From Lab Tests to Quality Scores

From Lab Tests to Quality Scores: A Neural Network Model for Red Wine Quality at GNAWC

Seif H. Kungulio

January 17, 2026

Contents

Business Understanding

Background

Great North American Wine Company (GNAWC) collects physicochemical lab measurements (e.g., acidity, sugar, alcohol, sulphates, pH) during production. Final quality ratings, however, are primarily determined by sensory evaluation, which can be subjective and costly.

Problem Statement

Build a neural network model that predicts wine quality scores using physicochemical tests only, enabling earlier, standardized, data-driven screening before bottling/distribution.

Objectives

- Predict quality scores accurately from lab measurements
- Detect potential low-quality batches earlier
- Reduce routine reliance on tasting panels for screening
- Improve quality decision consistency

Success Criteria

- **Primary:** Low prediction error (MAE/RMSE) on a held-out test set
- **Operational:** Strong detection of low-quality wines (e.g., quality 5) via recall/ROC-AUC (optional classification view)
- **Business:** Practical thresholding rules for early screening (flagging likely low-quality batches)

Data Understanding

Load Data

```
wine_data <- read_csv(data_path, col_names = TRUE, show_col_types = FALSE)
glimpse(wine_data)
```

```
## Rows: 1,599
## Columns: 1
## $ `fixed acidity;volatile acidity;citric acid;residual sugar;chlorides;free sulfur dioxide;total sul
```

```
head(wine_data, 10)
```

```
## # A tibble: 10 x 1
##    fixed acidity;volatile acidity;citric acid;residual sugar;chlorides;free su~1
##    <chr>
##  1 7.4;0.7;0;1.9;0.076;11;34;0.9978;3.51;0.56;9.4;5
##  2 7.8;0.88;0;2.6;0.098;25;67;0.9968;3.2;0.68;9.8;5
##  3 7.8;0.76;0.04;2.3;0.092;15;54;0.997;3.26;0.65;9.8;5
##  4 11.2;0.28;0.56;1.9;0.075;17;60;0.998;3.16;0.58;9.8;6
##  5 7.4;0.7;0;1.9;0.076;11;34;0.9978;3.51;0.56;9.4;5
##  6 7.4;0.66;0;1.8;0.075;13;40;0.9978;3.51;0.56;9.4;5
##  7 7.9;0.6;0.06;1.6;0.069;15;59;0.9964;3.3;0.46;9.4;5
##  8 7.3;0.65;0;1.2;0.065;15;21;0.9946;3.39;0.47;10;7
##  9 7.8;0.58;0.02;2;0.073;9;18;0.9968;3.36;0.57;9.5;7
## 10 7.5;0.5;0.36;6.1;0.071;17;102;0.9978;3.35;0.8;10.5;5
## # i abbreviated name:
## #   1: `fixed acidity;volatile acidity;citric acid;residual sugar;chlorides;free sulfur dioxide;total
```

Target Variable Check

```
wine_data %>%
  count(wine_data$quality) %>%
  arrange(wine_data$quality) %>%
  kable()
```

```
## Warning: There was 1 warning in `count()`.
## i In argument: `wine_data$quality`.
## Caused by warning:
## ! Unknown or uninitialised column: `quality`.
```

```
## Warning: There was 1 warning in `arrange()`.
## i In argument: `..1 = wine_data$quality`.
## Caused by warning:
## ! Unknown or uninitialised column: `quality`.
```

n
1599