biggest problem was related to deciphering the medical terminology within our dataset. Terms such as ST depression and thalassemia needed background research. The other complication was the limited dataset size of 303 patients.

Team Collaboration

It was overall an easy experience to work with Alfin. Sharing of duties was made clear at the beginning and this avoided unnecessary duplication. We communicated regularly to ensure our preprocessing processes and random seeds were consistent. His extensive analysis of data quality complemented my exploratory analysis nicely.

Contribution Evaluation

Team Member	**Contribution**
Junaidh Haneefa Muhammedhaneefa	50%
Alfin D Silva	50%

It was a true division of labour. All of our models required the preprocessing pipeline developed by Alfin, and the quality of his SVM and Decision Tree implementations were equally substantial as my work with Logistic Regression and Random Forest. This project was successful since we both fulfilled our promises.

Final Reflection

This project has strengthened my desire to use machine learning within the healthcare context. I take pride in the fact that we surpassed our 80 percent accuracy goal and delivered a holistic analysis consistent with global health goals.