For Paper: ANN

Vanessa Zepeda

1 Section 2:

The Wine Quality (Red) datset consists of 1,599 observations and 12 variables, including 12 predictors and the quality target. An initial data quality assessment found no missing values across all features. However, duplicate observations were present and removed to prevent redundant information from biasing the analysis.
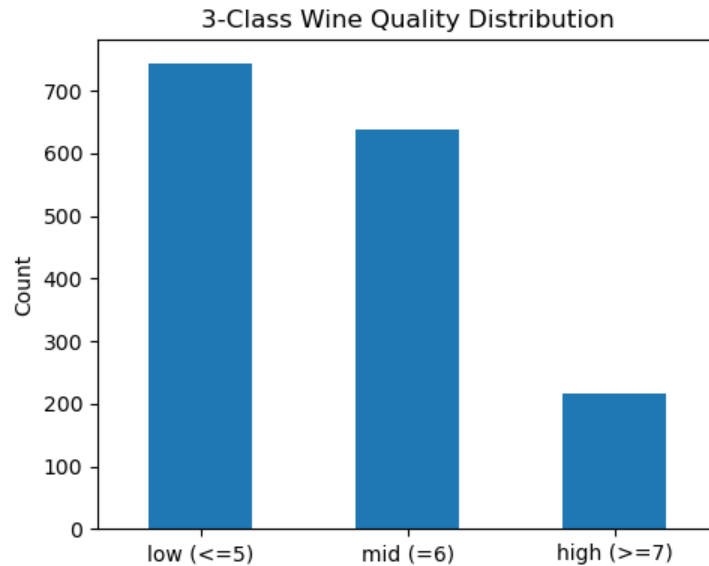

Figure 1: 3-class wine quality distribution

The original quality scores were visualized to assess their distribution which showed that most wines fall within a narrow range of ratings. To improve the interpretability, the quality variable was recoded into three classes: low (<=5), mid (6), and best (>=7). The resulting class distribution was imbalanced with 744 low-quality, 638 mid-quality, and 217 good quality observations. For each model, appropriate masures were taken to address the class imbalance.

Upon visual inspection of the predictor variables, their distributions revealed heterogeneity across variables. While density and pH were approximately normally distributed, most predictors were highly left or right skewed, indicated differences in scale and distributional shape. To address this, predictors were standardizaed during modeling to ensure that differences in scale and distribution did not disproportionately influence the results.

Box plots grouped by the three wine quality classes were used to visualize how predictor distributions differ across outcomes. For some variables, such as alcohol and volatile acidity, a noticeable shift in their median values across quality levels was obvious, suggesting a possible relationship between these features and wine quality where higher or lower values of these predictors tend to be associated with increases or decreases in
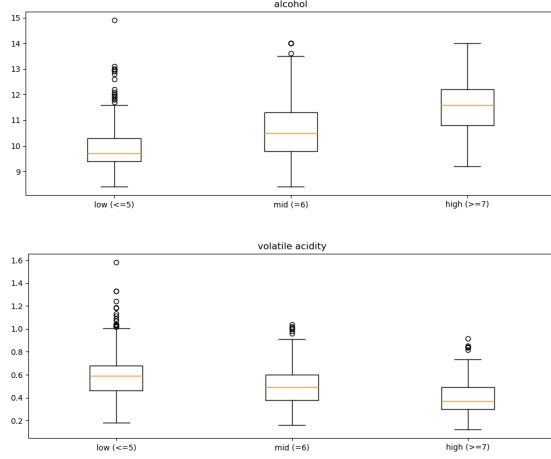
Figure 2: *

Box plots of alcohol (top) and volatile acidity
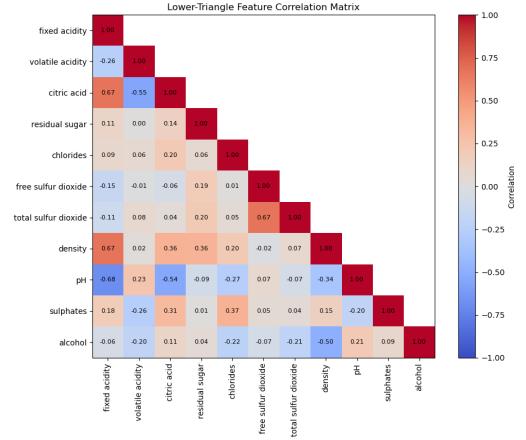(bottom) by wine quality class

Figure 3: *

Lower-triangle feature correlation matrix

Figure 4: Predictor distributions and correlation structure of features

quality. Other predictors showed substantial overlap between classes, indicating that these precidtors are less informative for differentiating between quality levels.

A correlation matrix was also used to visualize linear relationships among predictors. There were many moderate to strong correlations, including positive correlations between fixed acidity and citric acidc (0.67), fixed acidity and density (0.67), as well as negative correlations between pH and fixed acidity (-0.68) and between alcohol and density (0.50). Given the observed mutlicollinearity, predictors were standardized during model training to place them on a common scale, and models equipped to handle correlated inputs were used.

2 Section 3:

Wine quality was modeled as a three class classification problem, with labels defined as (low <=5), mid (6), and good (>=7). The dataset was split into training and testing sets using an 80/20 split, with stratification applied to preserve the original class proportions in the sets. Due to class imbalance in the training data, bootstrap resampling was used to balance the classes by oversampling the minority classes to match the size of the largest class.

An artificial neural network (ANN) classifier was implemented using a pipeline that combined feature standardization with a multilayer perceptron (MLP). The baseline ANN consisted of a single hidden layer with 50 neurons, ReLU activiation and the Adam optimizer, trained for up to 500 iterations. Standardization was applied to all predictors to account for differences in scale and skewed feature distributions that were identified during exploratory analysis.

In addition to the baseline model, a tuned ANN was developed using grid search with a 5 fold cross validation on the balanced training set. Hyperparameters explored included the number of hidden units, network depth, regulatization strength, and learning rate. Model selection was based on macro averaged F1 score to account for class imbalance.

3 Section 4:

The performance of the Artificial Neural Network (ANN) was evaluated using metrics that account for the three class imbalance, such as class wise precision, recall, and F1-score, along with balanced accuracy, Cohen's kappa, and Matthews correlation coefficient (MCC). Model performance was assessed using predictions on a test set and results were summarized using normalized confusion matrices and classification reports.

Table 1: Performance comparison of Baseline and Tuned ANN models

	Baseline ANN			Tuned ANN		
Class	Precision	Recall	F1-score	Precision	Recall	F1-score
Low ($\leq$5)	0.742	0.792	0.766	0.786	0.738	0.761
Mid (=6)	0.653	0.484	0.556	0.630	0.586	0.607
High ($\geq$7)	0.500	0.767	0.606	0.525	0.744	0.615
Macro Avg	0.632	0.681	0.643	0.647	0.689	0.661
Weighted Avg	0.674	0.666	0.661	0.688	0.678	0.680
Accuracy		0.666			0.678	
Balanced Acc.		0.681			0.689	
Cohen's Kappa		0.463			0.482	
MCC		0.469			0.484	

As shown in Table 1, the baseline ANN obtained an overall accuracy of 0.66 and a balanced accuracy of 0.681. Performance varied across classes, with the strongest results observed for the low-quality class (recall = 0.79) and the weakest for the mid-quality class (recall = 0.48). The confusion matrix (Figure 5) shows that missclassification occurred between adjacent quality levels the most. Low-quality wines were often misclassified as mid-quality, while high quality wines were sometimes predicted as mid-quality. This pattern suggests that ANN struggled the most with seperating the adjacent classes.
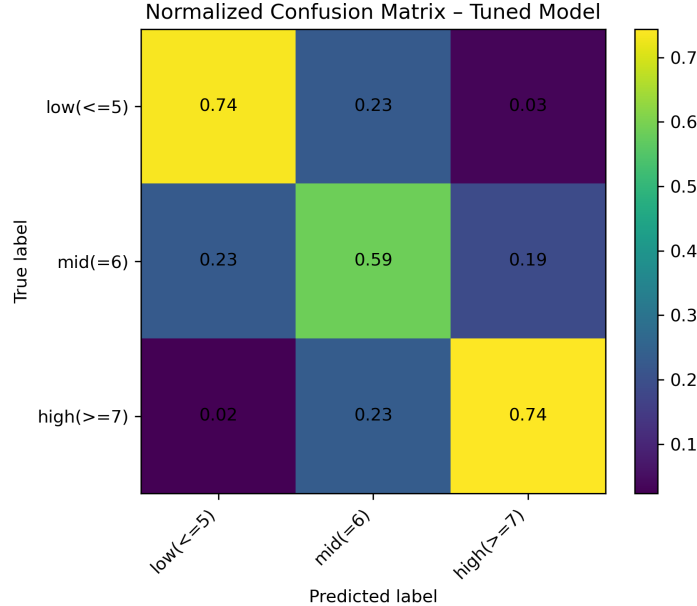

Figure 5: Tuned ANN Confusion Matrix

After hyperparameter tuning, the ANN demonstrated modest, but consistent improvements. The tuned model achieved an accuracy of 0.678 and a balanced accuracy of 0.689, with gains in macro averages precision, recall, and F1-score. Recall for the mid-quality class improved notable from 0.48 to 0.59, indicating better

identification of the central class, while maintaining performance for the low and high quality classes. The tuned model continued to show most misclassification occuring between adjacent quality levels, particularly between mid and high quality.

Overall, the tuned ANN demonstrated improved class balance and agreement beyond chance, as reflected by higher Cohen's kappa and MCC values. These results indicate that the ANN is effective at modeling nonlinear relationshups in the wine quality data, though class overlap remains a challenge.

4   Section 5: