Group- Arnay and Jonah

Project title- Predicting Wine Quality Using Supervised Machine Learning Models

1) Background and Project Motivation

Wine makers rely on decades-old tasting panels to grade product quality, yet modern data sets show that basic physicochemical measurements (acidity, sugar, sulphates, alcohol, etc.) already contain a great deal of information about perceived flavour. Our goal was to build supervised learning models that translate those eleven measurements into the same 0–10 quality scores assigned by professional tasters. Doing so can shorten R & D cycles, flag faulty batches early and, on the consumer side, give recommendation engines a low-cost proxy for flavour.

We used the UCI Wine-Quality data set: 1 599 red-wine samples and 4 898 white-wine samples, each described by the same 11 numeric predictors and a target label quality. Because quality is an ordinal integer, we treated the task as multi-class classification and compared seven models that include:

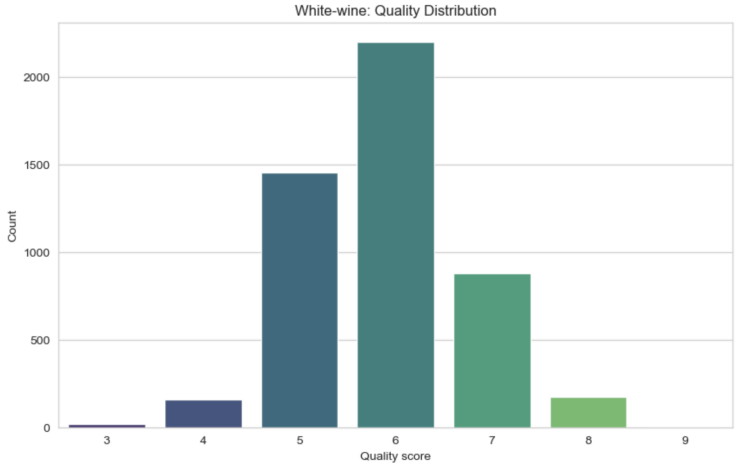
Linear Regression (rounded)
Logistic Regression
Support-Vector Machine (SVM)
k-Nearest-Neighbours (k-NN)
Gaussian Naïve Bayes
Decision Tree
Random Forest

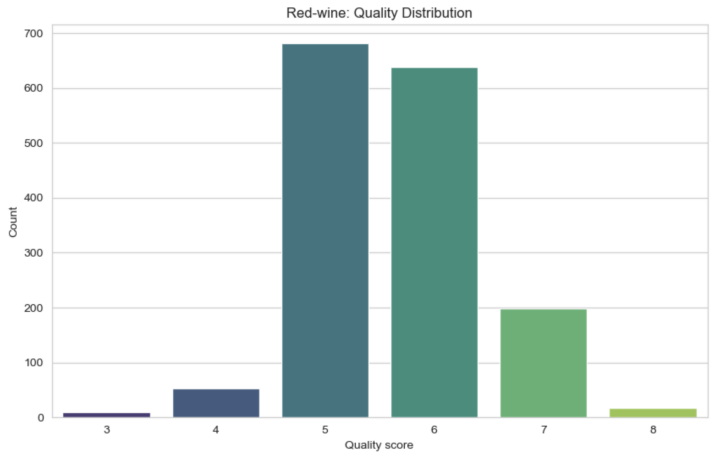
Hyper-parameter tuning was carried out with grid search and five-fold cross-validation, and model performance was evaluated on an 80/20 split, using accuracy, precision, recall and F1-score.

2) Exploratory Data Analysis

Label distribution.

Quality scores are strongly centre-weighted. For reds: 5 (681 samples) and 6 (638) make up 82 % of the data; only 28 red wines score 8 or better. Whites show a similar pattern, scores 5 (1 457) and 6 (2 198) cover 75 % of the set, with just five samples at the top score of 9. The class imbalance foreshadows why F1-scores lag behind raw accuracy.

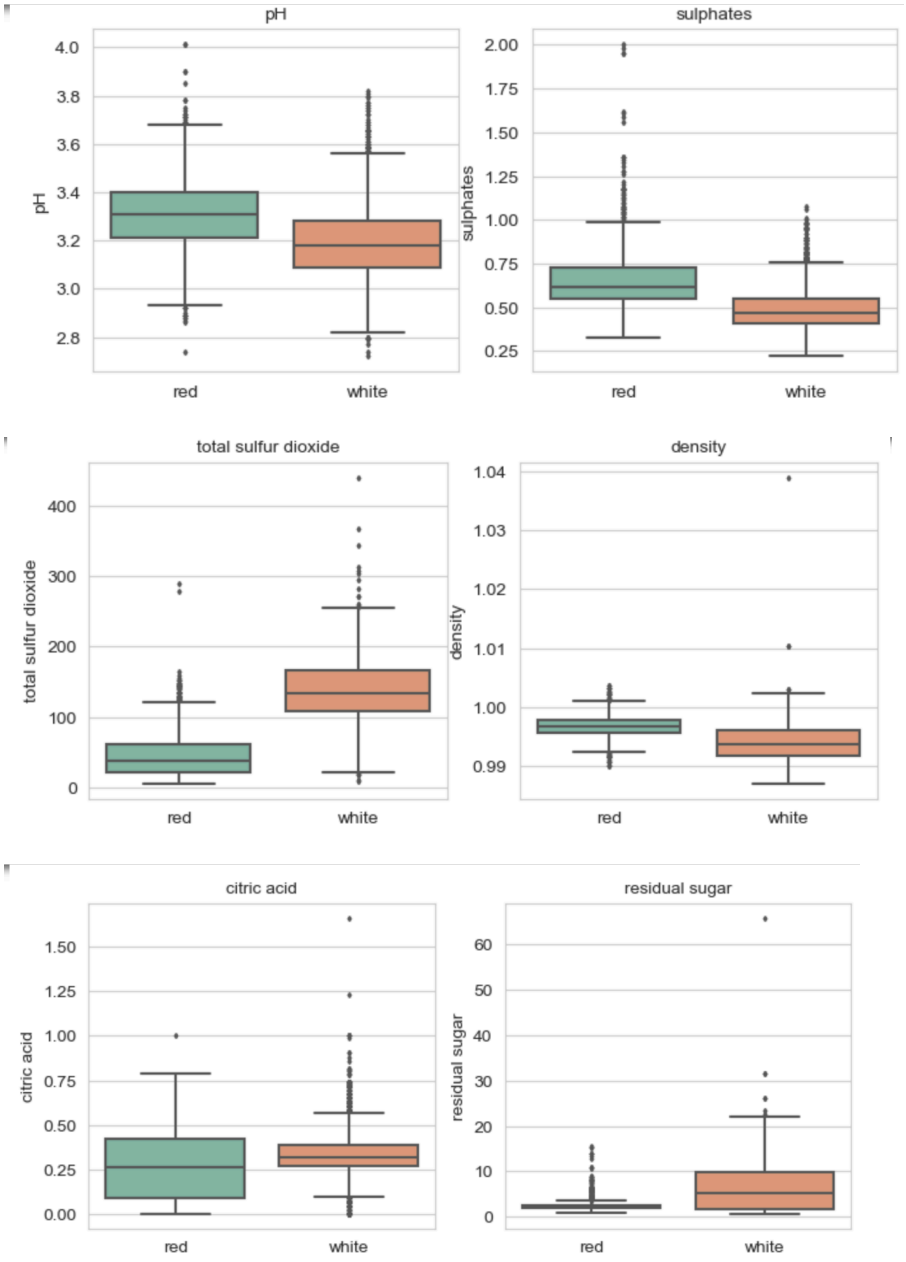


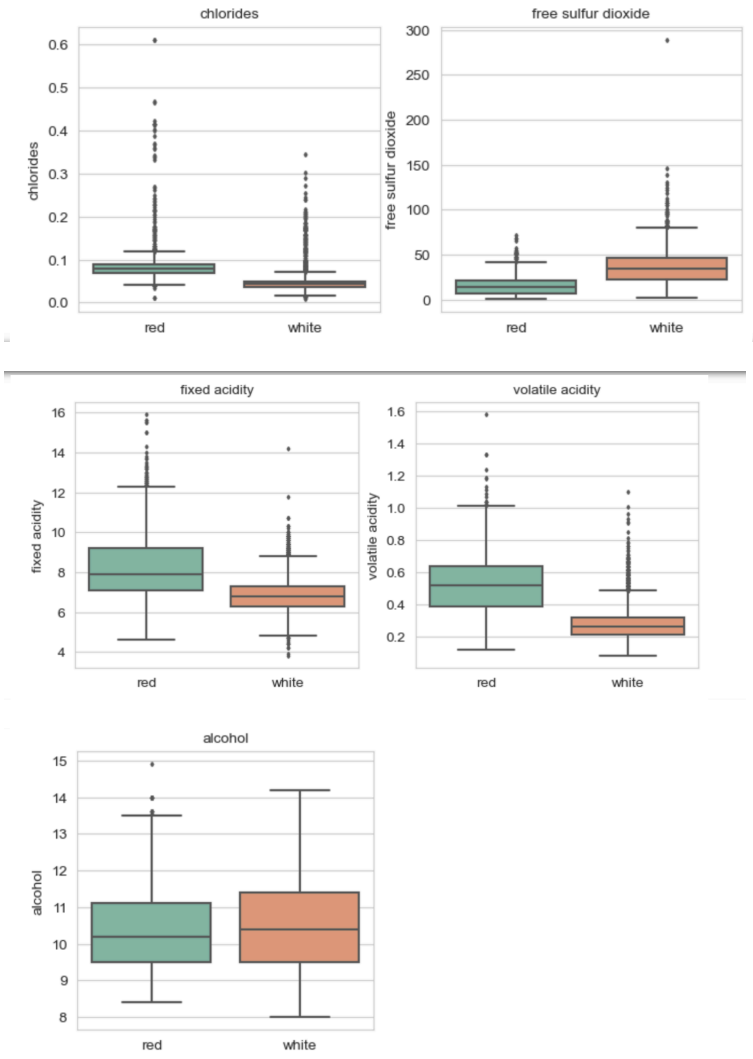


Feature ranges and outliers.

After constructing Boxplots of each predictor, they revealed a moderate skew in general. Red wines show higher medians for fixed and volatile acidity, chlorides, and sulphates, confirming their generally firmer structure and use of fuller‑bodied preservative practice. Whites carry

markedly more residual sugar and far larger ranges for free/total-sulphur dioxide, reflecting sweeter styles and heavier SO_2 protection against oxidation. Density tracks the sugar gap, while pH sits slightly lower (more acidic) in whites. Alcohol levels overlap but whites skew a touch higher in the upper quartile. Outliers are most extreme in white sugar and SO_2, whereas red wines display longer tails in acidity measures.





Correlation structure.

Alcohol (ρ = 0.48 for red, 0.44 for white) – higher alcohol content tends to earn higher quality scores. This confirms fuller-bodied wines tend to score higher with tasters.

Volatile acidity (|ρ| ≈ 0.39 red; smaller in whites) – excess volatile acidity degrades perceived quality.

In reds wines show a second tier of positive links in sulphates (+0.25) and citric acid (+0.23).

In whites, alcohol remains key, but the most pronounced negative correlations shift to density (–0.31) and chlorides (–0.21), signalling that leaner, less saline profiles are favoured.

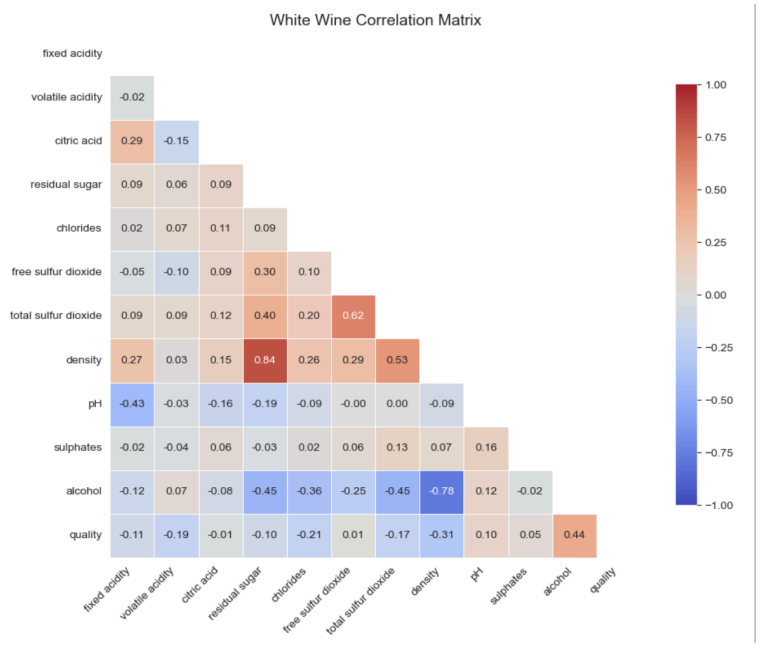
Among the predictors themselves, density is strongly inversely correlated with alcohol in whites (ρ ≈ –0.78).

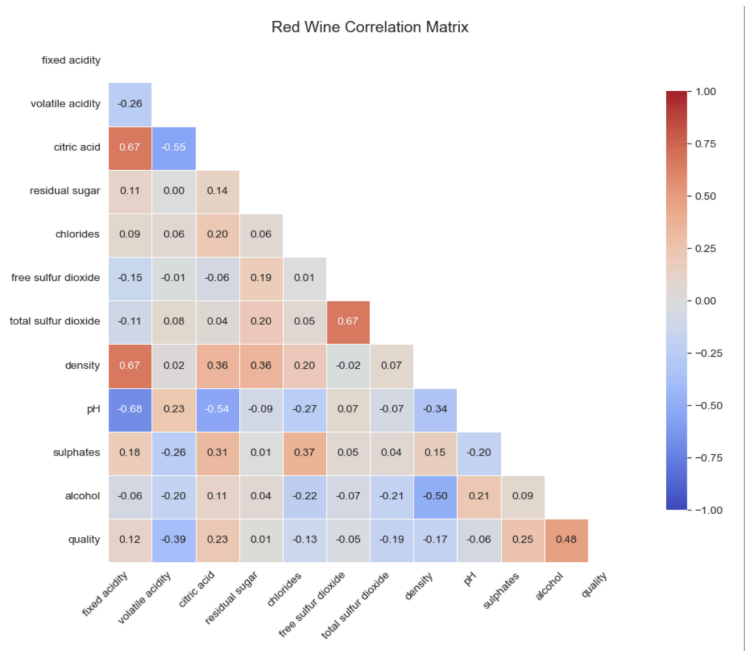
3) Data Pre-processing and Experimental Setup

Train–test split: 80 % training, 20 % test to preserve rare high-quality classes.

Scaling: StandardScaler fitted on training data

Feature selection: we inspected correlation heat-maps and confirmed that dropping any single variable reduced validation accuracy, so the full feature set was retained for parity across models.





Hyper-parameter grids:

SVM: kernels {linear, RBF, poly. Best for both colours was RBF.

k-NN: neighbours ∈ {1, 3, 5, 7, 9, 11} – the data favoured k = 1, consistent with tight clustering around the modal scores.

Decision Tree: max_depth ∈ {2, 4, 6, 8, 10}.

Random Forest: n_estimators ∈ {10, 50, 100, 200}. Red wines preferred 200 trees; whites 50 provided the best bias-variance trade-off.

4) Modelling Results (test set)

White Wine results:

	best_params	accuracy	precision	recall	f1_score
Linear Regression	N/A	0.515314	0.257201	0.181262	0.182414
Logistic Regression	N/A	0.537774	0.297414	0.222944	0.226083
SVM	{'kernel': 'rbf'}	0.555641	0.325271	0.225065	0.225045
K-Nearest Neighbors	{'n_neighbors': 1}	0.599541	0.37184	0.361353	0.366077
Naive Bayes	{'var_smoothing': 1e-09}	0.440531	0.236451	0.254455	0.233395
Decision Tree	{'max_depth': 10}	0.541092	0.31228	0.283098	0.292852
Random Forest	{'n_estimators': 50}	0.653395	0.504211	0.35307	0.389669

Red Wine results:

	best_params	accuracy	precision	recall	f1_score
Linear Regression	N/A	0.587867	0.329645	0.261271	0.262729
Logistic Regression	N/A	0.594121	0.492067	0.290727	0.307154
SVM	{'kernel': 'rbf'}	0.62414	0.314563	0.281309	0.282808
K-Nearest Neighbors	{'n_neighbors': 1}	0.649156	0.420479	0.39519	0.403832
Naive Bayes	{'var_smoothing': 1e-09}	0.562226	0.350981	0.343038	0.342809
Decision Tree	{'max_depth': 8}	0.590994	0.339511	0.303631	0.312994
Random Forest	{'n_estimators': 200}	0.693558	0.454847	0.355783	0.369635

5) Discussion

Tree ensembles win on accuracy. The Random Forest's ability to model nonlinear predictor interactions yields the highest accuracy for both wine types (~ 0.69 red, 0.65 white). However its F1-score trails k-NN on the red set because the forest is slightly more conservative in predicting minority high-quality classes.

k-NN excels at minority recall. With only one neighbour, k-NN memorises local structure; it therefore retrieves rare classes better, lifting the F1-score to 0.40 (red) and 0.37 (white). The trade-off is a small drop in overall accuracy.

Linear approaches under-fit. Both linear and logistic regression plateau near 0.59/0.54 accuracy: the relationship between physicochemical inputs and sensory quality is clearly non-linear.

SVM is competitive but sensitive. RBF kernels improve on logistic regression yet remain below tree models.

Class imbalance matters. Accuracy inflates performance because “5” and “6” dominate the label set. F1-score, being the harmonic mean of precision and recall, better penalises mis-classification of scarce high- and low-quality wines, which is why every model’s F1 is ≤ 0.40.

6) Conclusion and Future Work

These basic chemical values alone allow us to predict expert wine ratings with ~ 65–70 % accuracy. A couple possible next steps could include:

- Ordinal regression or cost-sensitive learning to exploit the ordered nature of the target and address imbalance without blunt rounding.
- Utilize Gradient-boosted trees or XGBoost, which often surpass random forests on tabular data.
- Feature engineering: ratios such as free/total sulphur, interaction terms, or log-scaling of skewed variables could expose additional structure.
- Domain-specific validation using sensory panels to judge prediction error impact; mis-labelling a “9” as a “7” may be more costly than confusing “5” and “6”.

With these extensions the pipeline could move from an academic exercise to a deployable tool for objective quality assessment.