### A

**GROUP PROJECT REPORT**

**ON**

.**WINE QUALITY PREDICTION.**

***Dissertation Submitted In Partial Fulfillment Of The Requirement For The Award Of***

### BACHELOR OF TECHNOLOGY IN

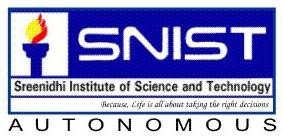
**INFORMATION TECHNOLOGY**

**By**

**Janagama ShivaCharan (18311A12D9)**

**Bommena Nithin Kumar (18311A12D2)**

**Chanda Sai Charan (18311A12D3)**



**DEPARTMENT OF INFORMATION TECHNOLOGY SREENIDHI INSTITUTION OF SCIENCE & TECHNOLOGY**

**(An Autonomous Institution) (AFFILIATED TO JNTU, HYDERABAD)**

**Yamnampet, Ghatkesar, R.R. Dist, Hyd-501301.**

**2021**

**WINE QUALITY PREDICTION**

***Dissertation submitted in partial fulfillment of the Requirement for the award of***

### BACHELOR OF TECHNOLOGY IN

**INFORMATION TECHNOLOGY**

**By**

### Janagama ShivaCharan(18311A12D9)

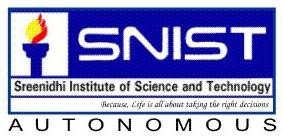
### Bommena Nithin Kumar (18311A12D2)

### Chanda Sai Charan (18311A12D3)

**Under the guidance of**

**Mr.R.Kameshwar Reddy**

Associate Professor, Dept of IT



**DEPARTMENT OF INFORMATION TECHNOLOGY SREENIDHI INSTITUTION OF SCIENCE AND TECHNOLOGY**

### (An Autonomous Institution) (AFFILIATED TO JNTU, HYDERABAD)

**Yamnampet, Ghatkesar, R.R. Dist, Hyd-501301**

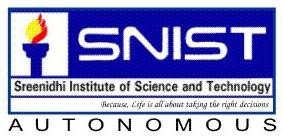
**2021**

### DEPARTMENT OF INFORMATION TECHNOLOGY

**SREENIDHI INSTITUTION OF SCIENCE AND TECHNOLOGY**

### (An Autonomous Institution)

**(Affiliated to JNT University; ISO Certified 9001:2000) Yamnampet, Ghatkesar, Hyderabad-501301**

**Ph.Nos:08415-200597, 08415-325444, 9395533303**

### 

### CERTIFICATE

This is to certify that the Dissertation entitled **“WINE QUALITY PREDICTION”** is bonafide workdone and submitted by **Janagama ShivaCharan (18311A12D9) , Bommena Nithin Kumar(18311A12D2) ,Chanda Sai Charan(18311A12D3) ,** in partial fulfillment of the requirement for the award of Degree of **Bachelor of Technology in Information Technology, SREENIDHI INSTITUTE OF SCIENCE AND TECHNOLOGY, Affiliated to Jawaharlal Nehru Technological University, Hyderabad** is a record of bonafide work carried out by us under the guidance and supervision from **May 2021** to **Sep**  **2021.**

The results presented in this dissertation have been verified and are found to be satisfactory. The results embodied in this dissertation have not been submitted to any other university for the award of any other degree or diploma**.**

**Project Internal Guide Head of the Department Mr.R.KAMESHWAR REDDY Dr. V.V.S.S BALRAM.**

Associate Professor, IT Dept, Department of IT, SNIST, Hyderabad SNIST, Hyderabad

.

**Project Co-ordinator Mr. P. SREEDHAR**

Associate Professor & Associate Head Department of IT, SNIST.

### ACKNOWLEDGEMENT

I thank **Dr. V.V.S.S BALARAM** , **Head of the Department of IT** and **Mr. R.Kameshwar Reddy,** for supporting us as our **Internal Guide** and providing seamless knowledge over the past one year, and also for providing right suggestion at every phase of the development of our project.

I extend my profound gratitude to thank **Mr. P. Sreedhar** , **(Project Cordinator)** , Department of IT, SNIST, for his valuable suggestion and support throughout the course.

I express a whole hearted gratitude to **Dr. P. Narsimha Reddy, Executive Director,** and **Dr. Shiva Reddy , Principal, Sreenidhi Institute of Science and Technology** for providing us the conducive environment for carrying through our academic schedules and projects with ease.

I convey my heartful thanks to my Parents, friends, Technical and Non-Technical staff of the college for their constant support in the successful completion of the project.

\

### Janagama ShivaCharan(18311A12D9)

### Bommena Nithin Kumar(18311A12D2)

### Chanda Sai Charan(18311A12D3)

**WINE QUALITY PREDICTION**

# ABSTRACT

In recent years, most of the industries are promoting their products based on the quality certification they received on the products. The traditional way of assessing the product quality is time consuming, however with the invent of machine learning techniques the processes has become more efficient and consumed less time than before. Our project explores the usage of machine learning techniques such as linear regression, neural network and support vector machine for product quality in two ways. Firstly, determine the dependency of target variable on independent variables and secondly, predicting the value of target variable. On the basis of computed dependency, important variables are selected those make significant impact on dependent variable. Further, neural network and support vector machine are used to predict the values of dependent variable. All the experiments are performed on Red Wine and White Wine datasets.

### WINE QUALITY PREDICTION

## Preface TableofContents

### 1.INTRODUCTION

* 1. INTRODUCTION TO PROJECT
  2. DATASET DESCRIPTION
  3. PURPOSE OF THE PROJECT
  4. EXISTING SYSTEM
  5. PROPOSED SYSTEM

### 2.SYSTEM ANALYSIS

* 1. FUNCTIONAL REQUIREMENTS
  2. SOFTWARE REQUIREMENTS
  3. HARDWARE REQUIREMENTS

### 3. SYSTEM DESIGN

* 1. ARCHITECTURE DESIGN

3.2 FLOWCHART

**4.SYSTEM IMPLEMENTATION**

**4.1** DATASET

**4.2** DATA PREPOCESSING

**4.3** RANDOM FOREST

**4.4** DECISION TREE

**5.OUTPUT SCREENSHOT**

**6.CONCLUSION**

**7.BIBLIOGRAPHY**

**WINE QUALITY PREDICTION**

### INTRODUCTION TO PROJECT

The quality of the wine is a very important part for the consumers as well as the manufacturing industries. Industries are increasing their sales using product quality certification. Nowadays, all over the world wine is a regularly used beverage and the industries are using the certification of product quality to increases their value in the market. Previously, testing of product quality will be done at the end of the production, this is time taking process and it requires a lot of resources such as the need for various human experts for the assessment of product quality which makes this process very expensive. Every human has their own opinion about the test, so identifying the quality of the wine based on humans experts it is a challenging task. Here we will predict the quality of wine on the basis of given 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 deal with both the white and red wine type quality, we use classification techniques to check further the quality of the wine i.e. is it good or bad. First of all, we need to install a bunch of packages that would come handy in the construction and execution of our code numpy will be used for making the mathematical calculations more accurate, pandas will be used to work with file formats like csv, xls etc. and sklearn (scikit-learn) will be used to import our classifier for prediction.

**WINE QUALITY PREDICTION**

**1.2DATASETDESCRIPTION**

The dataset providers alleges that due to privacy and logistic issues, only physicochemical and sensory variables are available.There’s no information about how the dataset was created, such as the distribution of grape types, wine brands and whether the same experts graded all wines. This information is important to identify possible bias that may distort the analysis results. For example, the physicochemical properties may vary among different wines of the same type, affecting the distribution in the dataset.The classes are ordered and not balanced. There are much more normal wines than excellent or poor ones.

Input variables based on physicochemical tests:

1.Fixed Acidity

2.Volatile Acidity

3.Citric Acid

4.Residual Sugar

5.Chlorides

6.Free Sulfur Dioxide

7.Total Sulfur Dioxide

8.Density

9.pH

10.Sulphates

11.Alcohol

Output variable based on sensory data:

12.Quality (score between 0 and 10)

**WINE QUALITY PREDICTION**

### 1.3 PURPOSE OF THE PROJECT

Wine quality prediction is very helpful for companies that produce red and white wine and also for the customers. Customers can refer to the quality of different wines present in the existing market and get to a decision of choosing the best wine possible according to their requirements.

**1.4 EXISTING SYSTEM**

Till now the quality of wine is based on conclusions made by any organization or a group of people around the world. This consumes a lot of time because there are various wines made by many companies across the world and it is definitely not possible to compare all of them. Comparing means the case where an organization classifies wines in an order based on it’s quality by taking into consideration pH,amount of sulfer dioxide etc. this becomes impossible in today’s world where there are lots of wines.Hence our project focuses on making it easier.

**WINE QUALITY PREDICTION**

**1.5 PROPOSED SYSTEM**

**Wine quality prediction is very important to predict the quality of wine for industries to promote their products and also good quality of wine is helpful to it’s consumers in maintaining good health. Our aim will be to train a variety of supervised learning algorithms on the data so that when a new data point arrives, our best performing classifier can be used to categorize the data point as a good or bad quality of wine. In problems of wine classification simply comparing the accuracy, that is ratio of correct predictions to total predictions is not enough. This is because depending on the context sometimes it is more important than algorithm doesn’t wrongly predict a good quality wine as bad quality.**

**Thus, here we will use confusion matrix which is basicallt the weighted harmonic mean of precision and recall.**

**WINE QUALITY PREDICTION**

**2.SYSTEM ANALYSIS**

System analysis is related to Requirement analysis. The Requirements defines the functional, non-functional, and technical requirements. Entire process is divided into steps to analyses the situation and analyses the project goals. Requirements documentation takes place in requirements analysis phase.

**2.1 FUNCTIONAL REQUIREMENTS**

**After careful analysis of the documentation the following modules are identified.**

**1.Administrator module**

The role of administrator is to run the python file in

jupyter notebook

**2.User module**

The role of the user is to submit the values for different factors

in order to retrieve the predicted value.

**WINE QUALITY PREDICTION**

**2.2 SOFTWARE REQUIREMENTS**

•**Operating system**: Windows7 and above.

•**Technology:** Anaconda navigator(jupyter notebook)

•**Software libraries**: Pandas, Seaborn, Numpy, Matplotlib,

Scikit-learn.

**2.3 HARDWARE REQUIREMENTS**

**• Processor:** INTEL i3 7th gen or higher

**• RAM:** 4GB or above

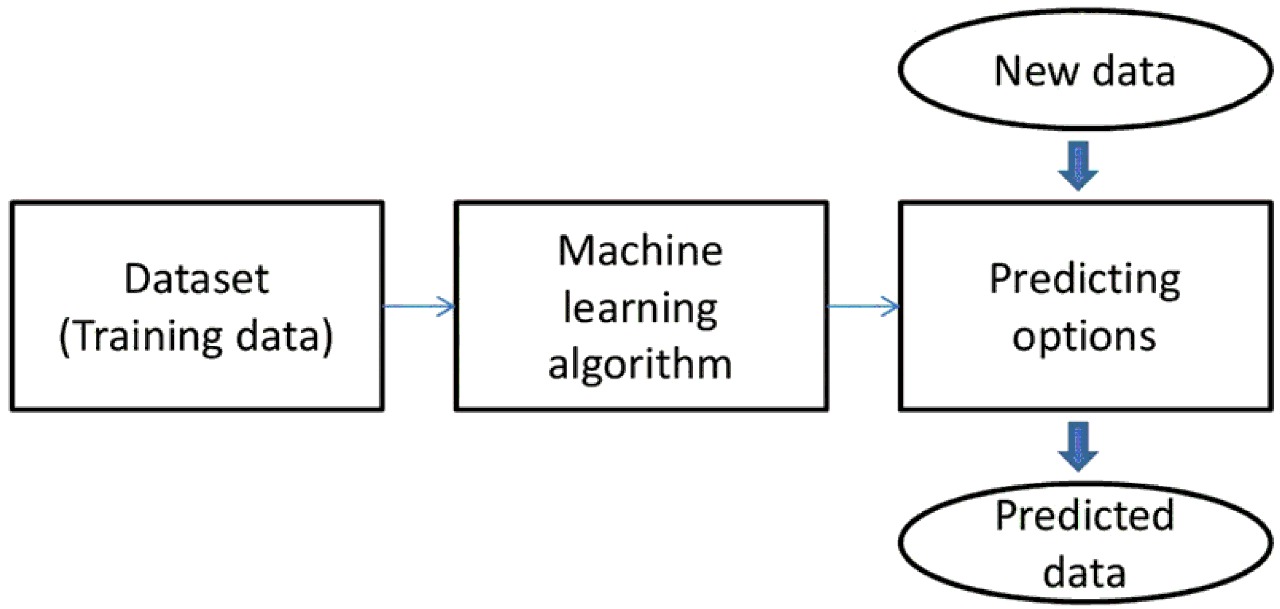
**• Hard disk:** 10GB

**WINE QUALITY PREDICTION**

**3.SYSTEM DESIGN**

Systems design is the process of defining the architecture, components, modules, interfaces, and data for a system to satisfy specified requirements. One could see it as the application of systems theory to product development. Object-oriented analysis and design methods are becoming the most widely used methods for computer systems design.

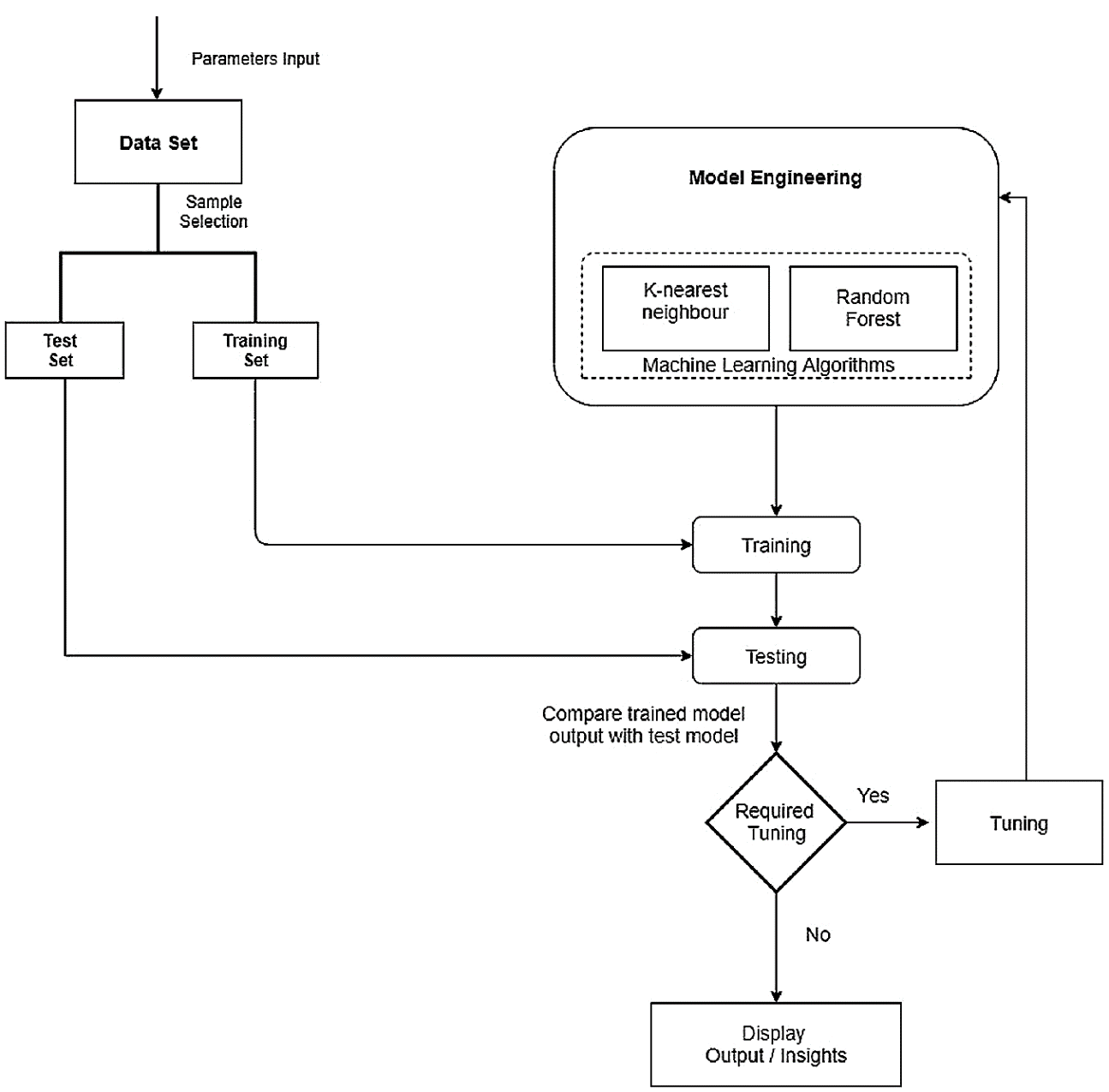
**3.1 ARCHITECTURE DESIGN**

****

A system architecture is the conceptual model that defines the structure, behaviour and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviour' s of the system. A system architecture is the conceptual model that defines the structure, behaviour and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviour's of the system

**WINE QUALITY PREDICTION**

* 1. **FLOWCHART**



A flowchart is simply a graphical representation of steps. It shows steps in sequential order and is widely used in presenting the flow of algorithms, workflow or processes. Typically, a flowchart shows the steps as boxes of various kinds, and their order by connecting them with arrows.

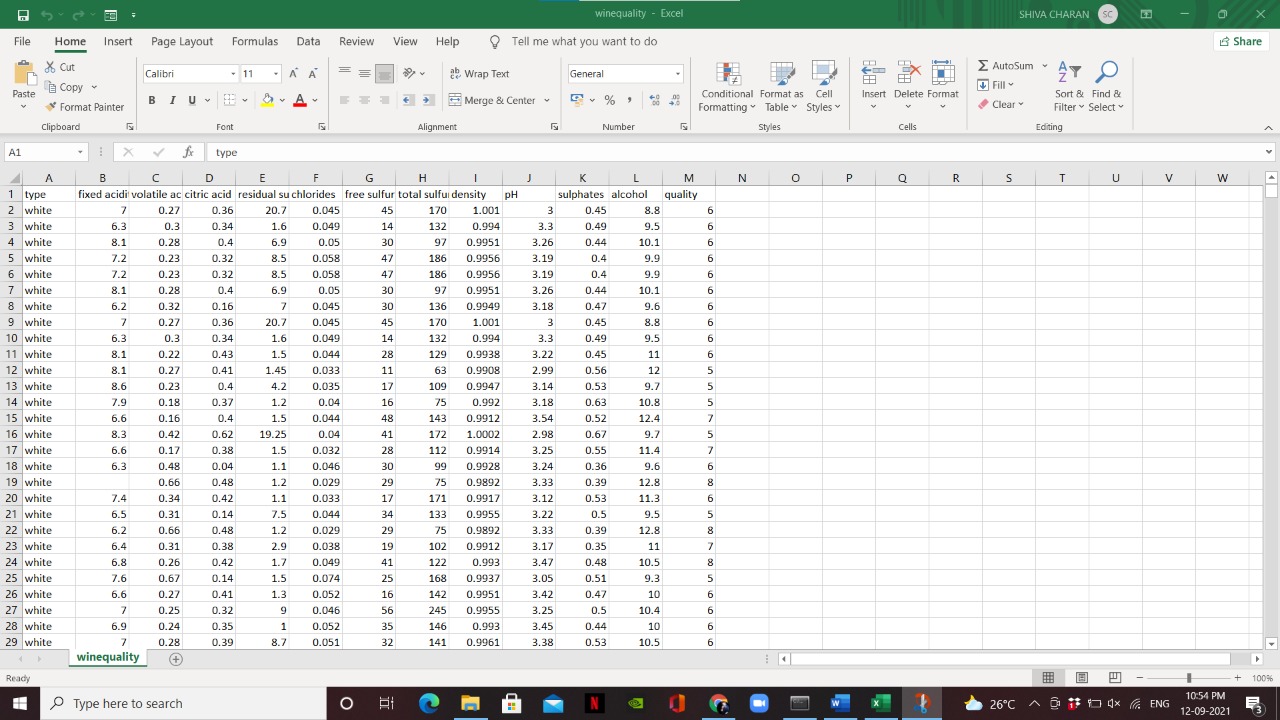
**WINE QUALITY PREDICTION**

**4.SYSTEM IMPLEMENTATION**

The implementation stage of any project is a true display of the defining moments that make a project a success or a failure. The implementation stage is defined as the system or system modifications being installed and made operational in a production environment. The phase is initiated after the system has been tested and accepted by the user. This phase continues until the system is operating in production in accordance with the defined user requirements.

**4.1 DATASET**

The data for our study was accessed from Kaggle Machine Learning Repository. The data set includes figures on 6496 different records and 12 attributes. In this dataset, the target variable is quality.

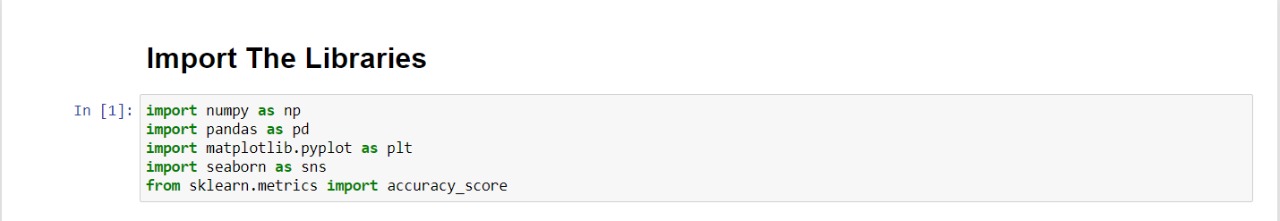


**WINE QUALITY PREDICTION**

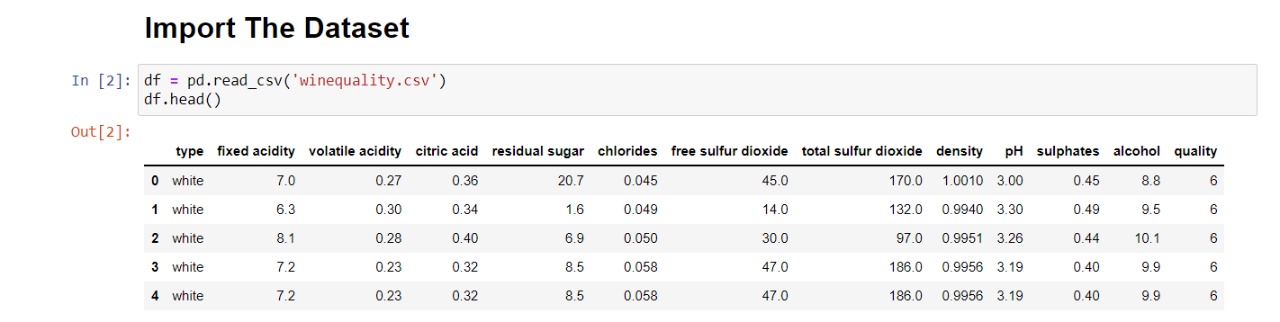
**4.2 DATA PREPROCESSING**

Before processing the Adult Dataset, cleaning the data with certain pre processing techniques becomes a necessity. This includes: 1) Handling Missing Values: The dataset contains certain set of missing values which has been dealt by replacing it with the mean of all values.

**IMPORTING LIBRARIES**

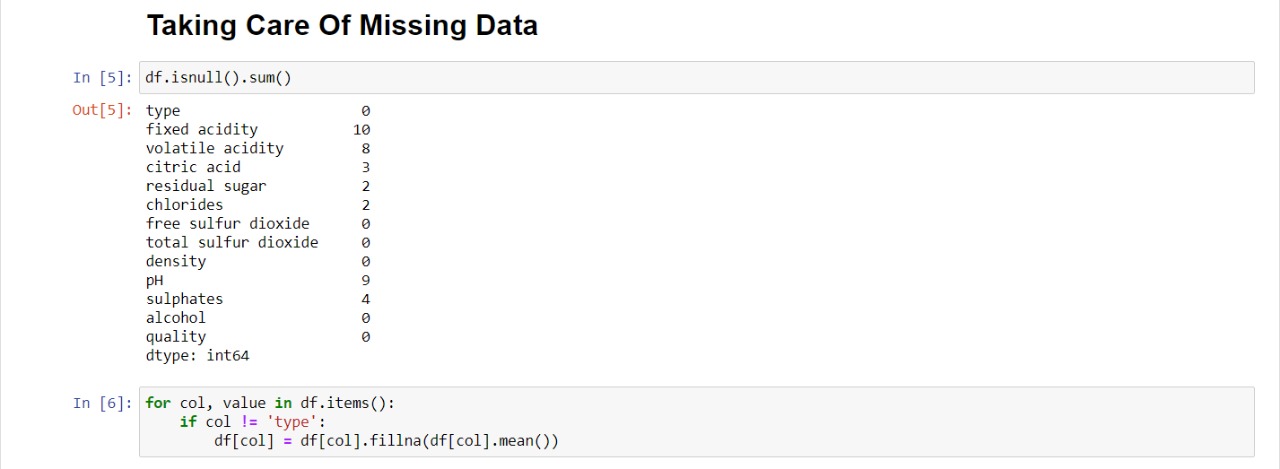
****

**IMPORTING DATASET**

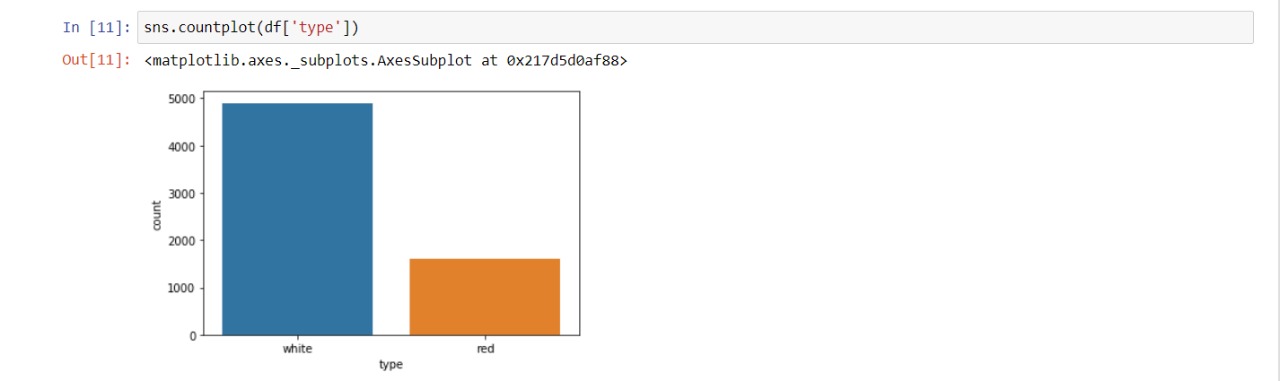
****

**WINE QUALITY PREDICTION**

**HANDLING MISSING VALUES**

****

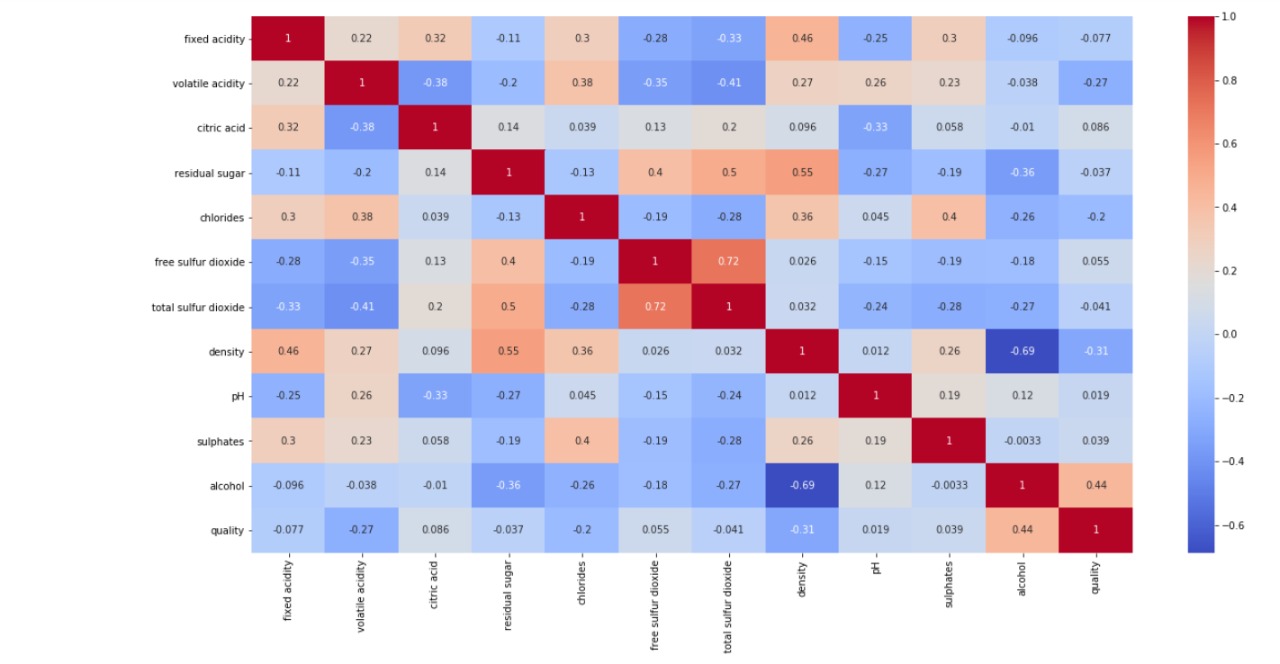
**WHITE AND RED WINE RECORDS COUNT**

****

**WINE QUALITY PREDICTION**

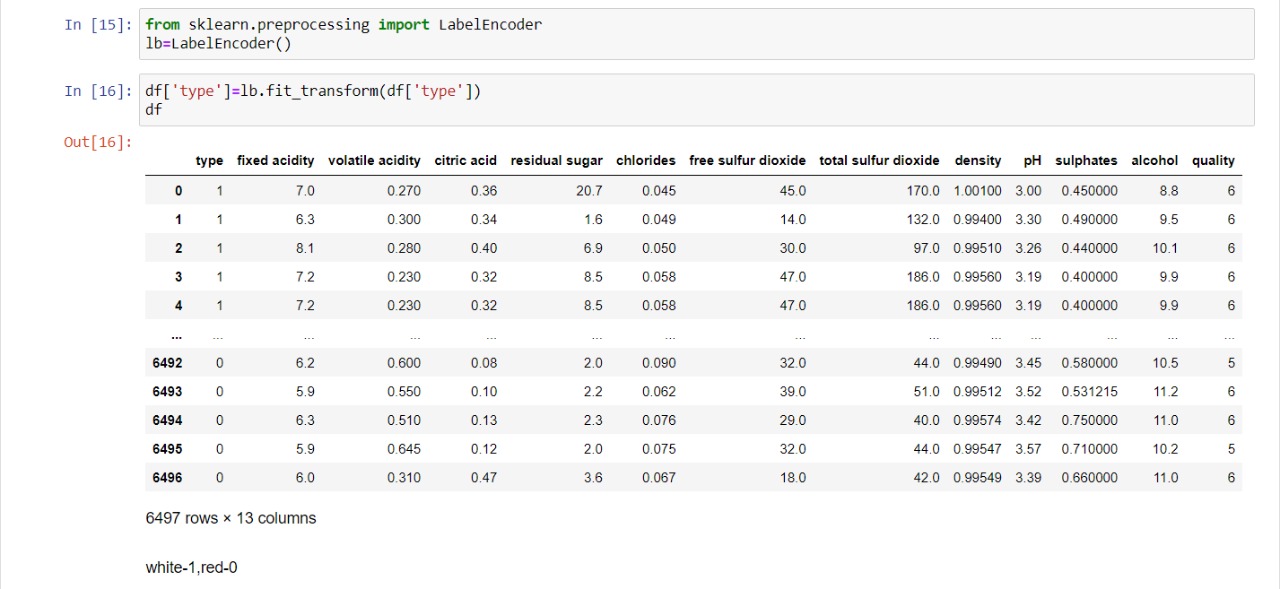
**CORRELATION MATRIX**

****

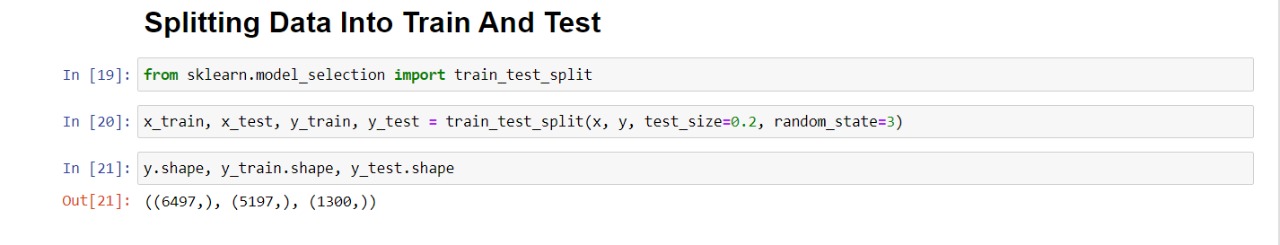


**WINE QUALITY PREDICTION**

**LABEL ENCODING**

****

**SPLITTING DATA INTO TRAIN AND TEST**

****

**WINE QUALITY PREDICTION**

**4.3 RANDOM FOREST**



***“*Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.”** Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

**The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.**

**4.4 DECISION TREE**



**WINE QUALITY PREDICTION**

* Decision Tree is a Supervised learning techniquethat can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, whereinternal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.
* In a Decision tree, there are two nodes, which are the Decision Node andLeafNode**.** Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.
* The decisions or the test are performed on the basis of features of the given dataset.

**SVM**

****

The support vector machine (SVM) is the most popular and most

widely used machine learning algorithm. It is a supervised learning

model that can perform classification and regression tasks. However,

it is primarily used for classification problems in machine learning.

**WINE QUALITY PREDICTION**

**KNN**



* K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.
* K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
* K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.

**LOGISTIC REGRESSION**



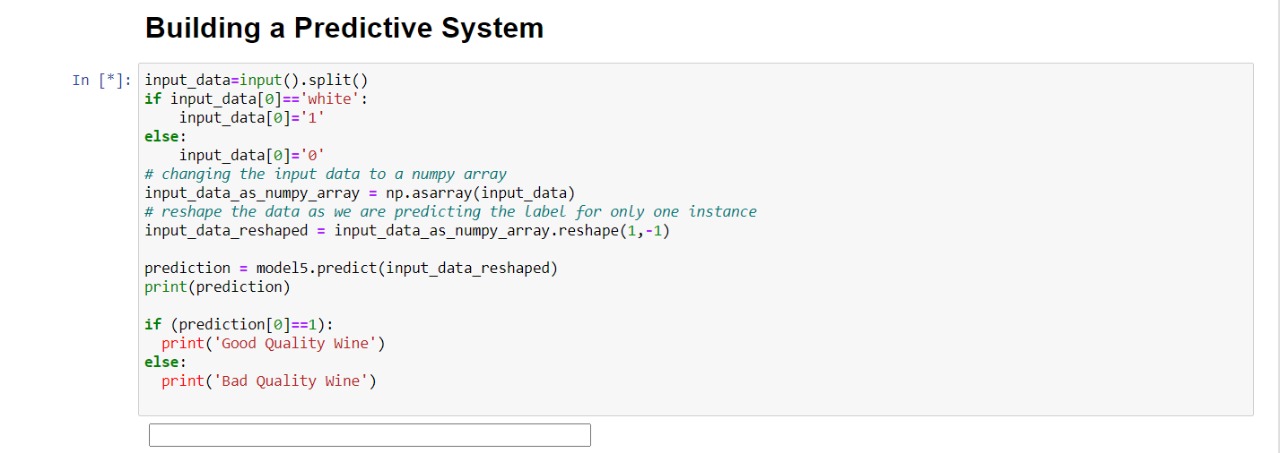
**WINE QUALITY PREDICTION**

Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes.

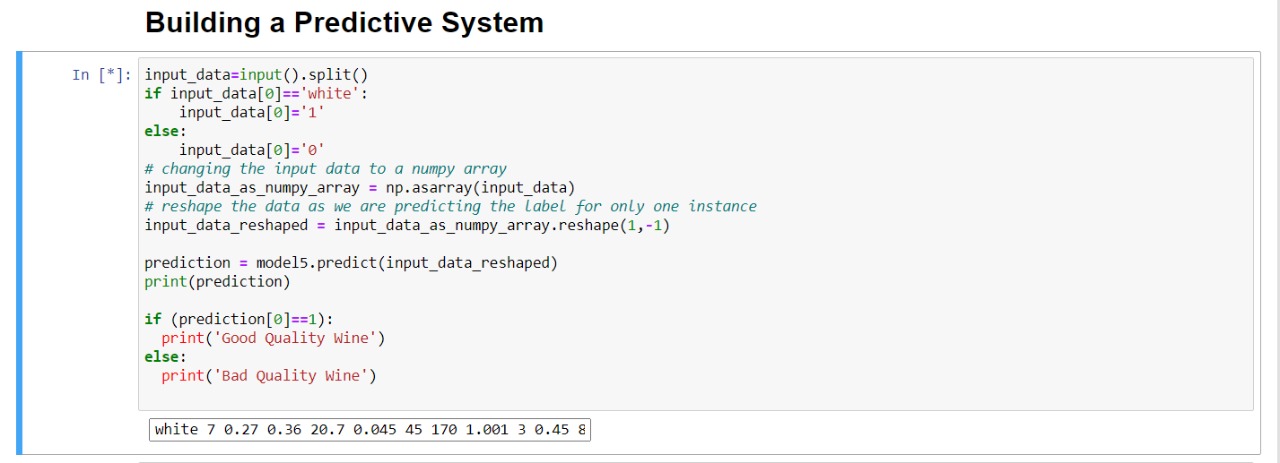
In simple words, the dependent variable is binary in nature having data coded as either 1 (stands for success/yes) or 0 (stands for failure/no).

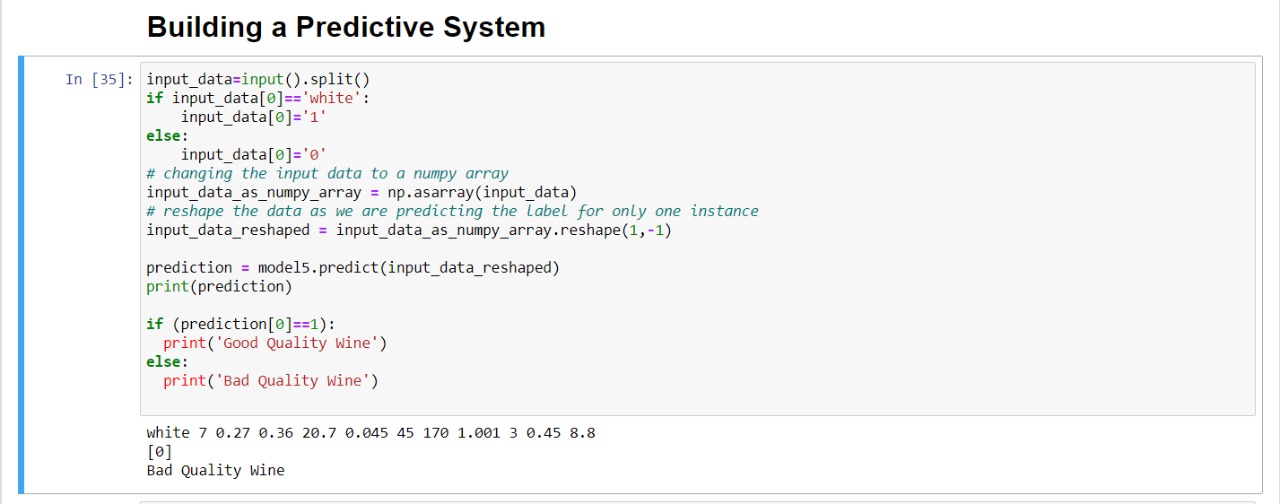
**5.OUTPUT SCREENSHOTS**

From the above results produced by all the mentioned machine learning algorithms we choose the algorithm that has the most accuracy in order to classify the wine into a good or bad quality wine. Random forest produces most accuracy, continued by decision tree.

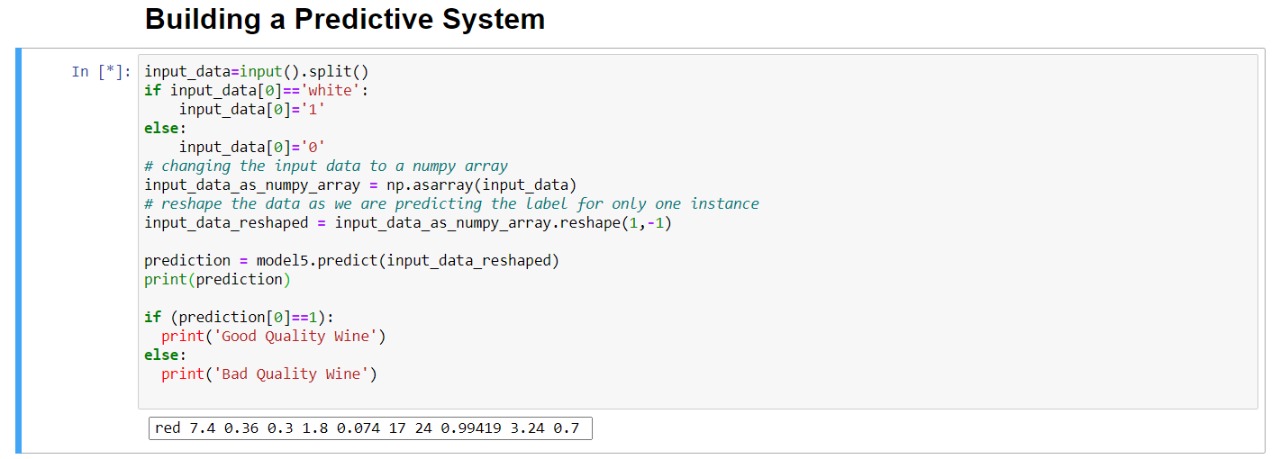
****

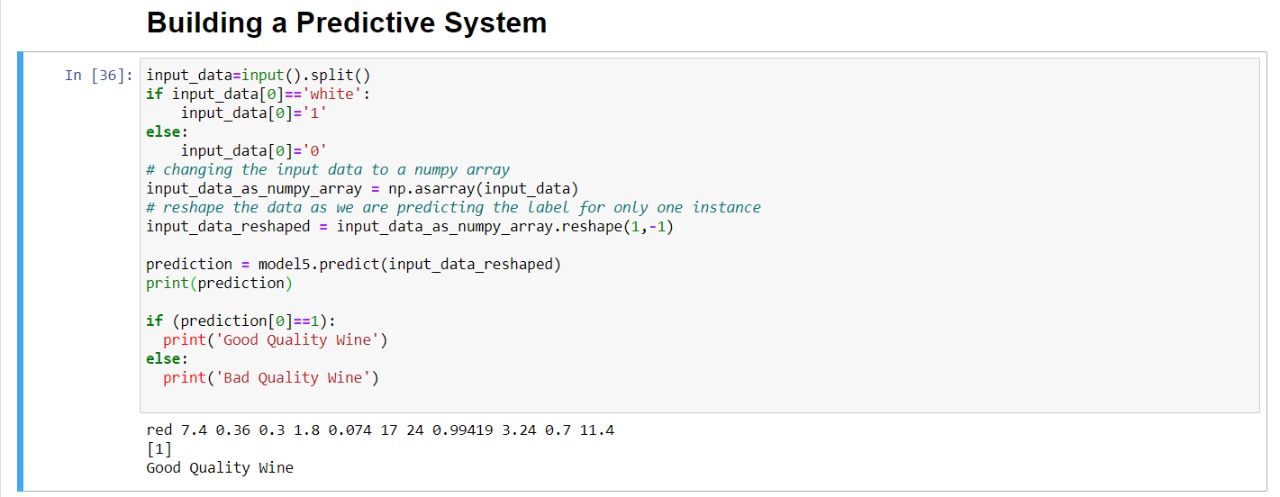
**WINE QUALITY PREDICTION**





**WINE QUALITY PREDICTION**





**WINE QUALITY PREDICTION**

**6.CONCLUSION**

This project mentioned about predicting the quality of Wine using various machine learning techniques. The feature select on algorithm provided a clear idea about the

importance of the attributes for prediction of quality, which was time consuming and expensive when done in the traditional way. We have also compared the accuracy of

each technique used in prediction of quality and it was found that these classifiers performed well. We have found that the Random Forest based feature sets performed better than

others. We have also found that the Random Forest classifier performed better compared to all other classifiers for wine data set.

In future we can try other performance measures and other machine learning techniques for better comparison on results. This analysis will help the industries to predict the quality of the different type of wines based on certain attributes and also it will helpful for them to make good product in the future.

**WINE QUALITY PREDICTION**

**7.BIBLIOGRAPHY**

* <https://www.youtube.com/watch?v=W25TEa93T_I>
* <https://www.kaggle.com/rajyellow46/wine-quality>
* <https://en.wikipedia.org/wiki/Decision_tree>
* <https://en.wikipedia.org/wiki/Random_forest>
* <https://www.researchgate.net/publication/350110244_Prediction_of_Wine_Quality_Using_Machine_Learning_Algorithms>