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 **1**

Background . . . 1

Problem Statement . . . 1

Objectives . . . 1

Success Criteria . . . 1

Data Understanding **2**

Load Data . . . 2

Initial Data Exploration . . . 2

Data Summary . . . 3

Missing Values . . . 4

. . . 4

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.df <- read.csv("resources/wine.csv")
```

Initial Data Exploration

Display the first 10 rows of the dataset to understand its structure and contents.

```
head(wine.df, 10)
```

```
##    fixed.acidity volatile.acidity citric.acid residual.sugar chlorides
## 1            7.4             0.70        0.00            1.9     0.076
## 2            7.8             0.88        0.00            2.6     0.098
## 3            7.8             0.76        0.04            2.3     0.092
## 4           11.2             0.28        0.56            1.9     0.075
## 5            7.4             0.70        0.00            1.9     0.076
## 6            7.4             0.66        0.00            1.8     0.075
## 7            7.9             0.60        0.06            1.6     0.069
## 8            7.3             0.65        0.00            1.2     0.065
## 9            7.8             0.58        0.02            2.0     0.073
## 10           6.7             0.58        0.08            1.8     0.097
##    free.sulfur.dioxide total.sulfur.dioxide density   pH sulphates alcohol
## 1                   11                   34  0.9978 3.51      0.56     9.4
## 2                   25                   67  0.9968 3.20      0.68     9.8
## 3                   15                   54  0.9970 3.26      0.65     9.8
## 4                   17                   60  0.9980 3.16      0.58     9.8
## 5                   11                   34  0.9978 3.51      0.56     9.4
## 6                   13                   40  0.9978 3.51      0.56     9.4
## 7                   15                   59  0.9964 3.30      0.46     9.4
## 8                   15                   21  0.9946 3.39      0.47    10.0
## 9                    9                   18  0.9968 3.36      0.57     9.5
## 10                  15                   65  0.9959 3.28      0.54     9.2
##    quality Id
## 1        5  0
## 2        5  1
## 3        5  2
## 4        6  3
## 5        5  4
## 6        5  5
## 7        5  6
## 8        7  7
## 9        7  8
## 10       5 10
```

View the data structure

```
glimpse(wine.df)
```

```
## Rows: 1,143
## Columns: 13
## $ fixed.acidity        <dbl> 7.4, 7.8, 7.8, 11.2, 7.4, 7.4, 7.9, 7.3, 7.8, 6.7~
## $ volatile.acidity     <dbl> 0.700, 0.880, 0.760, 0.280, 0.700, 0.660, 0.600, ~
## $ citric.acid          <dbl> 0.00, 0.00, 0.04, 0.56, 0.00, 0.00, 0.06, 0.00, 0~
## $ residual.sugar       <dbl> 1.9, 2.6, 2.3, 1.9, 1.9, 1.8, 1.6, 1.2, 2.0, 1.8,~
## $ chlorides            <dbl> 0.076, 0.098, 0.092, 0.075, 0.076, 0.075, 0.069, ~
## $ free.sulfur.dioxide  <dbl> 11, 25, 15, 17, 11, 13, 15, 15, 9, 15, 16, 9, 35,~
## $ total.sulfur.dioxide <dbl> 34, 67, 54, 60, 34, 40, 59, 21, 18, 65, 59, 29, 1~
## $ density              <dbl> 0.9978, 0.9968, 0.9970, 0.9980, 0.9978, 0.9978, 0~
## $ pH                   <dbl> 3.51, 3.20, 3.26, 3.16, 3.51, 3.51, 3.30, 3.39, 3~
## $ sulphates            <dbl> 0.56, 0.68, 0.65, 0.58, 0.56, 0.56, 0.46, 0.47, 0~
## $ alcohol              <dbl> 9.4, 9.8, 9.8, 9.8, 9.4, 9.4, 9.4, 10.0, 9.5, 9.2~
## $ quality              <int> 5, 5, 5, 6, 5, 5, 5, 7, 7, 5, 5, 5, 7, 6, 5, 5, 5~
## $ Id                   <int> 0, 1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 13, 16, 19, 21~
```

Data Summary

Provide a summary of the dataset to get an overview of the variables and their distributions.

```
summary(wine.df)
```

```
##  fixed.acidity    volatile.acidity  citric.acid     residual.sugar
##  Min.   : 4.600   Min.   :0.1200   Min.   :0.0000   Min.   : 0.900
##  1st Qu.: 7.100   1st Qu.:0.3925   1st Qu.:0.0900   1st Qu.: 1.900
##  Median : 7.900   Median :0.5200   Median :0.2500   Median : 2.200
##  Mean   : 8.311   Mean   :0.5313   Mean   :0.2684   Mean   : 2.532
##  3rd Qu.: 9.100   3rd Qu.:0.6400   3rd Qu.:0.4200   3rd Qu.: 2.600
##  Max.   :15.900   Max.   :1.5800   Max.   :1.0000   Max.   :15.500
##    chlorides       free.sulfur.dioxide total.sulfur.dioxide    density
##  Min.   :0.01200   Min.   : 1.00       Min.   :  6.00       Min.   :0.9901
##  1st Qu.:0.07000   1st Qu.: 7.00       1st Qu.: 21.00       1st Qu.:0.9956
##  Median :0.07900   Median :13.00       Median : 37.00       Median :0.9967
##  Mean   :0.08693   Mean   :15.62       Mean   : 45.91       Mean   :0.9967
##  3rd Qu.:0.09000   3rd Qu.:21.00       3rd Qu.: 61.00       3rd Qu.:0.9978
##  Max.   :0.61100   Max.   :68.00       Max.   :289.00       Max.   :1.0037
##        pH          sulphates         alcohol         quality
##  Min.   :2.740   Min.   :0.3300   Min.   : 8.40   Min.   :3.000
##  1st Qu.:3.205   1st Qu.:0.5500   1st Qu.: 9.50   1st Qu.:5.000
##  Median :3.310   Median :0.6200   Median :10.20   Median :6.000
##  Mean   :3.311   Mean   :0.6577   Mean   :10.44   Mean   :5.657
##  3rd Qu.:3.400   3rd Qu.:0.7300   3rd Qu.:11.10   3rd Qu.:6.000
##  Max.   :4.010   Max.   :2.0000   Max.   :14.90   Max.   :8.000
##        Id
##  Min.   :   0
##  1st Qu.: 411
##  Median : 794
##  Mean   : 805
##  3rd Qu.:1210
##  Max.   :1597
```

Missing Values

Check for missing values in the dataset.

```
colSums(is.na(wine.df))
```

```
##        fixed.acidity     volatile.acidity          citric.acid
##                    0                    0                    0
##       residual.sugar            chlorides  free.sulfur.dioxide
##                    0                    0                    0
## total.sulfur.dioxide              density                   pH
##                    0                    0                    0
##            sulphates              alcohol              quality
##                    0                    0                    0
##                   Id
##                    0
```