Predicting Wine Quality

A Machine Learning Approach by Peter Skotte

Overview

The wine market is extremely competitive, and a wine's quality is a key aspect in determining its success. To enhance their offerings, draw in clients, and boost sales, vineyard owners, winemakers, wine distributors, and merchants all have a stake in wine quality forecasting.

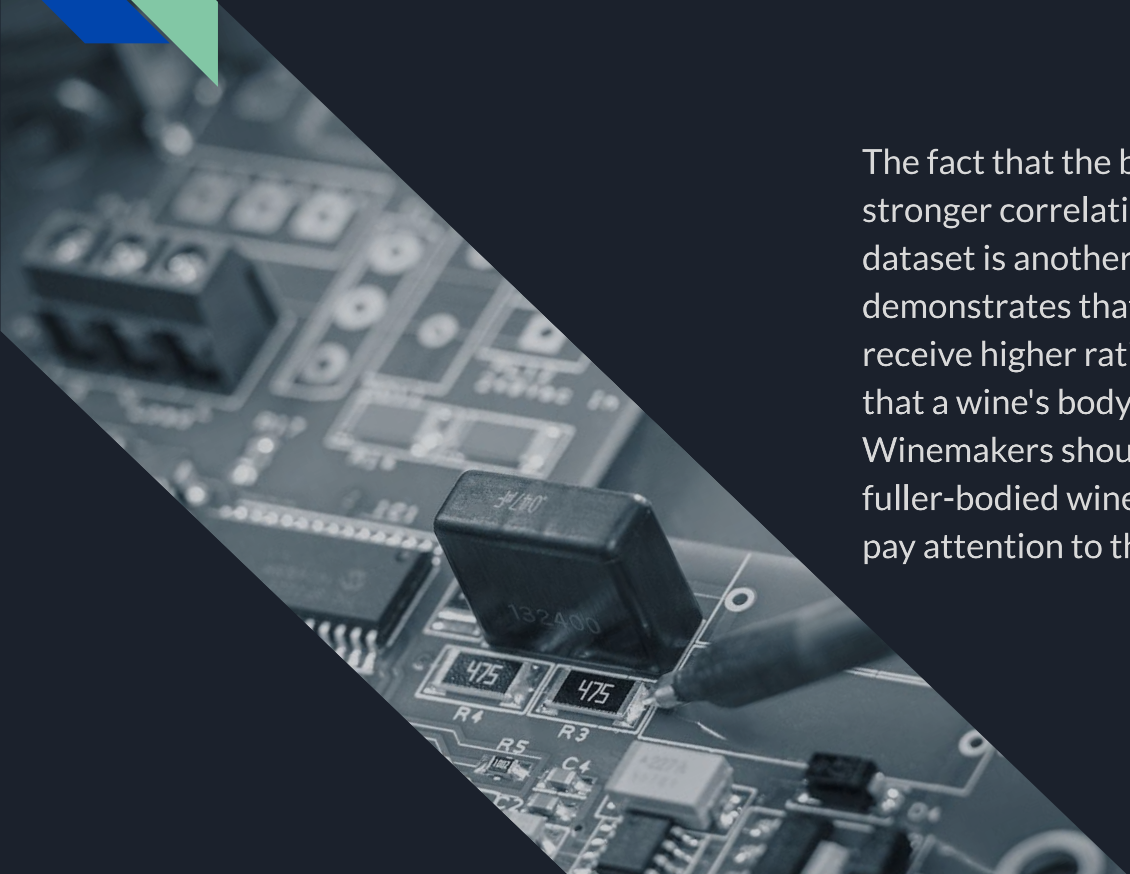
The data collection from Kaggle that we used for this experiment includes details about the red Spanish wine varieties. There are 11 features in all, including categorical and numerical variables. Our goal was to use a machine learning model to estimate the wine's rating based on these variables.

Key Findings - Price vs. Rating Correlation

01 The moderately positive link between the wine's rating and price (0.51) is one of our main findings. This scatter plot demonstrates that wines with greater pricing typically receive higher evaluations.

02 For stakeholders, the consequences of this study are crucial. First off, it shows that buyers may believe more expensive wines to be of greater quality. In order to balance profitability with perceived quality, winemakers and retailers should carefully analyze their pricing methods.

Rating vs. Body Correlation



The fact that the body of the wine and its rating have a stronger correlation (0.2) than other columns in the dataset is another important result. This scatter plot demonstrates that wines with a fuller body typically receive higher ratings. This discovery raises the possibility that a wine's body may affect its rating in some way. Winemakers should therefore consider making fuller-bodied wines to boost their ratings and sales and pay attention to the body of their wines.

Random Forest Model Evaluation

Our Random Forest model was able to account for 46% of the variance in wine ratings, as seen by its test MSE of 0.01 and R2 score of 0.46. This is a nice place to start when estimating wine quality. The model does, however, have significant drawbacks, such as the potential for overfitting and missing critical factors. To enhance the performance of the model, we suggest investigating additional feature engineering methods and different machine learning techniques, such as boosting or neural networks.

Regression Metrics and Implications

The mean squared error (MSE), which calculates the average squared difference between the predicted and actual values, is one of the metrics used to assess our regression model. Our Random Forest model has a test MSE of 0.01, demonstrating its ability to predict outcomes rather accurately. In light of this finding and the relationship between price and rating, it is likely that more expensive wines receive higher ratings. Stakeholders can improve the relationship between price and quality by carefully balancing pricing tactics with perceived quali

Recommendations and Conclusion

According to our data, stakeholders should concentrate on wine body because it positively correlates with wine rating. Higher ratings and greater sales may result from making wines with a fuller body. Furthermore, since more expensive wines typically receive higher ratings, stakeholders should carefully assess their pricing strategy to strike a balance between perceived quality and profitability. As a result, our analysis includes recommendations for enhancing wine quality and sales as well as insights into the elements affecting wine quality. We exhort stakeholders to use these data to make wise decisions and raise the caliber and profitability of their wine.