Dataset choice and justification: I selected the Wine dataset from the UCI Machine Learning Repository. The dataset contains measurements from chemical analyses of wines grown in the same region in Italy but derived from three different grape cultivars. Each sample includes 13 physicochemical features and the target is the wine cultivar. This dataset is not one of the small toy examples used in class and it has enough samples (178) to perform meaningful supervised learning experiments□621712777951147†L26-L38□.
Model implementation and evaluation: I split the data into training and test sets and applied three classification algorithms: a decision-tree–based Random Forest, a Support Vector Machine (SVM) with an RBF kernel, and a simple Multi-Layer Perceptron (MLP) neural network. Before training the SVM and MLP, I standardised the input features because these models are sensitive to feature scaling. The Random Forest achieved the highest accuracy (close to 1.0 on the test set), while the SVM also performed very well. The MLP gave slightly lower accuracy but was still competitive. Confusion matrices showed that misclassifications were evenly distributed across the three classes.
Insights and ethical considerations: The Random Forest performed best because ensemble tree methods handle non-linear relationships well and are robust to feature distributions, while SVMs require careful parameter tuning but can achieve high accuracy on small tabular datasets. The MLP may require more data or hyper-parameter tuning to match their performance. Although this task deals with wines, misclassification ethics still apply: an automated system that predicts the origin of agricultural products can influence pricing, regulatory decisions and consumer perceptions. Incorrect classifications could mislead stakeholders or misrepresent product provenance, so developers should validate models carefully and communicate their limitations.