WINE QUALITY PREDICTION

PROJECT

HTTPS://GITHUB.COM/SAUMYASTA/WINE-QUALITY-PREDICTION

WHAT IS WINE ?

Wine is an alcoholic beverage made from fermented grapes or other fruits. Grapes are the most common fruit used in winemaking, and the process involves crushing the grapes and allowing the juice to ferment with yeast, which converts the sugars in the juice into alcohol. The resulting liquid, called wine, can vary widely in color, flavor, and aroma, depending on the grape variety, the climate and soil in which the grapes were grown, and the winemaking techniques used. Wine has been produced and enjoyed by people for thousands of years, and it is a popular drink around the world.

WHAT IS WINE QUALITY PREDICTION ?

Wine quality prediction is the process of using machine learning algorithms to analyze data on various characteristics of a wine (such as its chemical composition, age, and vintage) in order to predict its overall quality or rating. This can be helpful for winemakers, wine sellers, and wine enthusiasts who want to better understand and appreciate the quality of different wines.

AIM OF MY PROJECT

Predict quality of wine

SOURCES

01 DATASET : https://www.kaggle.com/datasets/yasserh/wine-quality-dataset

02 MODELS USED : LOGISTIC REGRESSION, XGBCLASSIFIER, SVC

03 LANGUAGE USED : PYTHON

04 IDE USED : JUPYTER NOTEBOOK

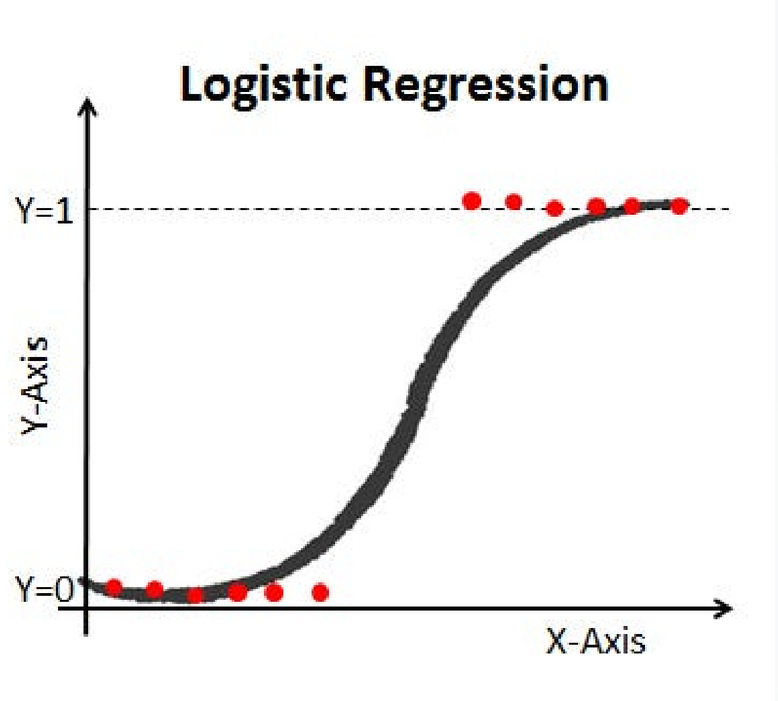
ABOUT THE DATASET :

The dataset contains 13 columns :
Fixed acidity: Higher levels of fixed acidity can give a wine a sharper taste and can be perceived as sour or tart.
Volatile acidity: Higher levels of volatile acidity can give a wine a vinegary or nail-polish-remover-like aroma and taste.
Citric acid: Citric acid can enhance the perceived freshness and fruity flavors in a wine.
Residual sugar: Higher levels of residual sugar can make a wine taste sweeter, while lower levels can make a wine taste drier.
Chlorides: Chlorides can contribute to the perceived saltiness or minerality of a wine.
Free sulfur dioxide: Free sulfur dioxide can help to preserve the wine and prevent oxidation, but high levels can cause a sulfuric aroma and taste.
Total sulfur dioxide: Total sulfur dioxide can also help to preserve the wine, but high levels can cause a sulfuric aroma and taste.
Density: Density is a measure of the wine's mass per unit volume and can be used to calculate the wine's alcohol content.
pH: pH can affect the perceived acidity of a wine, with lower pH values indicating higher acidity.
Sulphates: Sulphates can help to prevent microbial growth and oxidation, but high levels can cause a bitter or astringent taste.
Alcohol: Higher levels of alcohol can contribute to the perceived body and warmth of a wine.
Quality (score between 0 and 10): Quality is a subjective measure of how much a person enjoys the overall taste and aroma of a wine, based on sensory evaluation.

```
RangeIndex: 1143 entries, 0 to 1142
Data columns (total 13 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   fixed acidity         1143 non-null   float64
 1   volatile acidity      1143 non-null   float64
 2   citric acid           1143 non-null   float64
 3   residual sugar        1143 non-null   float64
 4   chlorides             1143 non-null   float64
 5   free sulfur dioxide   1143 non-null   float64
 6   total sulfur dioxide  1143 non-null   float64
 7   density               1143 non-null   float64
 8   pH                    1143 non-null   float64
 9   sulphates             1143 non-null   float64
 10  alcohol               1143 non-null   float64
 11  quality               1143 non-null   int64
 12  Id                    1143 non-null   int64
dtypes: float64(11), int64(2)
memory usage: 116.2 KB
```

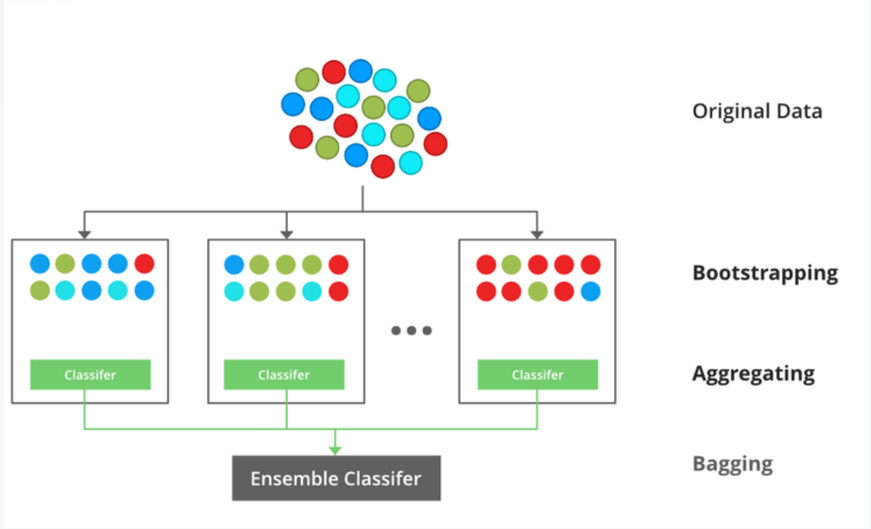
LOGISTIC REGRESSION

Logistic regression is a type of machine learning algorithm that can be used for binary classification tasks, such as predicting whether a wine is of high or low quality based on its physicochemical attributes.
In the case of wine quality prediction, logistic regression can be trained on a dataset of wines with known quality ratings, along with their corresponding physicochemical attributes. The algorithm will then learn to map these input features to the output quality rating, allowing it to predict the quality of new wines based on their attributes.
Logistic regression works by modeling the relationship between the input features and the output using a logistic function, which maps the input values to a probability of the output being either high quality or low quality. The algorithm then uses this probability to make a binary classification decision.



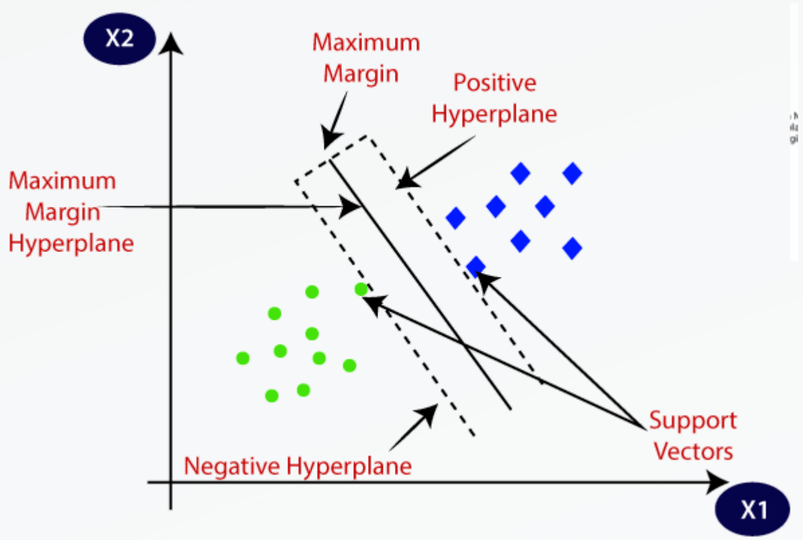
XGB CLASSIFIER

XGBoost (eXtreme Gradient Boosting) classifier is a popular machine learning algorithm used for wine quality prediction. It is a type of gradient boosting algorithm that combines multiple decision trees to make predictions, and is known for its high accuracy and efficiency. XGBoost can handle a large number of input features, making it well-suited for analyzing the many physicochemical attributes that can influence the quality of wine. Additionally, XGBoost allows for feature importance ranking, which can help identify which attributes have the most impact on the predicted wine quality.

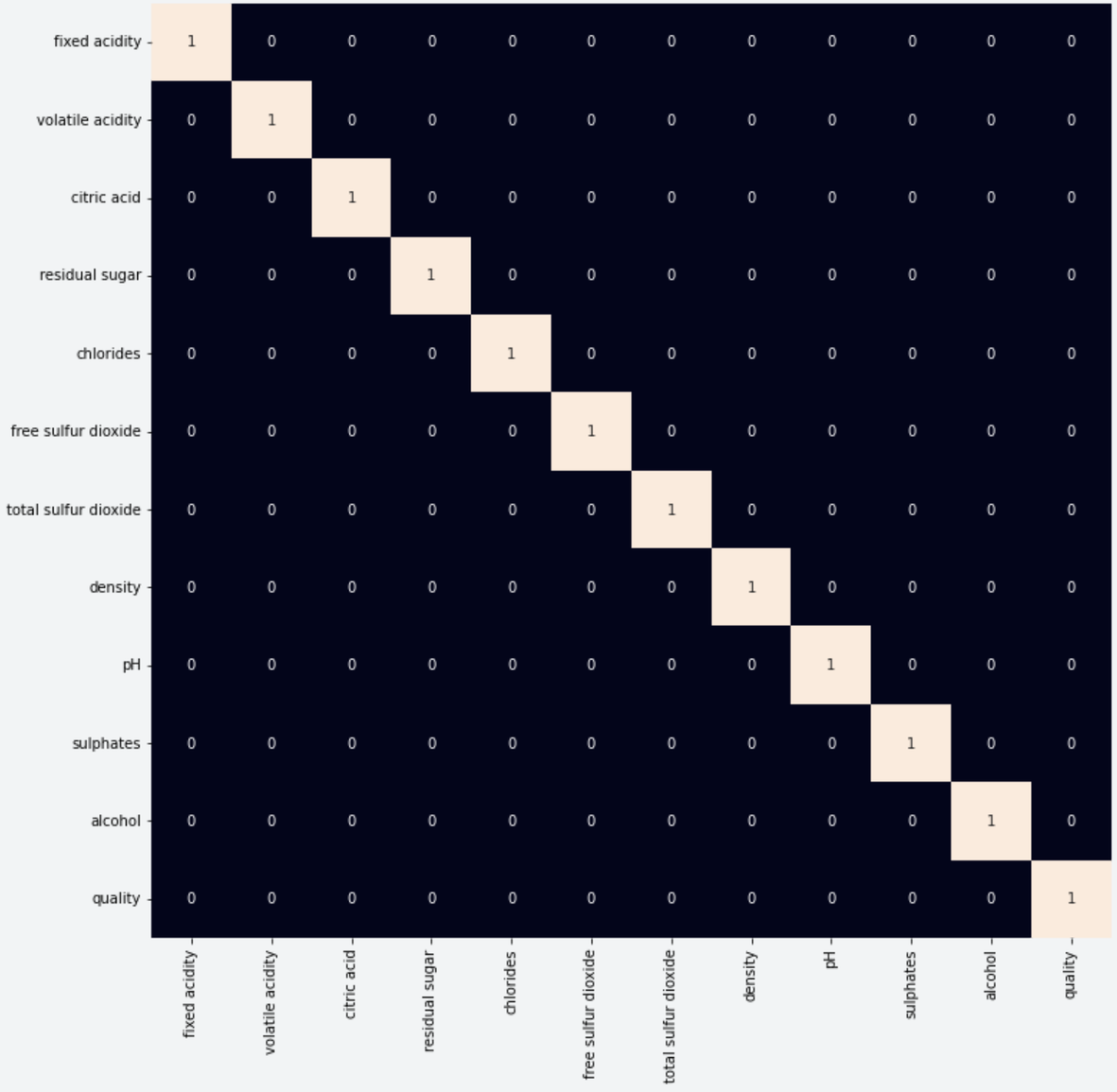


SVC

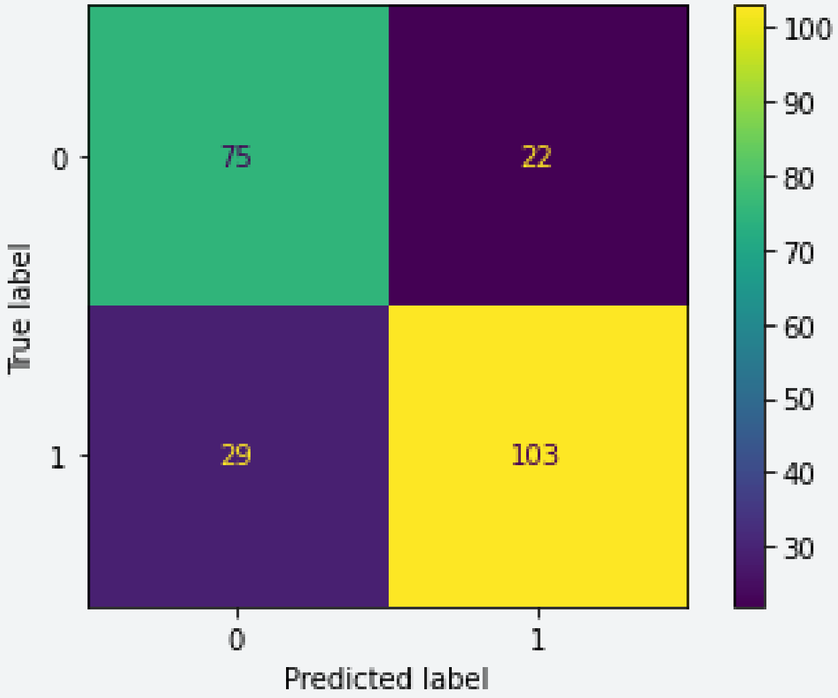
Support Vector Machine (SVM) is a type of supervised machine learning algorithm that can be used for wine quality prediction. SVM works by finding a hyperplane that best separates the data into different classes, based on the input variables (such as the physicochemical tests listed for wine quality prediction). In the case of wine quality prediction, the SVM model can be trained on a dataset of wines with known quality ratings, and then used to predict the quality rating of new wines based on their physicochemical characteristics. SVM is a popular algorithm for classification tasks and can be effective in identifying patterns and making accurate predictions, but may require careful tuning of its parameters to optimize its performance for a given dataset.



RESULT OBTAINED



RESULT OBTAINED



	precision	recall	f1-score	support
0	0.72	0.77	0.75	97
1	0.82	0.78	0.80	132
accuracy			0.78	229
macro avg	0.77	0.78	0.77	229
weighted avg	0.78	0.78	0.78	229

CONCLUSION

- The evaluation metrics suggest that the model used for wine quality prediction has an overall accuracy of 78%, meaning it correctly classified 78% of the wines in the dataset.
- Looking at the precision and recall scores, we can see that the model is slightly better at correctly identifying wines with a quality rating of "1" than "0". Precision measures the proportion of true positives (correctly classified "1" ratings) among all wines classified as "1", while recall measures the proportion of true positives among all wines that are actually "1".
- The F1-score is the harmonic mean of precision and recall and gives a balanced measure of the model's performance. In this case, the F1-score is 0.80 for rating "1" and 0.75 for rating "0", indicating that the model is slightly better at predicting the quality rating for wines with a rating of "1".
- Overall, the model's performance is reasonably good, but there may be room for improvement, particularly in correctly identifying wines with a quality rating of "0".

THANK'S FOR WATCHING