GLA UNIVERSITY, MATHURA

TOPIC: MINI PROJECT SYNOPSIS ON WINE

QUALITY ANALYSIS .

Submitted by:

Submitted to:

Faculty Name: Mr. Farmanul Haque
Technical Trainer

Name: Utkarsh Kulshrestha
Id: 191500879

Name: Rishika Sharma
Id: 191500656

DECLARATION

I hereby declare that this project work entitled "**Wine Quality Check**" has been prepared by our team during 2021-2022 under the guidance of **Mr. Farmanul Haque, Technical Trainer, Department of ComputerScience, GLA UNIVERSITY, MATHURA.** In the partial fulfillment of B.Tech degree prescribed by the college.

I also declare that this project is the outcome of effort of our team that it has not been submitted to any other university, college or any other institute for the award of any degree.

TEAM DETAILS

NAME	UNIVERSITY ROLL NO.
Utkarsh Kulshrestha	**191500879**
Rishika Sharma	**191500656**

The problem statement

The problems faced by we people and hope to be solved by the Wine Quality Analysis are described below:

- ❖ Prediction of quality on the basis of given features.
- ❖ Time consuming
- ❖ Mistakes occurring in prediction of fine substance.
- ❖ Lack of efficiency
- ❖ Data redundancy
- ❖ Data inconsistency

INTRODUCTION

Here we will predict the quality of wine on the basis of giving features. We use the wine quality dataset from Kaggle. This dataset has the fundamental features which are responsible for affecting the quality of the wine. By the use of several Machine learning models, we will predict the quality of the wine. Here we will only deal with the white type wine quality, we use classification techniques to check further the quality of the wine i.e. is it good or bad.

ABOUT THE PROJECT:-

For this project, I used Kaggle's Red Wine Quality dataset to build various classification models to predict whether a particular red wine is "good quality" or not. Each wine in this dataset is given a "quality" score between 0 and 10. For the purpose of this project, I converted the output to a binary output where each wine is either "good quality" (a score of 7 or higher) or not (a score below 7). The quality of a wine is determined by 12 input variables: Fixed Acidity, Volatile Acidity, Citric acid, Residual Sugar, Chlorides, Free Sulfur dioxide, Total Sulfur Dioxide, Density, pH, Sulphates, Alcohol, Quality.

Primary Reason to Choose This Project

As each and every sector of the market is growing, data is building up day by day, we need to keep the record of the data which can be helpful for the analytics and evaluation. Now we don't have data in gigabyte or terabyte but in zetta byte and petabyte and this data cannot be handled with the day to day software such as Excel or Matlab. Therefore in this report we will be dealing with large data sets with the high-level programming language 'Python'. The main goal of this project is to aggregate and analyze the data collected from the different data sources available on the internet. This project mainly focuses on the usage of the python programming language in the field of Wine Quality Check. This language has not only its application in the field of just analyzing the data but also for the prediction of the upcoming scenarios in the Wine Production.

The Main Objective of the Project

The objectives of this project are as follows

1. To experiment with different classification methods to see which yields the highest accuracy.

2. To determine which features are the most indicative of a good quality wine.

About Data

Fixed acidity: most acids involved with wine or fixed or nonvolatile (do not evaporate readily)

Volatile acidity: the amount of acetic acid in wine, which at too high of levels can lead to an unpleasant smell.

Vinegar taste citric acid: found in small quantities.

Citric acid can add 'freshness' and flavor to wines residual sugar: the amount of sugar remaining after fermentation stops, it's rare to find wines with less than 1 gram/liter and chlorides: the amount of salt in the wine freesulfur dioxide: the free form of SO2 exists in equilibrium between molecular SO2 (as a dissolved gas) and bisulfate ion; it prevents total sulfur dioxide: amount of free and boundforms of S02; in low concentrations, SO2is mostly undetectable in wine, but at free SO2 density: the density of water is close to that of water depending on the percent alcohol and sugarcontent pH: describes how acidic or basic a wine is on a scale from 0 (very acidic) to 14 (very basic); most wines are between 3-4 on the sulphates: a wine additive which can contribute tosulfur dioxide gas (S02) levels, which acts as an antimicrobial an alcohol: the percent alcohol content of the wine quality: output variable (based on sensory data, score between 0 and 10.

Working Methodology of the Project

1. Importing Library and Loading Database.
2. Information Of Dataset.
3. Checking For Null Value.
4. HeatMap for Correlation of Parameters.
5. Visualizing the Whole Database.
6. Quality Of Wine.
7. Splitting Data into Training Data and Testing Data.
8. Building Models.
9. Visualizing Confusion Matrix Using Seaborn.

Details about the Hardware and the Software

System Requirements: - **Windows 7/8/10**

Software Required:

- Technology Implemented: Machine Learning and Data Science.

- Language :   Python.

- Database: Google Collab, VS Code

- Browser: Google Chrome

Hardware Requirements: -

- Processor: Intel i3

- Operating System: Windows 7/8/10

- RAM: 4+GB

- Hard Disk: 64 GB

- Hardware Devices: Computer System.

What contribution would the project make and where?

This data will allow us to create different regression models to determine how different independent variables help predict our dependent variable, quality. Knowing how each variable will impact the red wine quality will help producers, distributors, and businesses in the red wine industry better assess their production, distribution, and pricing strategy.

FLOW DIAGRAM
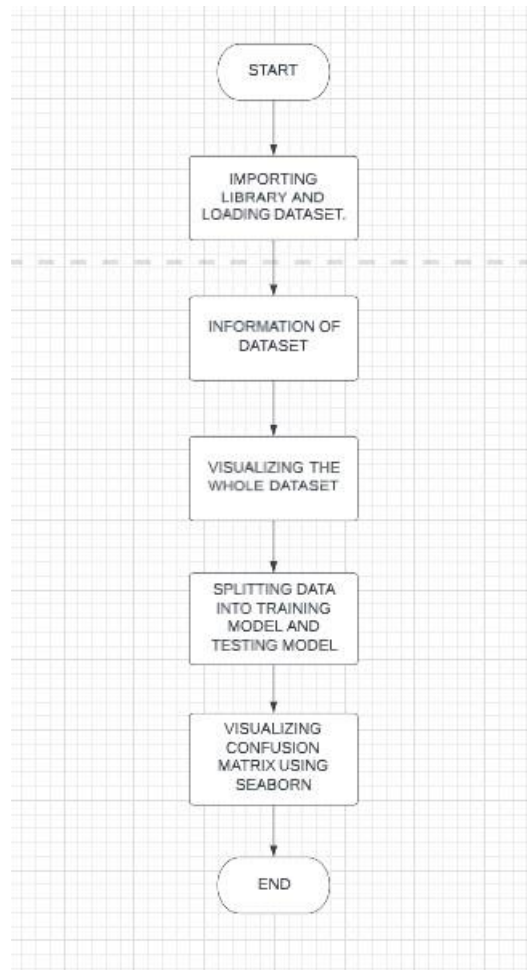
Conclusion

In conclusion we believe this project if properly utilized will save time, reduce the amount of work the administration has to do. Overall our logistic regression model performs quite well with: Accuracy, Precision and Recall but it still has something better to do so.

References

- Wikipedia
- Matplotlib https: //en.wikipedia.org/wiki/Matplotlib
- Numpy online https: //en.wikipedia.org/wiki/NumPy
- https://www.w3schools.com
- https://stackoverflow.com