Factors of Wine Quality

Authors:

Ethan Chase

Matthew McClure

Payton Parrish

Abstract

This report aims to analyze the various factors that impact wine quality using predictive models. Wine is a complex blend of multiple components, and to achieve quality, understanding the contributing factors is crucial. This study utilizes the Wine Quality dataset from UCI and investigates 11 input variables of white and red variations of the Portuguese “Vinho Verde” wine. The analysis will provide insights into the connection and importance between the input variables and wine quality, thereby helping to improve wine production processes.

Introduction

The production of high-quality wine involves numerous factors that vary significantly between batches and even individual bottles. Several elements, such as fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol, are crucial for wine quality. This research paper aims to investigate these factors and their impact on the quality of wine produced by a specific brand, Vinho Verde. Through the analysis of these parameters, we aim to gain insights into the characteristics that make a bottle of wine exceptional and to develop a better understanding of the wine production process.

Problem Statement

Wine quality depends on various factors, and identifying the crucial variables that contribute to it is crucial for the wine industry. This research paper aims to investigate the factors that impact wine quality and their relative importance. Additionally, we aim to analyze whether these factors differ between red and white wine. Through this research, we seek to contribute to the wine industry's understanding of the factors that impact wine quality and how these factors can be utilized to consistently produce high-quality wines. We also intend to develop a predictive model that can assess the quality of new wines added to the dataset.

Proposed Methodology

The proposed methodology for this research paper involves analyzing different regression and classification models to determine the most accurate model for predicting wine quality. The following methods will be utilized: random forest classification and Poisson regression since wine quality ranges from the discrete values of 1-10. We will also implement standard regression with rounding towards the nearest integer and other classification methods.

To identify the variables that contribute most to wine quality, we will use the forward stepping feature selection process. We will compare the features selected for each dataset to analyze the important variables between red and white wine. Finally, to determine the predictability of our model for new wines added to the dataset, we will analyze the accuracy of our models.

Variable Selection

Some of the questions we had for the dataset were related to which features of the data were the most important in predicting quality, and whether the important features were the same for red and white wines. We used SFS variable selection to investigate this. We found eight significant variables for the white wine data, and seven significant variables for the red wine data.

Out of all the variables, Citric Acid content was the only one that was not significant to either dataset. Chlorides and Total Sulfur Dioxide were significant to white wine but not red wine, and Fixed Acidity, Residual Sugar, and Density were significant to red wine but not white wine.

Regression Methods

We used several regression methods to try to predict the quality of the wine data based on input variables. The two most successful were Linear Regression on variables selected by SFS variable selection and Poisson Regression. Both methods of regression gave us a continuous number as output, so to analyze the predictions, we rounded the output to the nearest integer. The results of the regression techniques can be seen in the charts below.

|  |  |
| --- | --- |
|  |  |
|  |  |

The charts show that, when the results are rounded, normal regression methods perform slightly better than Poisson Regression on the data. On the other hand, when analyzing how close the continuous prediction was to the true value of each wine, we calculated the following Mean Squared Error values for the different models.

|  |  |
| --- | --- |
| White Wine SFS + Regression | 1.038 |
| Red Wine SFS + Regression | 0.446 |
| White Wine Poisson Regression | 0.949 |
| Red Wine Poisson Regression | 0.481 |

Looking at the values of the actual outputs, Poisson regression performed slightly better for the white wine data, and slightly worse for the red wine data.

Analysis and Results

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Conclusions

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lessons Learned

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Bibliography

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Appendix

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.