Untitled document

Camarines Sur Polytechnic Colleges

Document Details

Submission ID

trn:oid:::23367:87587719

Submission Date

Mar 24, 2025, 10:21 PM GMT+8

Download Date

Mar 24, 2025, 10:24 PM GMT+8

File Name

Untitled document

File Size

3.8 KB

3 Pages

547 Words

3,304 Characters

15% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Match Groups

8 Not Cited or Quoted 15%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

5%   🌐 Internet sources

8%   📖 Publications

12%   👤 Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

8   Not Cited or Quoted 15%
Matches with neither in-text citation nor quotation marks

0   Missing Quotations 0%
Matches that are still very similar to source material

0   Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0   Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

5%   🌐   Internet sources

8%   📖   Publications

12%   👤   Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1   Publication

R. N. V. Jagan Mohan, Vasamsetty Chandra Sekhar, V. M. N. S. S. V. K. R. Gupta. "Al...   **5%**

2   Submitted works

Colorado State University, Global Campus on 2021-02-04   **4%**

3   Submitted works

University College Dublin (UCD) on 2024-06-27   **2%**

4   Submitted works

University of Wollongong on 2024-02-13   **2%**

5   Submitted works

London Metropolitan University on 2024-12-06   **1%**

6   Internet

pkm.tunasbangsa.ac.id   **1%**

SEAN XANDER B. AQUINO BSCS 2A

Implementing Linear Regression to Wine Dataset

The various physicochemical properties in wine particularly acidity, sugar content, and alcohol concentration determine the acceptance of wine by consumers. Several interacting factors create a complex environment for wine quality estimation so machine learning techniques find this task interesting. Analysis of wine quality attributes uses linear regression on the Wine Quality Dataset found in the UCI Machine Learning Repository. This research deploys statistical modeling to determine how well the linear regression model performs at estimating wine quality assessment along with the main factors which drive predictions in the database.

The program draws its dataset from UCI Machine Learning Repository (ID: 186). The dataset consists of various physicochemical variables measured in wine samples with quality scores established by wine tasters acting as the target variable. The dataset contains these important characteristics:

- Fixed Acidity
- Volatile Acidity
- Citric Acid
- Residual Sugar
- Chlorides
- Free Sulfur Dioxide
- Total Sulfur Dioxide
- Density
- pH

- Sulphates

- Alcohol

- Quality (Target Variable)

First the dataset was loaded into the program for an inspection of missing data points before continuing with the analysis. Standardization procedures were applied to features before the data split created training and testing splits for accurate model assessment. Feature selection and data relationship assessment became possible through the generation of a correlation matrix heatmap which identified key influential features. The multiple linear regression model used the Scikit-learn library for its development. The information was divided into two groups where 80% served as training data and 20% served as testing data in order to conduct proper model assessment. The Linear Regression model received training from the available data before producing predictions for both training and testing periods. The model assessment metrics determined both the predictive accuracy and the ability to generalize its predictions.

The linear regression model evaluation used Mean Squared Error (MSE) along with R-squared (R^2) scores. The evaluation of model-data fit strength happened through calculations of training set R^2 along with testing set R^2:

- Training R^2 Score: 0.2992523560502254

- Testing R^2 Score: 0.2597673129771402

- Mean Squared Error: 0.546696441959444

Linear regression successfully demonstrated an ability to recognize broad patterns between actual and predicted quality scores but exhibited significant deviations in the

results. Wine quality seems to be impacted by non-linear patterns that the model probably failed to identify completely.

Both alcohol content and volatile acidity proved to be the main indicators for determining wine quality among the examined features. The linear model showed moderate prediction capability yet its residual discrepancy suggested other factors and potential nonlinear relationships exist in the data. Some features in the dataset showed weak correlations with quality based on the results from the correlation matrix which indicates possible data redundancy. More complex models including decision trees or neural networks should be applied for additional analysis to enhance prediction accuracy.

A linear regression model served to predict wine quality relying on the analysis of its physicochemical characteristics. The identified relationships demonstrate a satisfactory match between the model but performances indicate future models will achieve greater prediction accuracy. Further research should examine polynomial regression and ensemble methods and deep learning techniques because they could improve prediction power and capture the full range of wine quality assessment complexity.