

WINE QUALITY PREDICTION

FINAL EVALUATION PROJECT SUBMITTED BY:

MEHAK KANYAL(2210991916)

MEHAK BUDHIRAJA(2210991917)

LAVISHA GOEL(2210991852)

SUBMITTED TO: MR SACHIN GARG

TABLE OF CONTENT

- Introduction
- Problem Statement
- Objectives
- Applications
- Data Description
- Methodology
- Model Description
- Project Highlights
- Conclusion



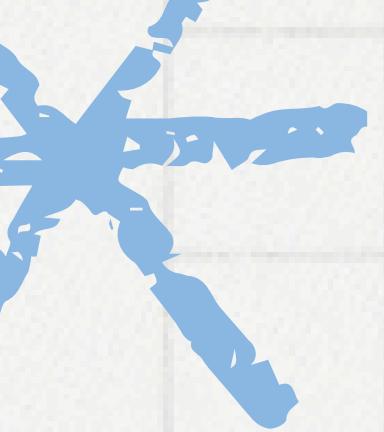
INTRODUCTION

- Wine is an alcoholic beverage made from fermented grapes or other fruits. It typically ranges in flavor, aroma, and complexity, influenced by factors such as grape variety, terroir, and winemaking techniques.
- The primary objective of this project is to build a predictive model capable of accurately estimating the quality of wines based on their chemical attributes. By leveraging machine learning techniques, we seek to identify the most influential factors affecting wine quality and create a tool that can assist winemakers in making informed decisions during the production process.



Problem Statement

In industries, understanding the demands of wine safety testing can be a complex task for the laboratory with numerous analytes and residues to monitor. But, our application's prediction, provide ideal solutions for the analysis of wine, which will make this whole process efficient and cheaper with less human interaction.



OBJECTIVES



01.

Our main objective is to predict the wine quality using machine learning through Python programming language.

02.

A large dataset is considered and wine quality is modelled to analyse the quality of wine through different parameters like fixed acidity, volatile acidity etc.

03.

All these parameters will be analysed through Machine Learning algorithms like random forest classifier algorithm which will help to rate the wine on scale 1 - 10 or bad - good. Output obtained would further be checked for correctness and model will be optimized accordingly.

APPLICATIONS

01

Wine Production Optimization:
By understanding which factors contribute most to wine quality, they can make informed decisions to optimize production efficiency and consistency.

02

Quality Control:
Early detection of quality deviations allows for timely interventions to maintain product standards and prevent spoilage, reducing the risk of costly recalls and customer dissatisfaction.

03

Wine Authentication and Fraud Detection:
Quality prediction models can help authenticate wines and detect counterfeit or fraudulent products by comparing their predicted quality profiles with known standards or historical data.

04

Consumer Recommendation Systems:
By matching consumers with wines predicted to align with their taste preferences and quality expectations, businesses can enhance the customer experience and foster brand loyalty.

DATA DESCRIPTION

- The dataset contains chemical descriptions of 6499 Portuguese "Vinho Verde" wines.
- There are 4899 entries for white wine, 1600 entries for red wines.
- The source of the data is taken from kaggle.com.

Attributes	Description
pH	To measure ripeness
Density	Density in gram per cm3
Alcohol	Volume of alcohol in %
Fixed Acidity	Impart sourness and resist microbial infection, measured in no. of grams of tartaric acid per dm3
Volatile Acidity	no. of grams of acetic acid per dm3 of wine
Citric Acid	no. of grams of citric acid per dm3 of wine
Residual Sugar	Remaining sugar after fermentation stops
Chlorides	no. of grams of sodium chloride per dm3 of wine
Free Sulfur dioxide	no. of grams of free sulphites per dm3 of wine
Total Sulfur dioxide	no. of grams of total sulfite (free sulphite+ bound)
Sulphates	no. of grams of potassium sulphate per dm3 of wine

METHODOLOGY

- It gives insights of the dependency of target variables on independent variables using machine learning techniques to determine the quality of wine because it gives the best outcome for the assurance of quality of wine
- The dependent variable is "quality rating" whereas other variables i.e. alcohol, sulphur etc. are assumed to be predictors or independent variables.
- While hindering the effectiveness of the data model, various types of errors have occurred like over fitting, introduced from having too large of a training set and bias occur due to too small of a test set.

MODELS DESCRIPTION



1. LOGISTIC REGRESSION:

1. What is Logistic Regression?

- Predicts probability of binary outcomes.
- Used for classification, not regression.

2. Training and Evaluation

- Trained with labeled data to minimize prediction errors.
- Evaluated using metrics like accuracy, precision, recall, and F1 score.

3. Applications

- Widely used in healthcare, finance, marketing, etc.
- Efficient, interpretable, and handles high-dimensional data well.

4. Conclusion

- Logistic Regression: Essential for binary classification tasks.
- Offers interpretable results and broad applicability.

MODELS DESCRIPTION



2. Support Vector Machines (SVM):

1. What is SVM?

- SVM is a powerful supervised learning algorithm used for classification and regression tasks.
- It finds the optimal hyperplane that best separates different classes in the feature space.

2. Training and Evaluation:

- Trained by finding the hyperplane that maximizes the margin while minimizing classification errors.
- Evaluated using metrics such as accuracy, precision, recall, and F1 score.

3. Applications

- Widely used in text classification, image recognition, bioinformatics, and financial forecasting.
- Effective for both linearly separable and non-linearly separable datasets.

4. Conclusion

- SVM: Versatile algorithm for classification and regression tasks.
- Offers robust performance and flexibility in handling complex data distributions.

MODELS DESCRIPTION



3. Random Forest Classifier:

1. What is Random Forest?

- Random Forest is an ensemble learning method used for classification and regression tasks.
- It consists of multiple decision trees trained on random subsets of the data and features.

2. Training and Evaluation

- Trained by growing a large number of decision trees and aggregating their predictions.
- Evaluated using metrics such as accuracy, precision, recall, and F1 score.

3. Applications

- Widely used in areas such as finance, healthcare, and marketing for tasks like fraud detection, disease diagnosis, and customer segmentation.
- Effective for handling high-dimensional data and capturing complex relationships between features.

4. Conclusion

- Random Forest: Powerful ensemble learning technique for classification and regression tasks.
- Offers robust performance, scalability, and resistance to overfitting.

PROJECT HIGHLIGHTS

Welcome to Wine Quality Predictor

Enter the fixed acidity:

4.6 to 16

Enter the volatile acidity :

0.12 to 1.60

Enter the citric acid:

0.00 to 1

Enter the residual sugar:

0.9 to 15.6

Enter the chlorides:

0.01 to 1

Enter the free sulfur dioxide :

1 to 75

Enter the total sulfur dioxide:

5 to 300

Enter the density:

Enter the chlorides:

0.01 to 1

Enter the free sulfur dioxide :

1 to 75

Enter the total sulfur dioxide:

5 to 300

Enter the density:

0.99 to 1.5

Enter the ph:

2.72 to 4.5

Enter the sulphates:

0.33 to 2

Enter the alcohol:

8 to 15

Predict

CONCLUSIONS

- The two most important features among all 12 attributes are Sulphur dioxide (both free and total) and Alcohol.
- Volatile acidity contributes to acidic tastes and have negative correlation to wine quality.
- Most important factor to decide the quality of wine is alcohol, higher concentration of alcohol leads to better quality of wine and lower density of wine.



**Thank you
very much!**