Machine Intelligence

Wine Quality Prediction

Anubuthi K, Attili Subha Vidisha, Devang Saraogi

PES University

Agenda

Introduction

Problem Statement

Dataset

Application & Uses

Scope & Novelty

Literature Survey

Proposed Approach

Comparative Analysis

Results & Discussion

Thank You

Introduction

Wine has become one of the most sought out beverages in the alcohol industry. Hence, winemakers are interested in procedures that can increase the taste, aroma and quality of wine produced. However, wine quality is not solely based on the alcohol content in its composition. It also depends on other factors such as citric acid content, residual chloride content, etc.

This study can help winemakers increase profits, by shedding light on the attributes that have a more significant impact on taste compared to other.

Problem Statement

Wine is becoming an increasingly prevalent and consumed drink. In the wine business, quality and taste makes a considerable difference. To produce high quality wine which tastes good, winemakers need to know which features influence the final product the most.

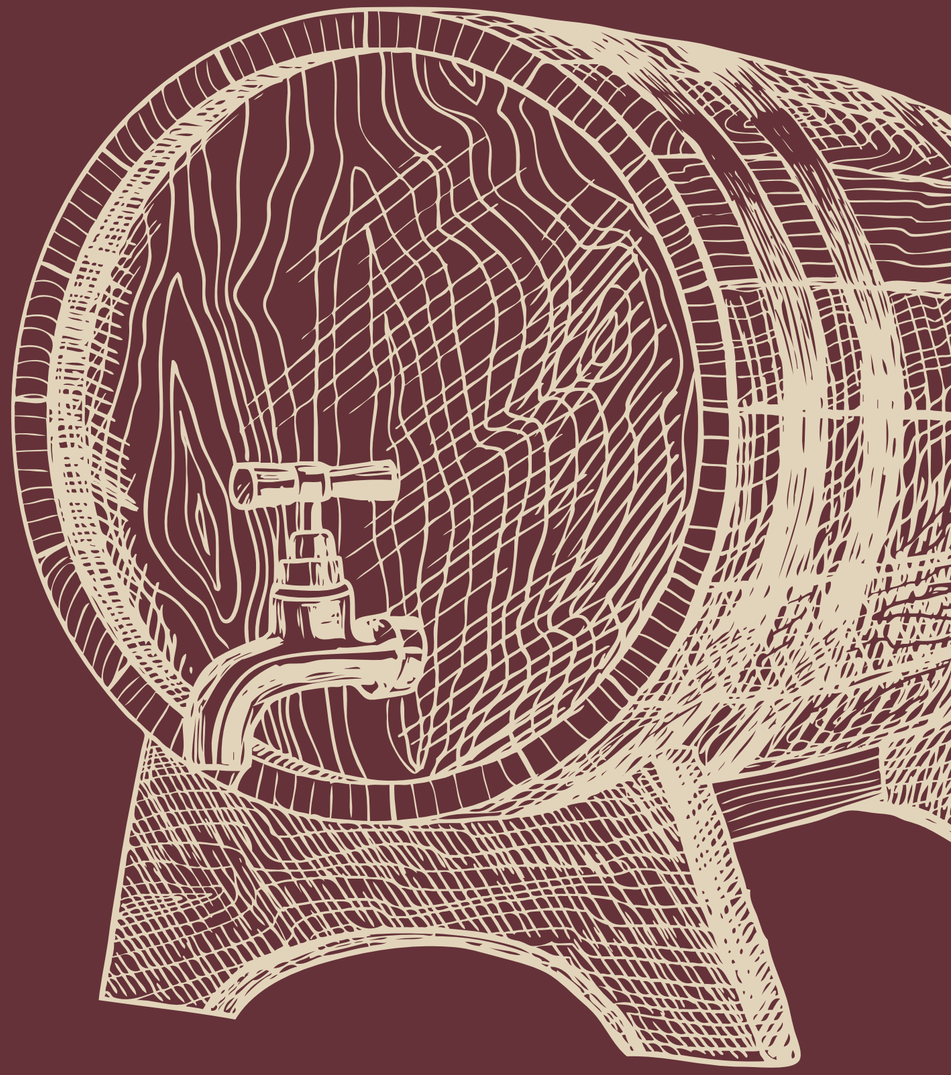
To solve this problem, we aim to build a model which highlights the features which influence the quality and also predict the quality of wine based on input features, therefore making it easy fpr the winemakers to assess how the quality of the final product would be.

Dataset

We have chosen the dataset form Kaggle. The dataset contains data about the various physiochemical factors such as alcohol level, various acidity levels etc. The quality of wine is based on these various factors and will also serve the purpose of our target variable.

Features of the Dataset

- total sulfur dioxide
- density
- pH
- sulphates
- alcohol
- quality
- fixed acidity
- volatile acidity
- citric acid
- residual sugar
- chlorides
- free sulfur dioxide



Objectives

01

Chalk out the most Influential Factors

There are many physiochemical factors which effect the quality of wine, we plan on highlighting the most influential factor.

02

Build a Model from Scratch

Implement a neural network from scratch to predict the quality of the data

03

Comparative Analysis

Comparative Analysis of pre-made models with our model (implemented from scratch).

Applications & Uses

Application

Our project can find its applications in the winemaking industry and in retail.

Uses

Helps winemakers know which physio-chemical factors influence the quality of wine made while also helping them maximise their profits.

Scope & Novelty

Scope

- Prediction of Wine Quality by using Artificial Neural Networks.
- Comparison of accuracies obtained after prediction from various in built predictors.
- Identification of features that are relatively more important than the others when it comes to quality of wine.

Novelty

Prediction of Wine Quality by using artificial neural networks built from inbuilt modules such as NumPy and Pandas and without using Keras and Tensorflow.

Literature Survey

The literature survey was performed on 6 papers with most of them being published in the past few years. The researchers employed various techniques/models such as Linear Regression, Random Forest, SVMs, Naïve Bayes. All of the techniques were available as function calls on libraries such as *sklearn* and *tensorflow*. They also highlighted the advantages of cleaning and preprocessing of the dataset.

None of the papers discussed the implementation of a model from scratch.

Title	Year	Journel/Conference	Advantages	Limitations
Red Wine Quality Prediction Using Machine Learning Techniques	2020	IEEE	• Dataset chosen has similar attributes. • Comparison of different models and mentions approach of how to perform analysis	• Inbuilt python libraries are used to implement the models • Does not illustrate the implementation
Prediction of Wine Quality using Machine Learning	2021	Journal of Emerging Technologies and Innovating Research	• The paper provides theoretical explanation of each model. • The significance of data cleaning, representation and correlation matrix are explained in a concise and clear manner.	• Inbuilt python libraries are used to implement the models • Does not illustrate the implementation
Wine Quality Classification with Multilayer Perceptron	2018	International Journal of Internet, Broadcasting and Communication	• Implementation of Deep Neural Networks is shown in detail	• Dataset used is very small

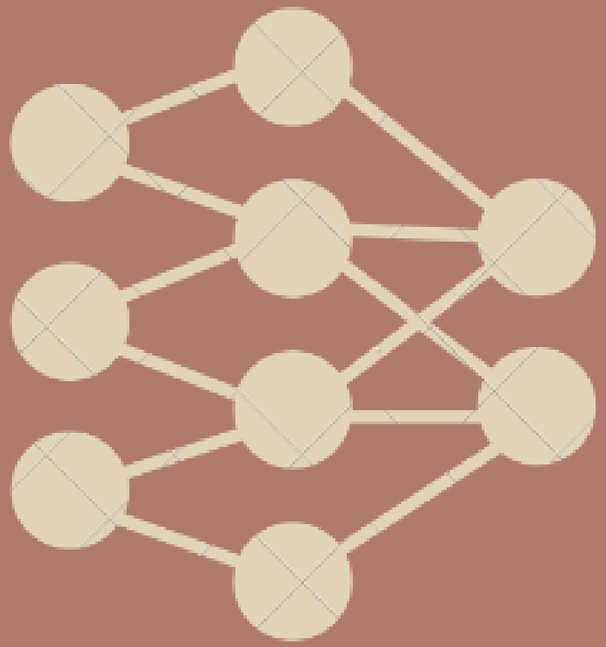
Title	Year	Journel/Conference	Advantages	Limitations
Selection of Important Features & Predicting Wine Quality using Machine Learning Techniques	2017	ICSCC	• Predicts qualities for Red Wine and White Wine using SVM and Neural Network regression. • Adequate mathematical proofs have been given behind the concepts of each model.	• Inbuilt python libraries are used to implement the models • Does not illustrate the implementation
Prediction of Wine Quality Using Machine Learning Algorithms (Different Paper)	2021	Conference: Open Journal of Statistics	• This paper provides a theoretical explanation before pursuing each model. • The significance of data cleaning, representation and correlation matrix are explained in a concise and clear manner. • Multiple models are used, giving us a wide spectrum of results. • A lot of performance metrics were used, giving a good view of the paper's methodology.	• Inbuilt python libraries are used to implement the models • Does not illustrate the implementation
Wine Quality and Taste Classification Using Machine Learning Model	2020	International Journal of Innovative Research in Applied Sciences and Engineering	• the have explored several ML techniques for evaluating wine quality based on different metrics and properties related to wine quality. • the have explored several machine learning techniques for evaluating wine quality based on different metrics and properties related to wine quality.	• dataset used is very small ,models used are not implemented from scratch

Title	Year	Journel/Conference	Advantages	Limitations
Prediction of Quality for Different Type of Wine based on Different Feature Sets Using Supervised Machine Learning Techniques	2018	ICACT Transactions on Advanced Communications Technology	• Used various performance techniques.	• Inbuilt python libraries are used to implement the models • Does not illustrate the implementation • Uses Genetic Algorithm
Red Wine Quality Prediction Using Machine Learning Techniques	2020	International Conference on Computer Communication and Informatics	• The steps involved in predicting wine are clearly illustrated. • SVM, Naive Bayes, Random Forest have been employed, shedding light on these methods.	• Accuracy of these models is below 70% which is relatively low when compared against other implementations of the same.
Wine Quality Prediction Using Data Mining	2019	IEEE	• The steps involved in predicting wine are clearly illustrated.	• Data Mining Techniques were used

Proposed Approach

Exploratory Data Analysis & Visualization

Model

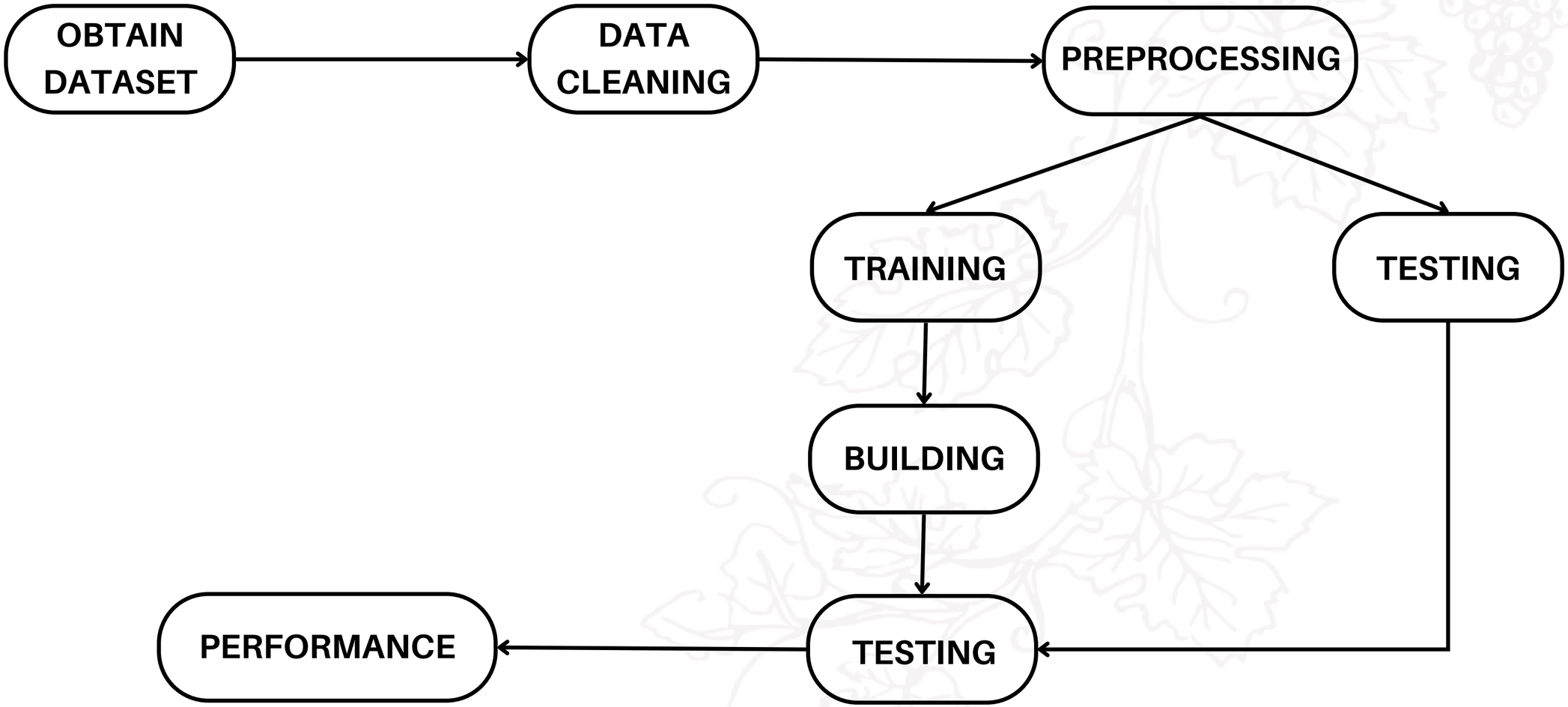


We are designing a neural network from scratch with the help of python libraries. A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates.

Neural networks can adapt to changing input; so the network generates the best possible result without needing to redesign the output criteria. Looking at the bigger picture, there are three main components, the input, Output and the Hidden Layers.

First, the input is the data entered into the network that is to be analyzed. Second, the processing layer utilizes the data (and prior knowledge of similar data sets) to formulate an expected outcome. That outcome is the third component, and this third component is the desired end product from the analysis.

Model Framework



Methodology

Xavier Initialization and Generation of Weights

Xavier Initialization is used to generate weights such that, these weights are able to propagate across all the layers of the network without causing the vanishing gradient problems and exploding gradient problem. The model improved over 300 epochs.

One Hot Encoding

We converted categorical data to binary data. We encoded the target variable i.e. the quality measure. The wines with a quality measure above 6 were encoded as 1 (or good) and the wines with a quality measure under 6 were encoded as 0 (or bad).

Our model gave an accuracy of 89%.

Comparitive Analysis

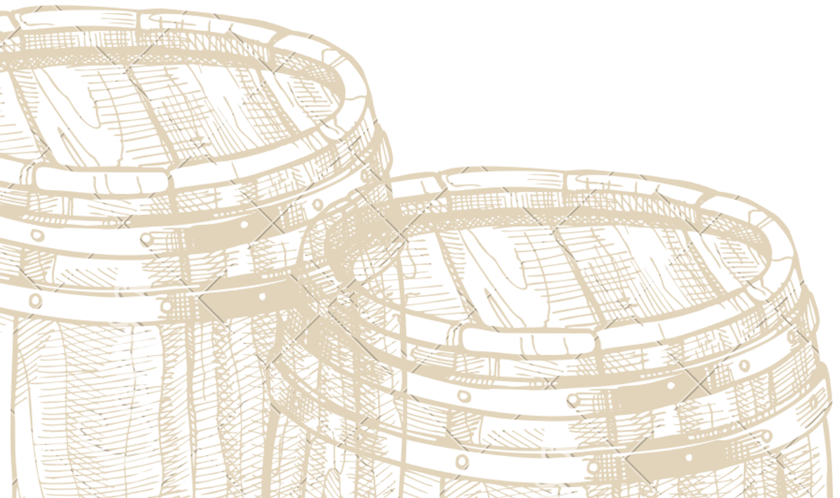
Our model was compared with in-built library models. Some models had a higher accuracy then our model whereas others had accuracies lesser than ours. The following models were part of the comparative analysis

- Logistic Regression
- K-Nearest Neighbours
- Support Vector Machines
- Decision Trees
- Naïve Bayes
- Random Forest
- XGBoost
- Deep Neural Network

Results and Discussion

Our models performance was far more superior to the performance of other models which were part of the comparative analysis. **Our model projected an accuracy of 89%.**

The only model to showcase a better performance than ours was Random Forest. **It had an accuracy of 90%.**



Model	Accuracy
Logistic Regression	88.04%
K-Nearest Neighbors	86.88%
Support Vector Machine	87.17%
Decision Tree	86.01%
Naive Bayes	85.13%
Random Forest	90.03%
XGBoost	88.62%
Neural Network	87%
Our Model	86%

References

https://koreascience.kr/article/JAKO201832073079660.pdf

https://ijirase.com/assets/paper/issue_1/volume_4/V4-Issue-4-715-721.pdf

researchgate.net/publication/344650812_Wine_Quality_and_Taste_Classification_Using_Machine_Learning_Model

scipy.org/journal/paperinformation.aspx?paperid=107796

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8702017

https://ieeexplore.ieee.org/document/9104095

jjetir.org/view?paper=JETIR2111328

researchgate.net/publication/341812162_Red_Wine_Quality_Prediction_Using_Machine_Learning_Techniques

xplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9063846

Thank You

Team

Anubuthi K

Attili Subha Vidisha

Devang Saraogi

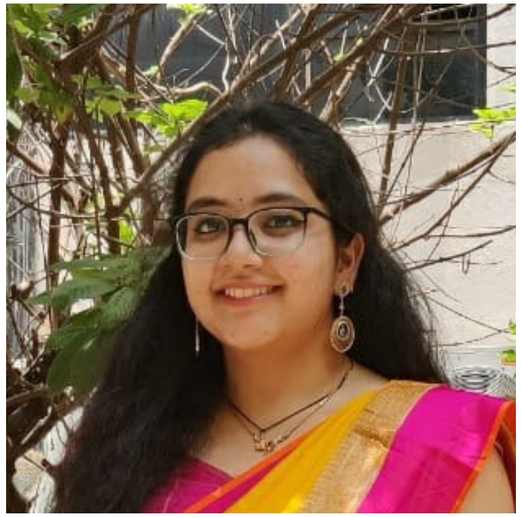
Mentor

Professor Samatha R Swamy

Members

Attili Subha Vidisha

PES1UG20CS091



Anubuthi K

PES1UG20CS065

Devang Saraogi

PES1UG20CS122