Exploring Model-Agnostic Interpretable Machine Learning for Red Wine Quality Assessment

ABSTRACT

The people of today's world like to live a luxurious life. They want to use the things for their purpose or the lifestyle showing off. Nowadays the consumption of wine is common almost in all countries. Wine is consumed by people for their stress release or their lifestyle or due to their addiction.

We need the wine quality detection for the quality check. Quality matters a lot in preserving human health. The bad quality wine may lead to health loss. It may cause diseases too. Wine is also known as alcohol. It's a drink which is made from grape juice. The alcohol is made from the juice of other fruits blackberries or plums.

The Machine Learning supervised algorithms are used in the project. The algorithms are used for the classification of the good or bad quality of red wine. The quality of the red wine is detected in this project. The dataset is pre-processed in the machine. The dataset is applied to the machine learning algorithms. The algorithms help to classify the data. The supervised classification machine learning algorithms are used in this project. The labelled data is used for the training of the dataset.

We are using the Machine Learning technique for the prediction of the wine quality. We are using a dataset. The dataset is imported into our algorithm. Then we are applying the Naïve Bayes, SVM, Linear Regression, and MLP algorithms. The red wine quality is predicted using Machine Learning algorithms.

TABLE OF CONTENTS

1. CHAPTER

1.1 INTRODUCTION

Wine is an alcoholic beverage which is being produced for thousands of years. This is made from grapes, fermented without the addition of sugar acids, enzymes, water, and other victims. Wines that are not made from grapes may include rice wine and fruit wines such as Cherry, pomegranate, and elderberry.

Wine quality assessment is one of the main key elements in this context, such type of quality assessment helps to assure the wine quality in the market and it can be decided through its attributes. A few years ago, with the availability of a lot of wine brands, it is difficult to identify good quality wines, Good Quality wine depends on important factors such as chemical, scientific and technical factors.

In the last few years, Machine Learning techniques can produce highly accurate result which is used to implement in predictive analytics, here we are considering Red Wine analysis. Red Wine is made from dark-coloured grape varieties. The quality of this wine can be predicted using different Machine Learning techniques, these techniques will allow building a model with a User Interface that predicts the wine quality by selecting the important parameters of wine which play a significant role in determining the quality of the wines and The proposed models such as Linear Regression, Naïve Bayes, SVM will determine the quality of wines on a scale of Good, Average and Bad. These proposed models can further be applied to several other products which require quality prediction. This process makes the process more efficient and cheaper with less human interaction

Nowadays the technologies which help us for the product quality check. The machines are used for the quality check of certain products. Red wine is consumed in some countries. So the human health is to be preserved. The wine quality should be checked before consumption. Thus the Machine Learning technique is used for the red wine quality prediction. The Machine Learning algorithms are applied for the red wine quality check.

Today, all type of industries is improving by adopting new technologies and applying these in all areas. These technologies are also helpful to enhance production and making the whole process smooth. But, still, there are different areas, which demand human expertise such as product quality assurance. Nowadays, it becomes an expensive process as the demand for the product is growing over time.

Therefore, this paper explores different machine learning techniques such as linear regression and support vector machines for product quality assurance. These techniques perform the quality assurance process with the help of available characteristics of a product and automate the process by minimizing human interfere. The work also identifies the important features to predict the values of dependent variables.

In this work, all the above-mentioned machine learning techniques are used to support the wine industry. Wine quality assessment is one of the key elements in this context and this assessment can be used for certification. Such type of quality certification helps to assure wine quality in the market. Wine has various characteristics like density, pH value, alcohol, and other acids. Wine quality can be assessed by two types of tests first is a physicochemical test and the second is the sensory test. The physicochemical test can be determined by lab tests and no human expert is required but for a sensory test, a human expert is required. Moreover, Wine quality assessment is very difficult as the relationships between the physicochemical and sensory analysis are complex and still not fully understood.

In literature, some researchers have used machine learning techniques to assess wine quality, but still, a huge scope is available for improvement. They used 170 samples of data from Germany for their experiments. They got a 100% predictive rate. Grape maturity level and chemical analysis are used for wine classification. A sample of 36 examples was used for experiments and achieved only 6% error.

In this paper, linear regression and SVM are implemented to determine the dependency of wine quality on different 11 physicochemical characteristics. Moreover, the predictions are also made for wine quality based on important variables/characteristics, selected according to their dependencies.

1.2 OBJECTIVE

- Nowadays, industries need product quality certifications to promote their products.
- This is a time taking process and requires the assessment given by human experts which makes this process very expensive.
- It is important to aspect to analyze the quality of red wine before its consumption to preserve human health.
- Hence this research is a step towards the quality prediction of the red wine using its various attributes.
- Machine Learning technology is used for the prediction of the red wine quality as good or bad.

1.3.EXISTING SYSTEM

The red wine quality prediction is done with the help of human experts. The wine quality is checked by the experts and the same is certified after examinations of the red wine. These technologies are also helpful to enhance production and making the whole process smooth.

But, still, there are different areas, which demand human expertise such as product quality assurance. Thus this is a time-consuming process and experts should be required to check the product quality.

Disadvantages:

1. Time-consuming.
2. Expensive.
3. Complex process.

1.4. PROBLEM STATEMENT

The red wine quality prediction is done with the help of human experts. The wine quality is checked by the experts and the same is certified after examinations of the red wine. Thus this is a time-consuming process and experts should be required to check the product quality.

The red wine is analyzed and predicted using the AI techniques that is Machine Learning used for the prediction of the red wine quality. The wine quality is detected using the dataset of various attributes. The attributes or parameters are being used in the dataset. The ML technique is used for the detection of the red wine quality based on the dataset attribute. The dataset is a numeric file that has the values in the CSV file.

It is important to aspect to analyze the quality of red wine before its consumption to preserve human health. Therefore, we use the Machine Learning technique for product quality assurance. Using this technique we are predicting the red wine quality.

1.5. AIM

The aim of our project is the prediction of red wine quality using the machine learning technique. The CSV dataset is used for the input. The red wine quality parameters are used for the prediction. The Linear Regression, SVM, Naïve Bayes & MLP algorithms are used to implement the project.

The CSV dataset is applied to the algorithm. Then the data is preprocessed and split into train and test sets. Finally, we apply the algorithms for the classification of the wine as good or bad. The red wine quality check is done by using the SVM, Naïve Bayes, MLP, and Linear Regression algorithms.

1.6. PROPOSED SPECIFICATION

The red wine is analyzed and predicted using the AI techniques that is Machine Learning used for the prediction of the red wine quality. The wine quality is detected using the dataset of various attributes. The attributes or parameters are being used in the dataset. The ML technique is used for the detection of the red wine quality based on the dataset attribute. The dataset is the numeric file that has the values in the CSV file.

It's important to aspect to analyze the quality of red wine before its consumption to preserve human health. Therefore, we use the Machine Learning technique for product quality assurance. Using this technique we are predicting the red wine quality.

The proposed method uses a dataset. The algorithms such as Naïve Bayes, Linear Regression, SVM, and MLP are used. The parameters or variables are used for the prediction of red wine quality. The CSV dataset is preprocessed at first. The preprocessed data is applied to the respective algorithms. The test data is passed for the detection of the quality of the red wine. Hence ML algorithms are used for the implementation of red wine quality prediction.

1.7 RESEARCH QUESTIONS

The Research questions have been prepared concerning the topic of the research. The research questions are as follows:

- What are the major and important factors that impact the wine quality preferences
- How understanding components behind wine quality is beneficial in wine production

These are the questions that will be answered throughout the investigation,However, it cannnt be started at this point that the research will lead to understanding all the questions that put forward for the study.

2. CHAPTER II

2.1.BACKGROUND RESEARCH

Individuals of the present world like to carry on with a luxurious life. They need to utilize the things for their motivation or for the way of life flaunting. These days the utilization of wine is normal practice in every one of the nations. The wine is consumed by individuals for their pressure discharge or for their way of life or because of their addiction.

We need the wine quality discovery for the quality check. Quality matters a great deal in protecting human wellbeing. The bad wine quality might cause infections as well. The wine is otherwise called liquor. It's a beverage that is produced using grape juice. The liquor is produced using the juice of different natural products blackberries or plums.

There are two kinds of wines. They are red wine and white wine. The wine grapes have toughness and greater seeds. The grape wine is ready by picking the aged grapes. They are picked by the hands or by the machines. After the gathering cycle, the grapes are squashed to deliver the juice. Before this, the stepping is finished by the unshod.

After aging, the wine is put away for some time called maturing. The wine can be stored steel holder or a barrel made of oak wood. On the off chance that they pick wood, the wine will retain a portion of the wood flavor, which wine consumers call "oakiness". This cycle can half a month, as long as quite a while, contingent upon what sort of flavor the wine producer needs. Finally, the wine is packaged and shipped off to a client, cafe or store to be appreciated.

Red wine is a kind of wine delivered utilizing dull tinted grape combinations. The certified shade of the wine can go from genuine violet, ordinary of young wines, through to obstruct red for foster wines and brown for more settled red wines. The juice from most purple grapes is greenish-white, the red concealing coming from anthocyanin conceals (in like manner called anthocyanins) present in the skin of the grape; uncommon cases are the by and large remarkable combinations, which produce a red-shaded juice.

An enormous piece of the red-wine creation measure consequently incorporates extraction of concealing and flavor portions from the grape skin. Red wine is used for consumption by people for their habits or addictions.

Clinical benefits of red wines are the critical conflict of those people who like red more than white. Yet red wine helps to prevent heart sicknesses and getting veins on account of its disease anticipation specialists, white wine contains fewer calories. Nevertheless, their level can move as shown by the

considerable wine. French people have less levels of heart disease on account of their affinity for drinking a glass of red wine every day. The most grounded cell fortifications contained in red wine, for instance, reservatrol make it more grounded than white one.

Reservatrol prevents the damage of veins and blood coagulation that are very hazardous for life. Thinking about the caloric substance, we can say that the two sorts have the same percent. In any case, generally, white wine contains fewer extents of calories than red wine. Coincidentally, this brand name doesn't give the veritable disputes in favor of white wine. Likewise, white wine is more acidic and riskier for the teeth.

In culinary, red wine is more standard than white one. There are docents of plans of cooking with red wine. As it had been referred to already, red wine has a more grounded flavor and scent. As cooks need to make their dish more focused and to work on its taste, aroma and concealing, they usually use red wine as a marinade, cooking liquid, or additional fixing to a finished dish.

Red wine is a kind of wine with a trademark red tone delivered from dark or darker looking grape assortments (varietals). Albeit most purple grape separates are greenish-white, the anthocyanin shades present in the grape skin give the red tone to the wine. Extraction of flavor and shading components from grape skin, thusly, includes an essential piece of the creation cycle. Commonly, red wine tones will in general change with age, for example, youthful wines are exceptional violet while experienced wines are block red, and more established red wines are brown.

The nature of wine recognition matters a ton here. So it became critical to break down the nature of red wine before its utilization to safeguard human wellbeing. Henceforth this exploration is a stage towards the quality expectation of the red wine utilizing its different traits.

The red wine is analyzed and predicted using the AI techniques that is Machine Learning used for the prediction of the red wine quality. The wine quality is detected using the dataset of various attributes. The attributes or parameters are being used in the dataset. The ML technique is used for the detection of the red wine quality based on the dataset attribute. The dataset is a numeric file that has the values in the CSV file.

2.2. LITERATURE SURVEY

Literature	Literature Review	Advantages	Disadvantages
1. PREDICTION OF QUALITY FOR DIFFERENT TYPE OF WINE BASED ON DIFFERENT FEATURE SETS USING SUPERVISED MACHINE LEARNING TECHNIQUES **Authors:** Satyabrata Aich, Ahmed Abdulhakim Al-Absi, Kueh Lee Hui, and Mangal Sain **Year**: 2018	• In recent years, most of the wine-producing industries have been promoting their products based on the quality certification they received on the products. • The products manufactured have a certification for the quality check. The old ways of assessing the product quality are time-consuming, however with the invention of machine learning techniques or machine learning technology the processes are consuming less time than before. • In this paper ,explored some of the machine learning algorithms to check the quality of the wine based on the attributes of wine that depend on quality. We have used red wine quality datasets for this research work. We have used different feature selection algorithms to check the prediction performance. • Different performance measures has been used such as sensitivity, specificity, accuracy, positive predictive value, negative predictive value for comparison using different feature sets and different supervised machine learning algorithms. • Used linear, nonlinear, and probabilistic classification methods. We have found that feature selection-based feature sets can provide better prediction than	Effective method.	Time-consuming.

	considering all the features for performance prediction. • Resulted the accuracy ranging from 95.23% to 98.81% with different feature sets. • This analysis will help the industries to check the quality of the products in a more efficient way and in less time.		
2. AHP AND MACHINE LEARNING TECHNIQUES FOR WINE RECOMMENDATION **Authors:** Kunal Thakkar, Jay Shah, Raghav Prabhakar, Anant Narayan, Anurag Joshi **Year:** 2016	• In today's modern competitive world it is very necessary to check the wine quality a company must keep its customers loyal and similarly equally important to provide better recommendations to its customers. • Artificial Intelligence computer tools and the particular machine learning are used to facilitate and enhance recommendations as it is promising and is assumed to be a modern world trending aspect. • A variety of these algorithms, including Support Vector Machines(SVM's) and Random Forest(RF's), have been widely used for the prediction of wine quality. It recommends the wine to be consumed or not. • The ML models are used for the accurate learning of input data and to analyze the pattern present in the dataset. • In our paper, we used modern and effective algorithms like SVM and Random Forest which group together the dataset and provides deep learning.	• It works well with a clear margin of separation • It is effective in high-dimensional spaces.	• It doesn't perform well when we have large data set because the required training time is higher.

	• In this domain we show that the performance of these methods is better than that of previous methods, therefore promising a more comprehensive and generic approach for recommending wine based on certain parameters.		
3. ASSESSING WINE QUALITY USING A DECISION TREE **Authors:** Seunghan Lee, Juyoung Park and Kyungtae Kang **Year**: 2015	• The consumers of wine are curious about the quality, as the wine taste matters a lot. • The quality of the wine consumed is to be a topic to discuss. • There are many methods used for the assessment of wine quality. • There is a proposed method to verify or test the wine quality using the Decision Tree algorithm. • We used a dataset which are having few attributes of wine. The results that we got are 60% with traditional assessment techniques.	• Decision trees are easy to interpret and visualize. • It can easily capture Non-linear patterns.	• Sensitive to noisy data. • It can overfit noisy data.
4. WINE QUALITY PREDICTION MODEL USING MACHINE LEARNING TECHNIQUES **Author:** Rohan Dilip Kothawade **Year:**2021	• The quality of a wine is an important aspect for the consumers as well as the wine industry. • The traditional way of measuring wine quality is a time-consuming process. • Nowadays, machine learning models are important tools to replace the tasks performed by humans. • In this case, there are several parameters to predict the wine quality but the entire parameters will not be relevant for better prediction.	• It works well with a clear margin of separation • It is effective in high-dimensional spaces.	• It also doesn't perform very well, when the data set has more noise i.e. target classes are overlapping.

	• Work is focused on what wine features or parameters are important to get the promising output. • A dataset of red and white wine. • The Main purpose of classification and evaluation using relevant parameters, we have used three algorithms namely support vector machine (SVM), naïve Bayes (NB), and artificial neural network (ANN). • In this study, we used two wine quality datasets red wine and white wine. • To verify the feature importance we used the Pearson coefficient correlation and performance measurement matrices such as accuracy, recall, precision, and f1 score for comparison of the machine learning algorithms. • A grid search algorithm is applied for improving the model accuracy. • Finally, we achieved the artificial neural network (ANN) algorithm has better prediction results than the Support Vector Machine (SVM) algorithm and the Naïve Bayes (NB) algorithm for both red wine and white wine datasets.		

5. RED AND WHITE WINE DATA ANALYSIS PREDICT QUALITY OF WINE **Author:** Gregory D. Nelson **Year:2020**	• Model analysis is a widespread method and has many applications in various fields. • The Main aim is to analyze the red and white wine and predict the quality of the wine. • This topic is helpful for the consumer where quality is necessary for red wine consumption. • The wine manufacturers are interested so to produce a wine for which the consumers are interested. The analysis is done by comparing several models, including traditional linear models, Least Absolute Shrinkage and Selection Operator (LASSO) models, and the more sophisticated Gaussian process. The model performance will be compared through a training/test set split, as well as K-fold cross-validation.	• It can do the feature selection. • It is fast in terms of inference and fitting.	• It can also produce models that make no sense. • It ignores nonsignificant variables that may, nevertheless, be interesting or important.

3. CHAPTER

3.1.CHOICE OF METHODS

We have used four algorithms for the project implementation. We used SVM, Naïve Bayes, Linear regression, and MLP algorithms. The algorithms compared to each other will help us to the red wine quality check. They have their structure and methods using the pandas' library for reading the dataset. These algorithms read the dataset. The data is preprocessed. Then the data is trained using these algorithms such as SVM, Naïve Bayes, Linear regression, and MLP.

These algorithms give us an accuracy score. Almost each of the algorithm accuracy scores is different from each other. Sometimes accuracy may be the same in different algorithms.

We used the Linear Regression algorithm for the implementation of the project. The algorithm was used for the classification of the red wine type. It's a supervised machine learning algorithm. The supervised machine learning algorithm has the labeled training dataset. The labeled training dataset is used for the classification of the red wine. The Linear Regression algorithm is used for the detection of the quality of the red wine. The algorithm produces an accuracy score. The accuracy score is the ratio of correct predictions to the test data input.

We used the Support Vector Machine algorithm for the implementation of the project. The algorithm was used for the classification of the red wine type. It's a supervised machine learning algorithm. The supervised machine learning algorithm has the labeled training dataset. The labeled training dataset is used for the classification of the red wine. The Support Vector Machine algorithm is used for the detection of the quality of red wine. The algorithm produces an accuracy score. The accuracy score is the ratio of correct predictions to the test data input.

We used the Naïve Bayes algorithm for the implementation of the project. The algorithm was used for the classification of the red wine type. It's a supervised machine learning algorithm. The supervised machine learning algorithm has the labeled training dataset. The labeled training dataset is used for the classification of the red wine. The Support Vector Machine algorithm is used for the detection of the quality of red wine. The algorithm produces an accuracy score. The accuracy score is the ratio of correct predictions to the test data input

We used the Naïve Bayes algorithm for the implementation of the project. The algorithm was used for the classification of the red wine type. It's a supervised machine learning algorithm. The supervised machine learning algorithm has the labeled training dataset. The labeled training dataset is used for the classification of the red wine. The Naïve Bayes algorithm is used for the detection of

the quality of red wine. The algorithm produces an accuracy score. The accuracy score is the ratio of correct predictions to the test data input

Thus the algorithms achieve a different accuracy score. It's the ratio of correct predictions to the given test data.

3.2.JUSTIFICATION AND SUPPORT OF CHOICES

The ML algorithms are used to predict future data. The future data is predicted based on the trained data. The data is a dataset that is numeric data. The dataset is having some variables or we call it parameters. The parameters are necessary for the project implementation.

The dataset parameters are considered and applied to the project. Thus the project is implemented with the training process. Then the test data is passed. Finally, accuracy is displayed and prediction is done. The red wine quality is predicted.

Classification:

The classification algorithm is used when the output variable is categorical.

- It is used for the prediction from the two classes such as 1/0, Yes/No, Male/Female, True/false. etc.
- The output variable must be a discrete value(Ex:0,1,2).
- Examples:

    a) Customer Purchase(has defined with the labels 0 or 1): If it's 1 means the customer will purchase and if it's 0 then the customer won't purchase the product.

    b) Classifying whether a patient has the disease or not.

    c) Classifying whether an email is a spam or not.

The following Machine Learning Algorithms are used to predict the quality of wine

1. Linear Regression:

Linear Regression is a type of supervised classification algorithm to predict the output.

Linear regression makes predictions for categorical or numeric variables. For example sales, salary, product price, age, etc. Linear Regression works on the categorical type of input.

Linear regression algorithm shows a linear relationship between a dependent and one or more independent variables, hence called linear regression. The Independent is the y variable and x is the dependent variable. Since linear regression shows the linear connection, which means it finds how the value of the dependent variable is varying based on the value of the independent variable.

Advantages:

- Simple and easy algorithms which work on regression.
- It is an efficient method to train.

Disadvantages:

- It is susceptible to outliers.
- It is susceptible to noise.

We chose this method because it performs well with the input training data. Thus the accuracy score is less compared to the SVM algorithm. So we chose this to prove the accuracy score.

2. Support Vector Machine (SVM) :

Support Vector Machine algorithm is a supervised machine learning classification algorithm that constructs a hyperplane to predict the output.

Support Vector Machine or SVM is one of the most common Supervised Learning models used for classification and regression analysis . The goal of the SVM algorithm is to create the best line, decision boundary the hyperplane that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the coming future. This best decision border is called a hyperplane.

Advantages:

- Good accuracy.
- Faster prediction.
- Less memory usage.

Disadvantages:

- High training time.

We chose this method because it performs well with the input training data. Thus the accuracy score is higher compared to all other remaining algorithms. So we chose this to prove the accuracy score.

Naive Bayes:

Naïve Bayes algorithm is a type of supervised machine learning classification algorithm which is constructed on Bayes Theorem to predict the output. Naïve Bayes algorithms are used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset.

We use the Naïve Bayes concept where such as if the fruit is identified based on color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identifying that it is an apple without depending on each other. It is called Bayes because it depends on the principle of Bayes' Theorem.

Advantages:

- Simple and most effective Classification algorithm.
- Fast and accurate technique.
- Efficiently works on a large dataset.

Disadvantages:

- NB is allowing for all the variables independent that contribute to the probability.

We chose this technique because it is predicting well for all the input variables. Thus Naïve Bayes is used for improved prediction with a dissimilar number of input variables.

Multi Layer Perceptons (MLP):

A multilayer perceptron (MLP) is a feed-forward artificial neural network. It produces a set of outputs from a set of inputs. An MLP is characterized by numerous layers of neural networks. The MLP algorithm uses the backpropagation method to exercise the network. MLP is a deep learning method.

The neural network works with the neurons present in the different layers. Deep learning has a neural network with an input layer, hidden layer, and output layer. The input is received by an input layer and a hidden layer is used to transfer the processed data. Finally, the classification is done in the output layer.

Advantages:

- It can be applied to non-linear problems.
- Quick prediction after training.
- Efficient predictions.

Disadvantages:

- Quality of data is required for the training.
- Time is required for training if data is huge.

We chose this method because it’s an efficient method for the prediction of the input. It is quick in prediction after the training process.

3.2.PROJECT DESIGN/DATA COLLECTION

The project which we are implementing uses Machine Learning technology. Machine Learning is used for the prediction of red wine quality. The dataset is used for the applying of the machine. The dataset is having numeric values. The dataset is used for the preprocessing and building of the prediction model.

The dataset is having certain red wine-related parameters. Each parameter has a meaning which will be used for the model prediction. Thus the ML technique is using a CSV dataset for the red wine quality check. The proposed model used the red wine dataset for the prediction of the quality in terms of good or bad.

The project flows as shown below steps:

Step1. Importing the libraries

In this project, we are mainly focusing on using Machine Learning Libraries which can be reusable in the code for calling the method multiple times

These Machine Learning libraries are certainly used in the code to reduce the code complexity and can be accessed by different modules. We are using the Pandas library which makes data exploration and manipulation easy.

Compared to languages like C++ or C, Python libraries do not pertain to any specific context in Python. Here, a 'library' loosely describes a collection of core modules. The library is installed with the python installs packages command.

The library is also known as the set of useful functions that eliminate the need for writing codes from scratch. We are using the libraries for the reusability in the next part of the code such as pandas. We use the pandas' library mainly in this project.

Step2. Importing the dataset

Dataset can be many formats such as CSV, TXT, URL, XSL, R. In this project, we are importing CSV format dataset, this dataset consists of comma-separated values.

The dataset parameters are loaded using python and can be applied to the machine. By using these parameters, we are fetching the data from the CSV file.

The Red Wine dataset has the parameters which we consider for the project, those parameters help us get the values from the CSV dataset.

A collection of instances is a dataset and when working with machine learning methods we typically need a few datasets for different purposes. Each column in the dataset is called as Feature or an Attribute.

Each Feature will have its own datatype,it can be real or integer value or may have a categorical or ordinal value.It can be either strings, dates, times, and more complex types, but typically they are reduced to real or categorical values when working with traditional machine learning methods.

Some Features may be inputs to a model (the predictors) and others may be outputs or the features to be predicted.

Step3. Splitting dataset into Train and Test

The main objective of splitting the dataset is to estimate the performance of the Machine Learning model on the new data, The Splitting dataset into Train and Test can be used only for the largest datasets.

Dataset can be partitioned into two different sets such as the preparation set which is also known as the Training set and the Test set. After splitting the dataset, the estimated performance of the dataset can be either overly optimistic and overly pessimistic

Splitting of dataset into two different sets is one of the pivotal strides of information pre-processing which can lead to improve the presentation of our AI model.

Assume, on the off chance that we have offered to prepare our AI model by a dataset and we test it by a unique dataset. Then, at that point, it will make challenges for our model to comprehend the connections between the models.

If we train our model very well and its training accuracy is also very high, but we provide a new dataset to it, then it will decrease the performance. So we always try to make a machine learning model which performs well with the training set and also with the test dataset.

We are splitting the data into train and test. The whole dataset is split into train and test using the split functions of the Machine Learning technique.

Step4. Applying the algorithm

The Classification algorithm is a Supervised Learning technique that is used to identify the category of new observations based on training data. In Classification, a program learns from the given dataset or observations and then classifies new observations into many classes or groups.

Such as, Yes or No, 0 or 1, Spam or Not Spam, cat or dog**,** etc. Classes can be called targets/labels or categories. The red wine quality dataset as imported earlier helps us for the prediction of the wine quality in the next step.

We are using Naïve Bayes, SVM, Linear Regression in the project for the implementation of the classification process. The red wine dataset is split into train and test. Then the algorithm is used to apply the data for the classification of the dataset.

Step5. Prediction

We are using the algorithms for the prediction. We used the Naïve Bayes, SVM, Linear Regression algorithms. The prediction is nothing but detecting the future output. Machine learning helps to predict future data.

We are passing the various dataset parameters in the project. The parameter values are passed for the splitting into train and test. The same is implemented in the algorithms. The algorithm finds the data in the dataset.

The red wine quality is predicted in this stage. It is good or bad. It depends on the dataset parameters. Hence Machine Learning technology is used for the prediction.

4. CHAPTER IV

QUALITY AND RESULTS

4.1 HARDWARE AND SOFTWARE REQUIREMENTS

The hardware and software requirements are for our project as shown below in Tables 1 & 2.

Table1. Hardware System Configuration:

The table below shows the Processor as 7 Cores. Speed should be 1.1 GHz. The system RAM should be 8 GB. The 256 SSD.hard disk is better for memory for the project.

Processor	7 Cores
Speed	1.1 G Hz.
RAM	8 GB.
Hard Disk	256 SSD.

Table2.Software System Configuration:

The Operation System used is Mac OS. We are using Python programming and Machine Learning technology. On the front end, we are using Graphical User Interface(GUI).

Operating System	Mac OS
Technology	Machine Learning.
Front End	GUI-tkinter.
IDLE	Python 3.7 or higher.

4.2 DATASET:

A collection of instances is a dataset and when working with machine learning methods we typically need a few datasets for different purposes. The dataset we use is a CSV file that is a comma-separated values file. The dataset parameters are loaded using python and applied to the machine.

Structure of DataSet

The dataset is having the following attributes or features such as follows:

a. **Fixed Acidity** : It is a non volatile acid which does not evaporate readily
b. **Volatile Acidity:** It is a hight acetic acide which leads to an unpleasant vinegar taste
c. **Citric Acid**: It acts as a preservative to increase acidity,it is added in small quantities as flavour to wines
d. **Residual Sugars :** The amount of sugar remains after fermentation process is completed.The main key is to have a balanced sweetness and sourness
e. **Chlorides:** Chlorides are the amount of salt present in the wine
f. **Free sulfur dioxide:** It prevents the oxidation of winr and microbil growth
g. **Total sulfur dioxide:** It consists of amount of SO2
h. **Density:** The sweeter wines having much density
i. **pH:** The acidity level in wines
j. **Sulphates:** It contributes the SO2 levels and acts as an antimicrobial and antioxidant
k. **Alcohol:** The amount of alcohol added in wine
l. **Quality:** One of depedent variable in the dataset is Quality which is used to get the quality of each model or attribute.

According to the dataset which is used in this project,figured out that the Average values of Quality in the dataset is **5 & 6**

5. CHAPTER V

5.1 SYSTEM ARCHITECTURE

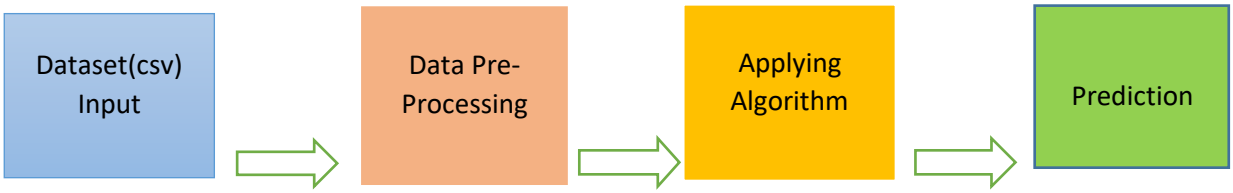

Figure 1.a : "Architecture of Red Wine Quality Prediction"

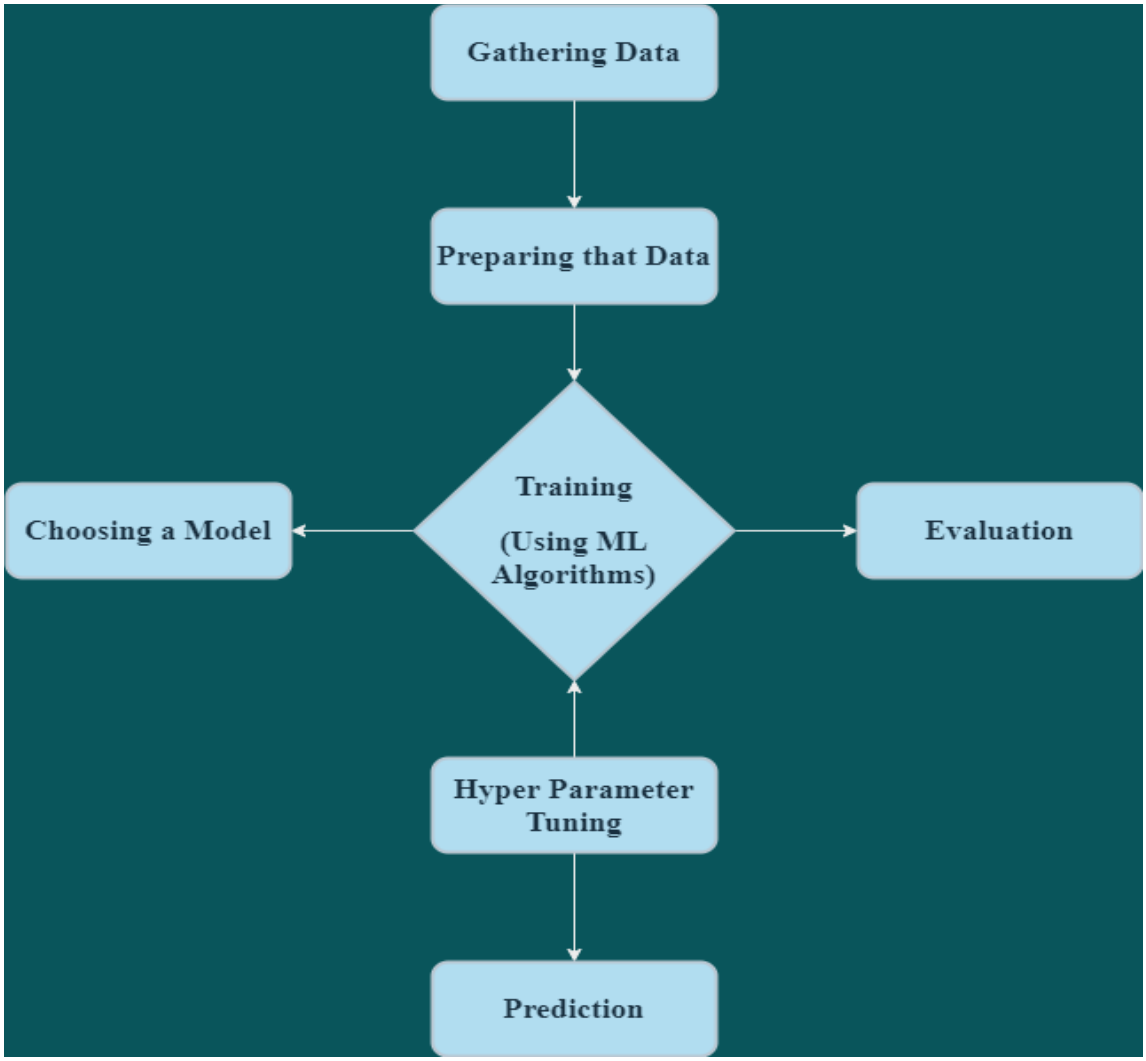

Figure 1.b: "Architecture of Red Wine Quality Prediction"

Dataset:

A collection of instances is a dataset and when working with machine learning methods we typically need a few datasets for different purposes. The dataset we use is a CSV file that is a comma-separated values file. The dataset parameters are loaded using python and applied to the machine.

A single column of data is called a feature. It is a component of observation and is also called an attribute of a data instance. Some features may be inputs to a model (the predictors) and others may be outputs or the features to be predicted.

Data Preprocessing:

In AI information pre-processing, we partition our dataset into a preparation set and test set. This is one of the pivotal strides of information pre-processing as by doing this, we can improve the presentation of our AI model.

If we train our model very well and its training accuracy is also very high, but we provide a new dataset to it, then it will decrease the performance. So we always try to make a machine learning model which performs well with the training set and also with the test dataset.

We are splitting the data into train and test. The whole dataset is split into train and test using the split functions of the Machine Learning technique.

Applying algorithm:

The Classification algorithm is a Supervised Learning technique that is used to identify the category of new observations based on training data. In Classification, a program learns from the given dataset or observations and then classifies new observations into many classes or groups.

We are using Naïve Bayes, SVM, Linear Regression in the project for the implementation of the classification process. The red wine dataset is split into train and test. Then the algorithm is used to apply the data for the classification of the dataset.

Prediction:

We are using the algorithms for the prediction. We used the Naïve Bayes, SVM, Linear Regression algorithms. The prediction is nothing but detecting the future output. Machine learning helps to predict future data.

We are passing the various dataset parameters in the project. The parameter values are passed for the splitting into train and test. The same is implemented in the algorithms. The algorithm finds the data in the dataset.

The red wine quality is predicted in this stage. It is good or bad. It depends on the dataset parameters. Hence Machine Learning technology is used for the prediction.

5.2.DATA FLOW DIAGRAM(DFD)

A data flow diagram shows the processes in the project. It shows the flow of the project block by block as shown in the below diagram.

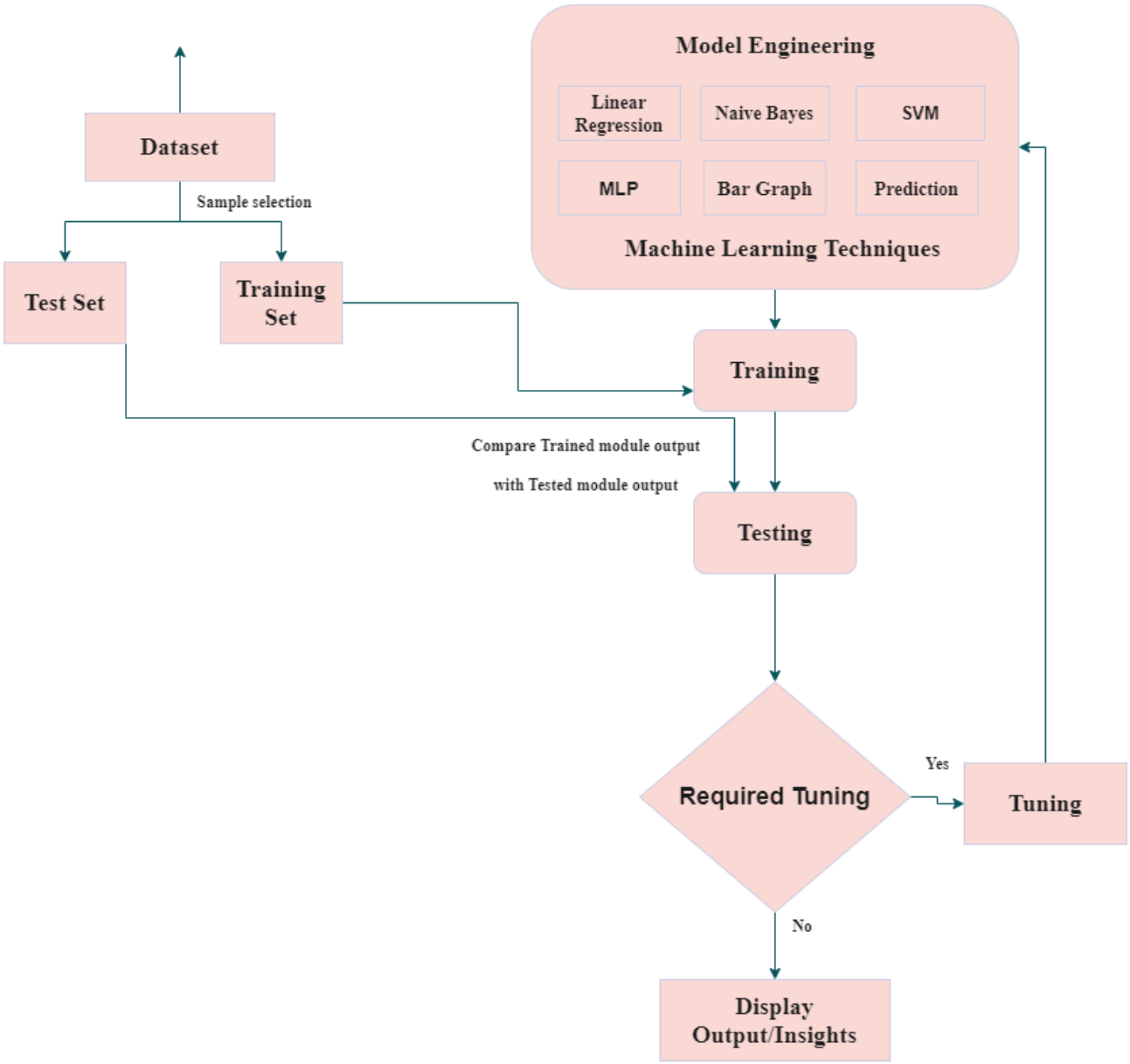

Figure 2:" Data Flow Diagram for Red Wine Quality Prediction"

The Naïve Bayes, Support Vector Machine, Linear Regression, and Multi-Layer Perceptron algorithms are applied for the prediction. The input is given with the parameters of the red wine dataset. Finally, the prediction is done. The red wine quality is detected i,e good or bad.

6. CHAPTER VI

6.1 ALGORITHMS

1. **Linear Regression:** Linear Regression is a type of supervised machine algorithm used for classification. Linear Regression is a supervised type of machine learning algorithm.

   Linear regression is one of the calmest and most commonly used Machine Learning algorithms. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as sales, salary, product price, age, etc.

   Linear regression algorithm shows a linear relationship between a dependent and one or more independent variables, hence called linear regression. The Independent is the y variable and x is the dependent variable. Since linear regression shows the linear connection, which means it finds how the value of the dependent variable is varying based on the value of the independent variable.

   The linear regression model provides a sloped straight line representing the relationship between the variables. The below image shows the Linear Regression algorithm:

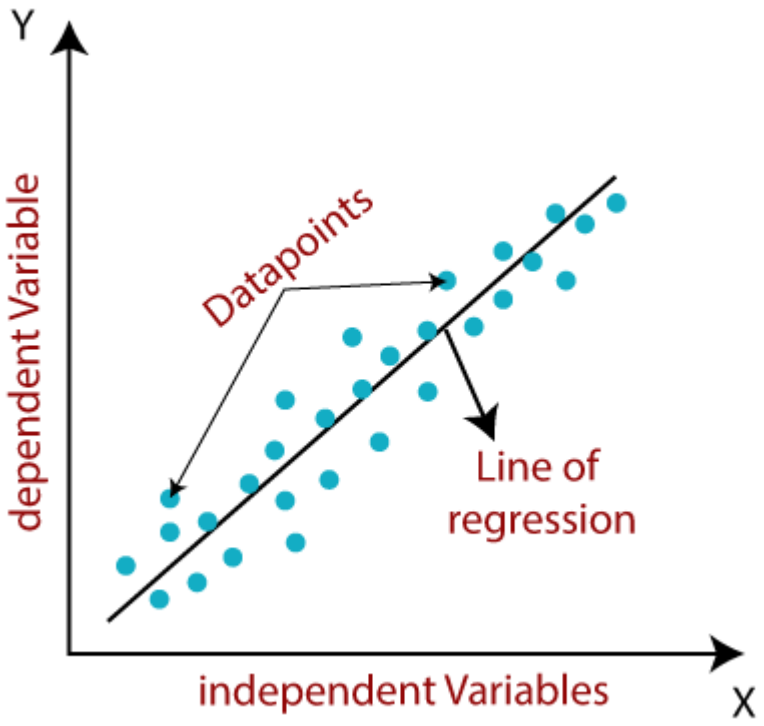

Figure 3: "Linear Regression Model"

2. **Support Vector Machine (SVM):** SVM is a supervised type of machine learning algorithm. It is used for the classification of the input data. Hence we call it's a supervised classification type of machine learning algorithm.

   Support Vector Machine or SVM is one of the most common Supervised Learning algorithms used. The SVM algorithm is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. We apply SVM for classification and finally, the input data is classified.

   The goal of the SVM algorithm is to create the best line, decision boundary the hyperplane that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the coming future. This best decision border is called a hyperplane.

   SVM selects the extreme points or the vectors that help in creating the hyperplane line. These extreme cases are called support vectors, and hence algorithm is termed a Support Vector Machine. Let's consider the below diagram in which two different groups are categorized using a decision boundary or the hyperplane:

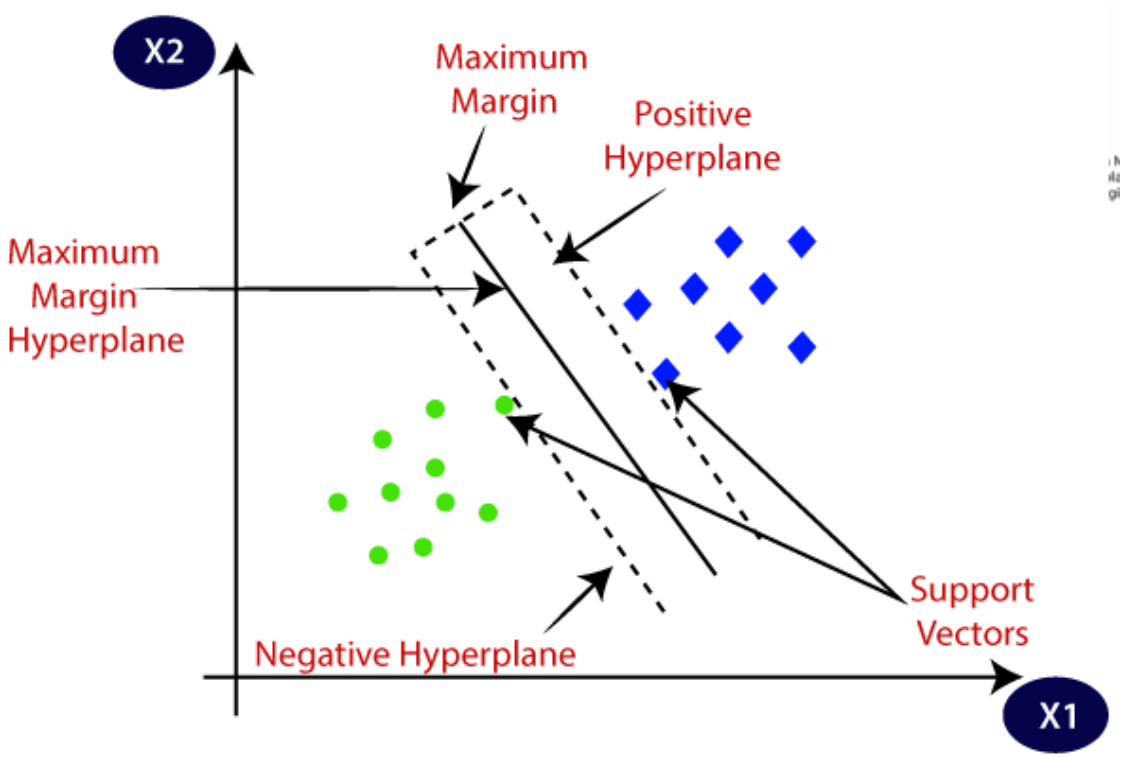

Figure 4: "Linear Regression Model"

3. **Naïve Bayes:** Naïve Bayes is a supervised type of machine learning algorithm. It is used for the classification of the input data. Hence we call it's a supervised classification type of machine learning algorithm.

   - Naïve Bayes algorithm is a supervised learning algorithm, which is based on the Bayes theorem and used for solving classification problems.
   - It is mainly used in text classification that includes a high-dimensional training dataset.
   - Naïve Bayes Classifier is one of the modest and best effective Classification algorithms which helps us in constructing the fast machine learning models that can make the fastest predictions.
   - Naïve Bayes is a probabilistic classifier.
   - The popular applications of the Naive Bayes Algorithm are Spam filtration, Sentimental analysis, and classifying of the articles.

   Why is it called Naïve Bayes?

   The Naïve Bayes algorithm is comprised of two words Naïve and Bayes, which can be explained as:

   - It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features or parameters. Such as if the fruit is identified based on color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identifying that it is an apple without depending on each other.
   - It is called Bayes because it depends on the principle of Bayes' Theorem.

   Bayes' Theorem:

   - Bayes' theorem is also known as Bayes' law or Bayes' Rule, which is used to determine the probability. It depends on conditional probability.
   - The formula for Bayes' theorem is given as:

$$P\left(\frac{A}{B}\right) = \frac{P(B/A)P(A)}{P(B)}$$

Where

A, B = events

P(A|B) = probability of A given B is true

P(B|A) = probability of B given A is true

P(A), P(B)= the independent probabilities of A and B

4. Multilayer Perceptron:

A multilayer perceptron (MLP) is a feed-forward artificial neural network. It produces a set of outputs from a set of inputs. An MLP is characterized by numerous layers of neural networks. The MLP algorithm uses the backpropagation method to exercise the network. MLP is a deep learning method.

The neural network has an input, hidden, and output layer. The input layer takes the input of different forms and passes to a hidden layer of neurons for processing. Finally, the output layer presents the output for the classification of the new input.

The pictorial demonstration of multi-layer perceptron learning is as shown below-

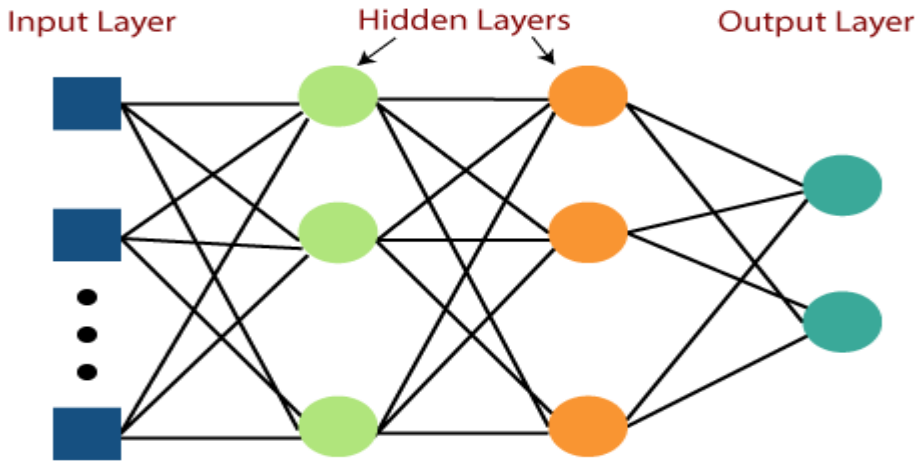

Figure 5: "Multilayer Perceptron"

7. CHAPTER VII

7.1.PYTHON AND MACHINE LEARNING

Python is a general popular programming language. It was created by Guido van Rossum, and released in the year 1991.

What can Python do?

- Python can be used on a server to create web-based applications.
- Python can connect to database systems. It can also read and modify files.
- Python can be used to handle big data.
- Python can perform complex math, mathematical calculations.
- Python can be used for fast prototyping or production-ready software development.

Why Python?

1. Python program mechanisms on different platforms (Windows, Mac, Linux, Raspberry Pi, etc).
2. Python program has a humble syntax similar to the English language.
3. Python program has a syntax that allows inventors to transcribe programs with fewer lines than some other programming languages.
4. Python runs on an interpreter system, which means that code can be executed as soon as it is written. This means that prototyping can be very quick.
5. Python can be treated procedurally, in an object-oriented way, or in a functional way.

- The most recent major version of Python is Python 3. However, Python 2, although not being simplified with whatever other than security modernizes, is still quite popular.
- Python will be written in a text editor. It is possible to write Python in an Integrated Development Environment, such as Thonny, Pycharm, Netbeans, or Eclipse which are particularly useful when managing larger collections of Python modules.
- Python was calculated for readability and has some similarities to the English language.
- Python uses new lines to complete a command, as conflicting to other programming languages which repeatedly use semicolons or parentheses.
- Python depends on indentation, using whitespace, to describe scope; such as the opportunity of loops, functions, and classes.

Python Syntax:

Python syntax can be executed by writing directly in the command prompt:

print(“Hello World”)

Or by creating a python file on the server, using the .py file extension, and running it in the command prompt:

C:\Users\Your Name>python myfile.py

7.1.2.MACHINE LEARNING

Machine learning is a growing technology that helps computers to learn automatically from past information. Machine learning exercises various algorithms for building mathematical models and making predictions using historic data or information**.** Presently, Machine Learning is being used for numerous tasks such as email filtering**,** image recognition**,** speech recognition, Facebook auto-tagging**,** recommender system, and many more tasks.

What’s Machine Learning

Machine Learning is a subset of artificial intelligence and that is mainly concerned with the development of ML algorithms that allow a computer to learn from past data and past experiences on its effort. The term machine learning was introduced first by Arthur Samuel in 1959. We can define in a brief way as:

With the usage of historical data or examples, we call training data. The training data is also known as the training dataset. The machine learning algorithms build a mathematical model which helps in making the predictions or the decisions without being explicitly programmed by the program. Machine learning constructs or uses the algorithms which learn from previous or historical data. The more the dataset input the more is the performance of the machine learning model.

A more amount of input dataset is required for the machine learning system. Once the input dataset is given then the machine learning learns from the input. The dataset may be CSV, image, audio, etc. Thus the machine learning system learns from the past data. Finally, it produces the predicted output.

Why Machine Learning?

A machine learning system learns from historical or past data. It builds the prediction models. Whenever it receives new data it predicts the output. The accuracy score of the predicted output depends on the amount of data. A large amount of the data help to build a better ML model. Then the output is predicted accurately.

Let's assume we have an issue where we need to play out certain predictions, so better composing or creating the code for it, we simply need to take care of the information to conventional calculations, and with the help of these calculations, the ML system foresees the output. The AI system has changed our perspective about the issue. The below block diagram shows and explains the working of the machine learning algorithm.

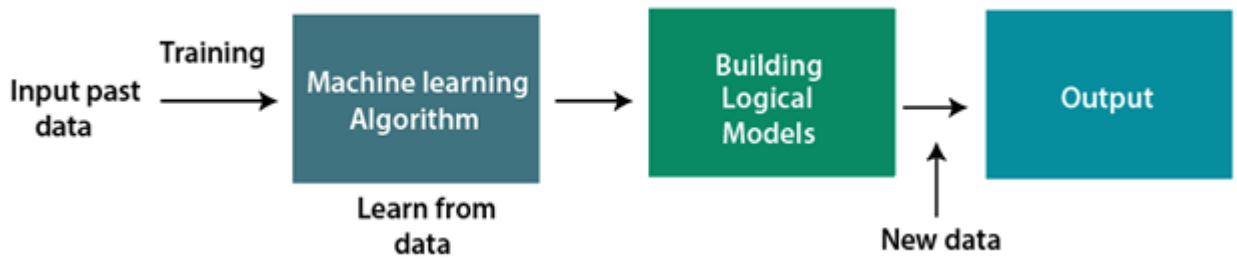

Figure 6 : "Architecture of Machine Learning"

Features of Machine Learning in

- Machine learning uses data to notice various outlines in a given dataset.
- Machine learning can learn from past data or information and advances robotically.
- Machine learning is a data-driven technology.
- Machine learning is much similar to data mining as it also deals with a huge amount of datasets.

Need for Machine Learning

Machine learning is capable of doing complex tasks. The ML field is used every day nowadays. The complex tasks cant be implemented by the human. As a human, we cant access or obtain the huge amount of the data or information and learn from that in short duration of time. So we need the help

of some faster computer systems which learn rapidly. The ML system does the required calculations and produces the output within a very short duration of time.

We can train machine learning algorithms by providing them a huge amount of data and let them explore the data, construct the models, and predict the required output automatically. The performance of the machine learning algorithm depends on the amount of data, and it can be determined by the cost function. With the help of machine learning, we can save both time and money.

We can train the machine learning system by providing them a huge amount of the data or information and allow them to explore the information from the input data. Then the ML machine constructs the model. Finally predicts the out automatically.

The performance of the ML model depends on the amount of input data. Hence the cost function determines the performance of the ML model. We can save time and money using machine learning technology.

The advantages of AI system is it can be used in digital misrepresentation discovery, face acknowledgment, self-driving vehicles, and companion ideas by Facebook, and so forth Different top organizations, for example, Amazon and Netflix have gathered the AI models that are utilizing a huge measure of information to analyze the user intrigue and suggests item as it needs accordingly.

Some key points which show the importance of Machine Learning is as shown below:

- Fast increment in the production of data
- Solving complex problems, which are difficult for a human.
- The decision-making in various sectors including finance.
- The finding of hidden patterns in the input data.
- Extracting useful information from the input data.

Classification of Machine Learning

At a broad level, machine learning can be classified into three types:

1. **Supervised learning**
2. **Unsupervised learning**
3. **Reinforcement learning**

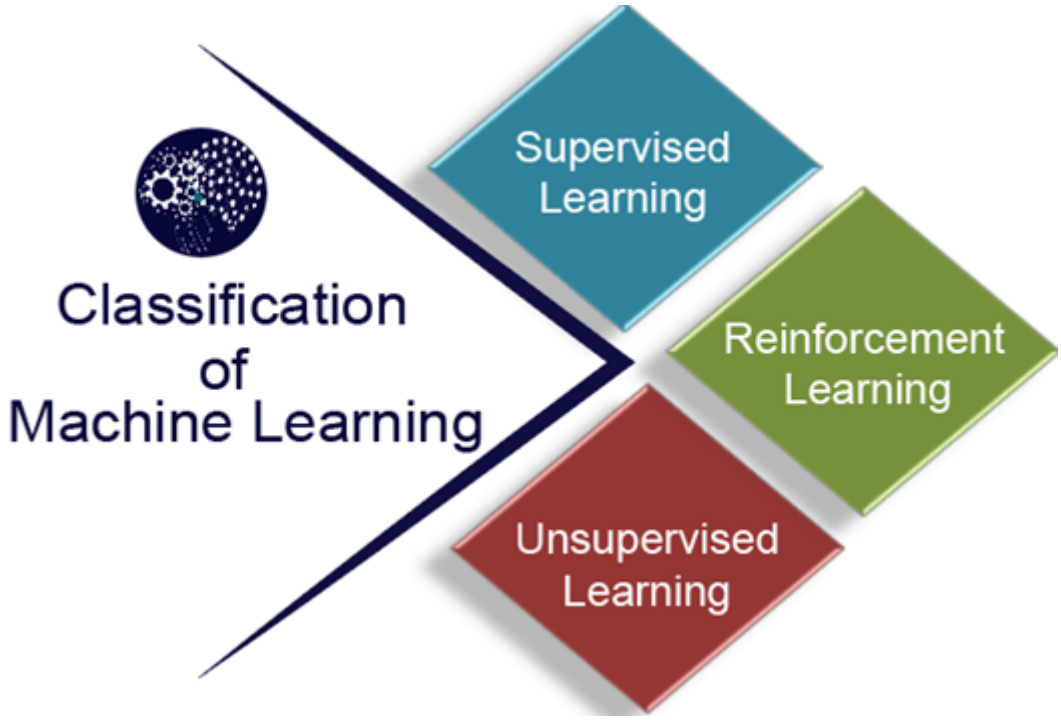

Figure 7: "Classification of Machine Learning"

1) Supervised Learning

Supervised machine learning is a type of machine learning in which we give labeled dataset input. The labeled dataset is trained in the ML model. The trained model is used to predict the output. The performance of the machine is also measured using the accuracy score.

The ML system creates a model using the labeled training dataset. The dataset information is recognized by the system. We can test the ML model by providing the test data to verify or check whether it forecasts the exact output or not.

The goal of supervised learning is to map input data with the output data. Supervised learning is based on supervision, and it is the same as when a student learns things under the supervision of the teacher. An example of supervised learning is spam filtering.

Supervised learning aims to map input data with the output data. Supervised learning is based on supervision, and it is the same as when a student learns things under the supervision of the teacher. Spam filtering is an example of a supervised type of machine learning.

Supervised learning can be grouped further in two categories of algorithms:

- Classification
- Regression

2) Unsupervised Learning

The unsupervised machine learning technique in which a machine learns from the unlabelled training data. The input data is unlabelled.

The machine gets trained from the input data. The input data hasn’t been labeled, classified, or categorized. The algorithms act on the input data without the help of any labeling. Unsupervised type of machine learning aims to restructure the input data into a new group of objects with similar patterns.

In unsupervised learning, we don't have a predetermined result. The machine tries to find useful insights from the huge amount of data. It can be further classifieds into two categories of algorithms:

The unsupervised type of machine learning tries to find useful information from the huge amount of input data. Unsupervised machine learning is categorizd into two types. They are:

- Clustering
- Association

7.2. GUI

Python programming offers various options for working with GUI(Graphical User Interface). Tkinter is the most used technique out of all GUI methods. Its standard Python interface to the GUI toolkit shipped with Python. Python with Tkinter is the fastest and easiest way to create GUI applications.

To create a Tkinter app:

1. First import module I,e Tkinter
2. Then create the main window.
3. Then add any number of widgets to the main window.
4. We can apply the event Trigger on the widgets.

The Tkinter importing is the same as importing any other module in Python programming. We have to note that the name of the module in Python 2.x is 'Tkinter' and in Python 3.x it is 'Tkinter.

```
import Tkinter
```

There are two main methods used in GUI. These methods the user needs to remember while creating the Python application with GUI.

1. **Tk():** For creating the main window, Tkinter offers a method called Tk(). The Tk() function is used to initiate the GUI using Python. We already have inbuilt support to python for GUI implementation. Just we need to import the Tkinter module.

2. **The main loop():** There is a method recognized by the name main loop() that is used when your application is ready to run. mainloop() is an infinite loop used to run the application, pause for an event to occur, and process the event as long as the window is not closed.

   ```
   m.mainloop()
   ```

   The main loop() method is used to run the window pause for an event. It helps to process an event unless and until we close the window. Hence it's the last line of GUI tkinter.

The Tkinter also offers access to the geometric configuration of the widgets which can organize the widgets in the parent windows. There are mainly three geometry manager classes class.

1. **pack() method:** This is a method that organizes the widgets in blocks before placing them in the parent widget.
2. **grid() method:** This is a method that organizes the widgets in a grid (table-like structure) before placing them in the parent widget.
3. **place() method:** This is a method that organizes the widgets by assigning them to specific positions focused by the programmer.

8. CHAPTER VIII

8.1 SUMMARY OF ACHIEVEMENTS

The project which we are implementing uses Machine Learning technology. Machine Learning is used for the prediction of red wine quality. The dataset is used for the applying of the machine. The dataset is having numeric values. The dataset is used for the preprocessing and building of the prediction model.

The dataset is having certain red wine-related parameters. Each parameter has a meaning which will be used for the model prediction. Thus the ML technique is using a CSV dataset for the red wine quality check. The proposed model used the red wine dataset for the prediction of the quality in terms of good or bad.

Step1. Importing the libraries

The Machine Learning libraries we are using for the project implementation and we apply for the project. The library is a reusable chunk of code that we use for calling the method.

Compared to languages like C++ or C, Python libraries do not pertain to any specific context in Python. Here, a 'library' loosely describes a collection of core modules. The library is installed with the python installs packages command.

The library is also known as the set of useful functions that eliminate the need for writing codes from scratch. We are using the libraries for the reusability in the next part of the code such as pandas. We use the pandas library mainly in this project.

Step2. Importing the dataset

A collection of instances is a dataset and when working with machine learning methods we typically need a few datasets for different purposes. The dataset we use is a CSV file that is a comma-separated values file. The dataset parameters are loaded using python and applied to the machine.

A single column of data is called a feature. It is a component of observation and is also called an attribute of a data instance. Some features may be inputs to a model (the predictors) and others may be outputs or the features to be predicted.

The dataset we use is a red wine dataset. The dataset is having the parameters which we consider for the project. The dataset uses the CSV format. The parameters help us to get the values from the CSV dataset.

Step3. Splitting dataset into Train and Test

In AI information pre-processing, we partition our dataset into a preparation set and test set. This is one of the pivotal strides of information pre-processing as by doing this, we can improve the presentation of our AI model.

Assume, on the off chance that we have offered to prepare our AI model by a dataset and we test it by a unique dataset. Then, at that point, it will make challenges for our model to comprehend the connections between the models.

If we train our model very well and its training accuracy is also very high, but we provide a new dataset to it, then it will decrease the performance. So we always try to make a machine learning model which performs well with the training set and also with the test dataset.

We are splitting the data into train and test. The whole dataset is split into train and test using the split functions of the Machine Learning technique.

Step4. Applying the algorithm

The Classification algorithm is a Supervised Learning technique that is used to identify the category of new observations based on training data. In Classification, a program learns from the given dataset or observations and then classifies new observations into many classes or groups.

We are using Naïve Bayes, SVM, Linear Regression in the project for the implementation of the classification process. The red wine dataset is split into train and test. Then the algorithm is used to apply the data for the classification of the dataset.

Step5. Prediction.

We are using the algorithms for the prediction. We used the Naïve Bayes, SVM, Linear Regression algorithms. The prediction is nothing but detecting the future output. Machine learning helps to predict future data.

We are passing the various dataset parameters in the project. The parameter values are passed for the splitting into train and test. The same is implemented in the algorithms. The algorithm finds the data in the dataset.

The red wine quality is predicted in this stage. It is good or bad. It depends on the dataset parameters. Hence Machine Learning technology is used for the prediction.

9. Chapter IX

9.1 DISCUSSION AND EVALUATION

We are using Machine Learning technology. We used four algorithms like Linear Regression, Naïve Bayes, Support Vector Machine, and Multilayer Perceptron. We achieved the project output using these algorithms. The prediction is done using the dataset.

The results are as follows:

Results:

1. Linear Regression

We used the Linear Regression algorithm for the implementation of the project. The algorithm was used for the classification of the red wine type. It's a supervised machine learning algorithm. The supervised machine learning algorithm has the labeled training dataset. The labeled training dataset is used for the classification of the red wine. The Linear Regression algorithm is used for the detection of the quality of the red wine. The algorithm produces an accuracy score. The accuracy score is the ratio of correct predictions to the test data input.

Hence the Linear Regression algorithm helped me to achieve the accuracy score below. Linear Regression achieves less accuracy score in this project compared to Linear Regression, Support Vector Machine, and Naive Bayes algorithms.

```
------------------LinearRegression------------------------------

Accuracy Score: 0.3283887639580224
```

Source: ***[Referred to Appendix 1]***

2. Support Vector Machine

We used the Support Vector Machine algorithm for the implementation of the project. The algorithm was used for the classification of the red wine type. It's a supervised machine learning algorithm. The supervised machine learning algorithm has the labeled training dataset. The labeled training dataset is used for the classification of the red wine. The Support Vector Machine algorithm is used for the detection of the quality of red wine. The algorithm produces an accuracy score. The accuracy score is the ratio of correct predictions to the test data input.

Hence the SVM algorithm helped me to achieve the accuracy score below.

```
------------------SVM---------------------------------
Accuracy Score: 0.64375
```

Source: ***[Referred to Appendix 1]***

3. Multi-Layer Perceptons

We used the MLP algorithm for the implementation of the project. The algorithm was used for the classification of the red wine type. It's a supervised machine learning algorithm. The supervised machine learning algorithm has the labeled training dataset. The labeled training dataset is used for the classification of the red wine. The MLP algorithm is used for the detection of the quality of red wine. The algorithm produces an accuracy score. The accuracy score is the ratio of correct predictions to the test data input

Hence the MLP algorithm helped me to achieve the accuracy score below.

```
------------------MLP-------------------------------

Accuracy Score: 0.596875
```

Source: ***[Referred to Appendix 1]***

4. Naïve Bayes

Naïve Bayes algorithm shows the accuracy.

We used the Naïve Bayes algorithm for the implementation of the project. The algorithm was used for the classification of the red wine type. It's a supervised machine learning algorithm. The supervised machine learning algorithm has the labeled training dataset. The labeled training dataset is used for the classification of the red wine. The Naïve Bayes algorithm is used for the detection of the quality of red wine. The algorithm produces an accuracy score. The accuracy score is the ratio of correct predictions to the test data input

Hence this algorithm helped me to achieve the accuracy score below.

```
------------------NaiveBayes-------------------------------

Accuracy Score: 0.540625
```

Source: ***[Referred to Appendix 1]***

5. Bar Graph

Below is the bar graph output of this project. It uses the Matplotlib library for building the project.

a. Fixed Acidity vs Volatile Acidity :

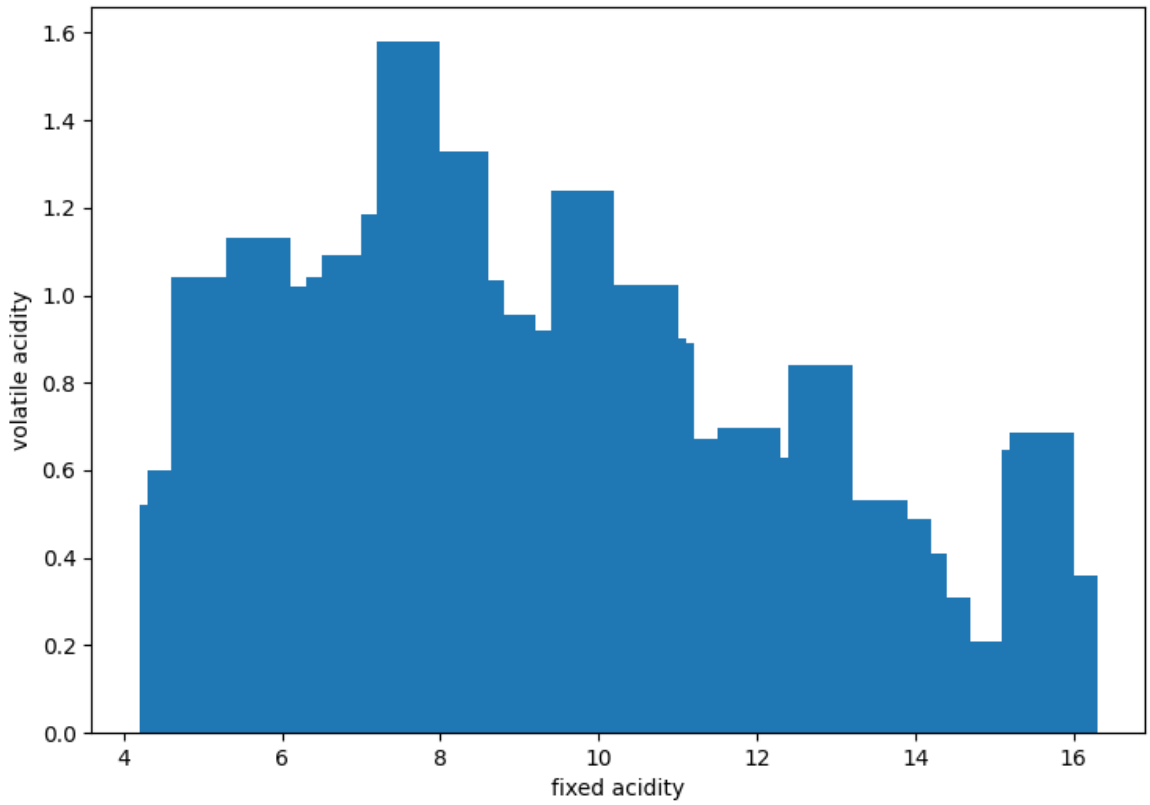

Figure 8.a : Bar Graph

Source: ***[Referred to Appendix 1]***

b. Alcohol vs Quality

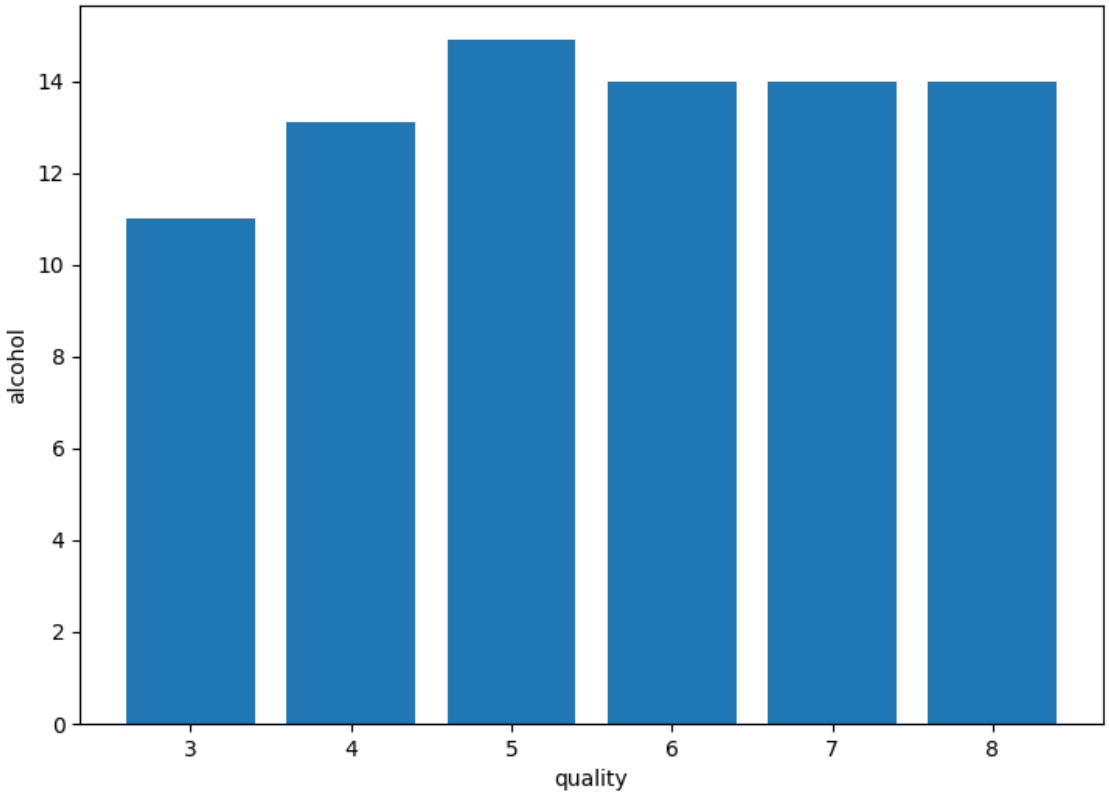

Figure 8.b : Bar Graph

Source: ***[Referred to Appendix 1]***

6. Prediction

Below is the prediction of certain parameters input by the user.

```
----------Please enter the details below:-----------


fixed acidity:7.4
volatile acidity:0.7
citric acid:0.0
residual sugar:1.9
chlorides:0.076
free sulfur dioxide:11
total sulfur dioxide:34
density:0.9978
pH:3.51
sulphates:0.56
alcohol:9.4

------------Prediction started------------
Red Wine Quality is Bad
```

10. CONCLUSION

We developed the project using Linear Regression, Naïve Bayes, SVM, and MLP algorithms. The interest has been increased in the wine industry in recent years which demands growth in this industry. Therefore, companies are investing in new technologies to improve wine production and selling. The algorithms help us to determine import features and for the prediction of the red wine quality. The ML techniques have a better learning capability on the input dataset. Thus finally accuracy score is displayed. In the future, a large dataset can be taken for experiments and other machine learning techniques may be explored for wine quality prediction.

11. BIBLIOGRAPHY AND REFERENCING

11.1.BIBLIOGRAPHY

DATASET: The dataset is used from the below website for the implementation of the project.

Link: https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009

11.2. REFERENCES

[1] P. Cortez, A. Cerderia, F. Almeida, T. Matos, and J. Reis, "Modelling wine preferences by data mining from physicochemical properties," *In Decision Support Systems, Elsevier*, 47 (4): 547-553. ISSN: 0167-9236.

[2] S. Ebeler, "Linking Flavour Chemistry to Sensory Analysis of Wine," in *Flavor Chemistry*, Thirty Years of Progress, Kluwer Academic Publishers, 1999, pp. 409-422.

[3] V. Preedy, and M. L. R. Mendez, "Wine Applications with ElectronicNoses," in *Electronic Noses and Tongues in Food Science*, Cambridge, MA, USA: Academic Press, 2016, pp. 137-151.

[4] A. Asuncion, and D. Newman (2007), UCI Machine Learning Repository, University of California, Irvine, [Online]. Available: http://www.ics.uci.edu/~mlearn/MLRepository.html

[5] S. Kallithraka, IS. Arvanitoyannis, P. Kefalas, A. El-Zajouli, E. Soufleros, and E. Psarra, "Instrumental and sensory analysis of Greek wines; implementation of principal component analysis (PCA) for classification according to geographical origin," Food Chemistry, 73(4): 501-514, 2001.

[6] N. H. Beltran, M. A. Duarte- MErmound, V. A. S. Vicencio, S. A. Salah, and M. A. Bustos, "Chilean wine classification using volatile organic compounds data obtained with a fast GC analyzer," Instrum. Measurement,IEEE Trans., 57: 2421-2436, 2008.

[7] S. Shanmuganathan, P. Sallis, and A. Narayanan, "Data mining techniques
for modeling seasonal climate effects on grapevine yield and wine quality," IEEE International Conference on Computational Intelligence Communication Systems and Networks, pp. 82-89, July 2010.

[8] B. Chen, C. Rhodes, A. Crawford, and L. Hambuchen, “Wineinformatics: applying data mining on wine sensory reviews processed by the computational wine wheel,” IEEE International Conference on Data Mining Workshop, pp. 142-149, Dec. 2014.

[9] K. Agrawal and H. Mohan, "Cardiotocography Analysis for Fetal State Classification Using Machine Learning Algorithms," 2019 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, Tamil Nadu, India, 2019, pp. 1-6.

[10] K. Agrawal and H. Mohan, "Text Analysis: Techniques, Applications, and Challenges," presented in 2019 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, Tamil Nadu, India.

12.APPENDICES

#IMPORTING THE REQUIRED LIBRARIES

```
from tkinter import *
import tkinter.messagebox
##import mysql.connector
from PIL import Image,ImageTk
```

#Creating the GUI form

```
def MAIN():
  R1=Tk()
  R1.geometry('800x800')
  R1.title('WELCOME-2')
  image=Image.open('a2.png')
  image=image.resize((900,700))
  photo_image=ImageTk.PhotoImage(image)
  l=Label(R1, image=photo_image)
  l.place(x=0, y=0)


  l=Label(R1, text="Red Wine Quality Prediction", font=('algerain',18,'bold'))
  l.place(x=220, y=40)
b1=Button(R1, text="Linear Regression",width=18,height=2,font=('algerain',14),bg="lightblue",
fg="red", command=LinearRegression)
  b1.place(x=250, y=120)
 b2=Button(R1, text="SVM",width=18,height=2,font=('algerain',14),bg="lightblue", fg="red",
command=SVM)
  b2.place(x=250, y=220)
b3=Button(R1, text="Naive Bayes",width=18,height=2,font=('algerain',14),bg="lightblue",
fg="red", command=NaiveBayes)
  b3.place(x=250, y=320)
```

```
b4=Button(R1,  text="MLP",width=18,height=2,font=('algerain',14),bg="lightblue",  fg="red",
command=MLP)
  b4.place(x=250, y=420)

  b5=Button(R1,  text="Bargraph",width=18,height=2,font=('algerain',14),bg="lightblue",  fg="red",
command=Bargraph)
  b5.place(x=250, y=520)
  b6=Button(R1, text="Prediction",width=18,height=2,font=('algerain',14),bg="lightblue", fg="red",
command=Prediction)
  b6.place(x=250, y=620)
  R1.mainloop()
```

#IMPLEMENTING LINEAR REGRESSION ALGORITHM

```
def LinearRegression():
  print('\n\n------------------LinearRegression----------------------------\n')
  import pandas as pd
  import warnings
  warnings.filterwarnings('ignore')
  #Import dataset
  dataset = pd.read_csv('winequality-red.csv', encoding='latin1')
  X = dataset.iloc[:, [0,1,2,3,4,5,6,7,8,9,10]].values
  y = dataset.iloc[:, 11].values
  #Split train and test data
  from sklearn.model_selection import train_test_split
  X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20, random_state = 0)
  from sklearn.preprocessing import StandardScaler
  sc = StandardScaler()
  X_train = sc.fit_transform(X_train)
  X_test = sc.transform(X_test)
  #Apply algorithm
  from sklearn.linear_model import LinearRegression
  classifier =  LinearRegression()
  classifier.fit(X_train, y_train)
```

```
  #accuracy score
  from sklearn.metrics import accuracy_score
  A = classifier.score(X_test,y_test)
  print('Accuracy Score: {}\n'.format(A))
```

#IMPLEMENTING SVM ALGORITHM

```
def SVM():
  print('\n\n----------SVM----------\n')
  import pandas as pd
  import warnings
  warnings.filterwarnings('ignore')
  #Import dataset
  dataset = pd.read_csv('winequality-red.csv', encoding='latin1')
  X = dataset.iloc[:, [0,1,2,3,4,5,6,7,8,9,10]].values
  y = dataset.iloc[:, 11].values
  #Split train and test data
  from sklearn.model_selection import train_test_split
  X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20, random_state = 0)
  from sklearn.preprocessing import StandardScaler
  sc = StandardScaler()
  X_train = sc.fit_transform(X_train)
  X_test = sc.transform(X_test)
  #Apply algorithm
  from sklearn.svm import SVC
  classifier =  SVC()
  classifier.fit(X_train, y_train)
  y_pred = classifier.predict(X_test)
  #Accuracy score
  from sklearn.metrics import accuracy_score
  A = accuracy_score(y_test,y_pred)
  print('Accuracy Score: {}\n'.format(A))
```

#IMPLEMENTING NAIVEBAYES ALGORITHM

```
def NaiveBayes():
  print('\n\n----------NaiveBayes----------\n')
  import pandas as pd
  import warnings
  warnings.filterwarnings('ignore')
  #Import dataset
  dataset = pd.read_csv('winequality-red.csv', encoding='latin1')
  X = dataset.iloc[:, [0,1,2,3,4,5,6,7,8,9,10]].values
  y = dataset.iloc[:, 11].values
  #Split train and test data
  from sklearn.model_selection import train_test_split
  X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20, random_state = 0)
  from sklearn.preprocessing import StandardScaler
  sc = StandardScaler()
  X_train = sc.fit_transform(X_train)
  X_test = sc.transform(X_test)
  #Apply algorithm
  from sklearn.naive_bayes import GaussianNB
  classifier =  GaussianNB()
  classifier.fit(X_train, y_train)
  y_pred = classifier.predict(X_test)
  #Accuracy score
  from sklearn.metrics import accuracy_score
  A = accuracy_score(y_test,y_pred)
  print('Accuracy Score: {}\n'.format(A))
```

#IMPLEMENTING MLP ALGORITHM

```
def MLP():
  print('\n\n------------------MLP-----------------------------\n')
  import pandas as pd
  import warnings
  warnings.filterwarnings('ignore')
  #Import dataset
  dataset = pd.read_csv('winequality-red.csv', encoding='latin1')
  X = dataset.iloc[:, [0,1,2,3,4,5,6,7,8,9,10]].values
  y = dataset.iloc[:, 11].values
  #Split train and test data
  from sklearn.model_selection import train_test_split
  X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20, random_state = 0)
  from sklearn.preprocessing import StandardScaler
  sc = StandardScaler()
  X_train = sc.fit_transform(X_train)
  X_test = sc.transform(X_test)
  #Apply algorithm
  from sklearn.neural_network import MLPClassifier
  classifier =  MLPClassifier()
  classifier.fit(X_train, y_train)
  y_pred = classifier.predict(X_test)
  #Accuracy score
  from sklearn.metrics import accuracy_score
  A = accuracy_score(y_test,y_pred)
  print('Accuracy Score: {}\n'.format(A))
```

```
-----------------------------------------------------------------------------------------------------------
#IMPLEMENTING BARGRAPH ALGORITHM
-----------------------------------------------------------------------------------------------------------
 def Bargraph():
    print('\n------------Plotting Bargraph------------')
    import pandas as pd
    import matplotlib.pyplot as plt
    dataset = pd.read_csv('winequality-red.csv', encoding='latin1')
    plt.figure(figsize=[15,6])
   #Input Values
    a1 = str(input('enter for x-axis:'))
    a2 = str(input('enter for y-axis:'))
    plt.bar(dataset[a1],dataset[a2])
    plt.xlabel(a1)
    plt.ylabel(a2)
    plt.show()
-----------------------------------------------------------------------------------------------------------
#IMPLEMENTING PREDICTION ALGORITHM
-----------------------------------------------------------------------------------------------------------
def Prediction():
  import numpy as np
  import matplotlib.pyplot as plt
  import pandas as pd
  import warnings
  warnings.filterwarnings('ignore')
  #Import dataset
  dataset = pd.read_csv('winequality-red.csv', encoding='latin1')
  X = dataset.iloc[:, [0,1,2,3,4,5,6,7,8,9,10]].values
  y = dataset.iloc[:, 11].values
  #Split train and test data
```

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20, random_state = 0)


# Feature Scaling
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)
#Apply algorithm
from sklearn.naive_bayes import GaussianNB
classifier =  GaussianNB()
classifier.fit(X_train, y_train)
# Predicting the Test set results
y_pred = classifier.predict(X_test)
#quality > 6.5 => "good"
print('\n----------Please enter the details below:----------\n')
a = float(input('\nfixed acidity:'))
b = float(input('volatile acidity:'))
c = float(input('citric acid:'))
d = float(input('residual sugar:'))
e = float(input('chlorides:'))
f = int(input('free sulfur dioxide:'))
g = int(input('total sulfur dioxide:'))
h = float(input('density:'))
i = float(input('pH:'))
j = float(input('sulphates:'))
k = float(input('alcohol:'))
print('\n------------Prediction started------------')
z = classifier.predict([[a,b,c,d,e,f,g,h,i,j,k]])
print(z)
if z>=5 and z<=6.5:
  print('Red Wine Quality is Bad')
elif z>6.5 and z<=10:
```

```
    print('Red Wine Quality is Good')
MAIN()
```

Appendix 1: List of few attributes of data for Red wine consists of statistical values from which Dataset is created.

Fixed acidity	Volatile acidity	Citric acid	Residual sugar	Chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
7.4	0.7	0	.19	0.076	11	34	0.9978	2	0.56	9.4	5
7.8	0.88	0	2.6	0.098	25	67	0.9968	1	0.68	9.8	5
7.8	0.76	0.04	2.3	0.092	15	54	0.997	2	0.65	9.8	5
11.2	0.28	0.56	1.9	0.075	17	60	0.998	3	0.65	9.8	5