IntelliML Report

Sample Datset

fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
7.4	0.7	0.0	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5
7.8	0.88	0.0	2.6	0.098	25.0	67.0	0.9968	3.2	0.68	9.8	5
7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.997	3.26	0.65	9.8	5
11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.998	3.16	0.58	9.8	6
7.4	0.7	0.0	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5

Feature Description

The dataset contains 11 features of wine.

Fixed acidity measures the amount of tartaric acid in wine. It is expressed in grams per liter.

Volatile acidity measures the amount of acetic acid in wine. It is expressed in grams per liter.

Citric acid is a type of acid found in citrus fruits. It is expressed in grams per liter.

Residual sugar is the amount of sugar that remains in wine after fermentation. It is expressed in grams per liter.

Chlorides are salts of hydrochloric acid. They are expressed in grams per liter.

Free sulfur dioxide is a type of antioxidant that is added to wine to protect it from spoilage. It is expressed in milligrams per liter.

Total sulfur dioxide is the sum of free and bound sulfur dioxide. It is expressed in milligrams per liter.

Density is the ratio of the mass of a substance to its volume. It is expressed in grams per milliliter.

pH is a measure of the acidity or alkalinity of a solution. It is expressed on a scale from 0 to 14, with 0 being the most acidic and 14 being the most alkaline.

Sulphates are salts of sulfuric acid. They are expressed in grams per liter.

Alcohol is the percentage of alcohol by volume in wine. It is expressed as a percentage.

Quality is a subjective measure of the overall quality of the wine. It is expressed on a scale from 0 to 10, with 0 being the worst and 10 being the best.

Insights on dataset

The dataset contains 1599 instances and 12 features. The features are: fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol, and quality. The mean, standard deviation, minimum, 25th percentile, 50th percentile, 75th percentile, and maximum of each feature are reported.

Fixed acidity has a mean of 8.319637, a standard deviation of 1.741096, a minimum of 4.600000, a 25th percentile of 7.100000, a 50th percentile of 7.900000, a 75th percentile of 9.200000, and a maximum of 15.900000.

Volatile acidity has a mean of 0.527821, a standard deviation of 0.179060, a minimum of 0.120000, a 25th percentile of 0.390000, a 50th percentile of 0.520000, a 75th percentile of 0.640000, and a maximum of 1.580000.

Citric acid has a mean of 0.270976, a standard deviation of 0.194801, a minimum of 0.000000, a 25th percentile of 0.090000, a 50th percentile of 0.260000, a 75th percentile of 0.420000, and a maximum of 1.000000.

Residual sugar has a mean of 2.538806, a standard deviation of 1.409928, a minimum of 0.900000, a 25th percentile of 1.900000, a 50th percentile of 2.200000, a 75th percentile of 2.600000, and a maximum of 15.500000.

Chlorides has a mean of 0.087467, a standard deviation of 0.047065, a minimum of 0.012000, a 25th percentile of 0.070000, a 50th percentile of 0.079000, a 75th percentile of 0.090000, and a maximum of 0.611000.

Free sulfur dioxide has a mean of 15.874922, a standard deviation of 10.460157, a minimum of 1.000000, a 25th percentile of 7.000000, a 50th percentile of 14.000000, a 75th percentile of 21.000000, and a maximum of 72.000000.

Total sulfur dioxide has a mean of 46.467792, a standard deviation of 32.895324, a minimum of 6.000000, a 25th percentile of 22.000000, a 50th percentile of 38.000000, a 75th percentile of 62.000000, and a maximum of 289.000000.

Density has a mean of 0.996747, a standard deviation of 0.001887, a minimum of 0.990070, a 25th percentile of 0.995600, a 50th percentile of 0.996750, a 75th percentile of 0.997835, and a maximum of 1.003690.

pH has a mean of 3.311113, a standard deviation of 0.154386, a minimum of 2.

Insights on Null Values in the dataset

The dataset does not contain any null values. This is a desirable property as it means that all of the data is available for analysis. However, it is important to note that this does not necessarily mean that the data is of high quality. For example, the data may be incomplete or inaccurate. It is important to carefully examine the data to ensure that it is fit for purpose before using it for analysis.

Feature Distribution

The distribution of each feature in the dataset is as follows:

fixed acidity: slightly right skewed
volatile acidity: slightly left skewed
citric acid: slightly left skewed
residual sugar: moderately right skewed
chlorides: moderately right skewed
free sulfur dioxide: slightly left skewed
total sulfur dioxide: slightly left skewed
density: slightly left skewed
pH: slightly left skewed
sulphates: moderately right skewed
alcohol: slightly right skewed
quality: slightly left skewed

The skewness of the data has several consequences. For example, the moderately right skewed distribution of residual sugar means that there are more wines with high residual sugar than wines with low residual sugar. This could make it difficult to compare wines with different residual sugar levels. Similarly, the slightly left skewed distribution of pH means that there are more wines with low pH than wines with high pH. This could make it difficult to compare wines with different pH levels.