Individual Learning Journal

Data Mining and Machine Learning
Heart Disease Classification Project

Alfin D Silva
MSc Computing Science
Griffith College

December 2025

My Contribution to the Project

The scope of my involvement in this project was literature review, data preparation, and the application of two of our four classification models. I started with the literature review, identifying and analysing 5 research papers related to predicting heart diseases using machine learning. This assisted in setting benchmarks on accuracy (78–93%) and informed our algorithm choice.

The data quality analysis was my responsibility and I reviewed details of missing values, outliers and feature distributions. This included visualising patterns of missing values and applying preprocessing steps with SimpleImputer and StandardScaler. It was essential to get the train-test split correct using stratification for equitable model evaluation.

I constructed and tested the Support Vector Machine (SVM) and Decision Tree classifiers. The SVM had to be tuned with GridSearchCV exploring various kernels, C values, and gamma values. For the Decision Tree, I also generated a visualisation of the decision rules which was useful in interpretation discussions.

My responsibilities also included generating confusion matrices for all four models and performing feature importance analysis. For the report, I composed the Data Preparation, Modelling, Evaluation, and Conclusions sections. I made slides 6–10 for our video presentation.

What I Learned

The insight that has had the most impact on me was that machine learning success relies on successful data preprocessing. Before this project, I believed preprocessing was a dull initial step. Today I know that choices regarding missing values, feature scaling, and train-test splitting directly influence model performance. Especially significant was the lesson about fitting the scaler only on training data to avoid information leakage.

The interpretability-accuracy trade-off also impressed me. My Decision Tree was not as accurate (81.97%)

as Random Forest (90.16%), but the ability to illustrate decision making has true value in medical practice where physicians must understand the model's logic.

Challenges Faced

The particularly challenging situation was the SVM training time in hyperparameter tuning - testing various kernel and parameter combinations took longer than expected. Sometimes, coordination on preprocessing with Junaidh was a problem as we needed identical random seeds and scaling techniques for fair model comparison.

Team Collaboration

Junaidh and I worked well together as we defined task ownership from the start. His EDA work was a source of useful information that informed my modelling choices. Regular check-ins were used to discuss progress and integration problems. I liked how Junaidh properly documented his code, making it easy for me to produce the combined confusion matrix.

Contribution Evaluation

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

The work was shared equally among us. Both contributions were vital to the project's success and neither could have completed the assignment single-handedly.

Final Reflection

This project provided me with real-life experience of the entire CRISP-DM process. I have gained confidence in applying machine learning to real-world classification problems. Above all, I learned that good data science involves not only technical expertise but also effective communication and collaboration.