**SMARTBRIDGE SUMMER INTERNSHIP PROGRAM-2019**

**Team Name : Young Achievers**

**Team Members:**

**1.Kailasa Arun**

**2.Aleti Sukruthi**

**3.Penchala Sindhuja**

**4.Korrapolu Anusha devi**

**QUALITY OF WINE PREDICTION**

INTRODUCTION:

Nowadays, industries are using product quality certification to promote their products. This is a time taking process and requires the assessment given by human experts which makes this process very expensive. This explores by usage of machine learning techniques such as linear regression, neural networks and support vector machine for product quality in two ways. Firstly, determine the dependency of target variable on independent variables. Secondly, predicting the value of target variable. In this linear regression is used to determine the dependency of target variable on independent variables. On the basis of computed dependency, important variables are selected those make significant impact on dependent variable. Further, neural networks and support vector machine are used to predict the values of dependent variable. In this machine learning techniques are used to support wine industry. Wine quality assessment is one of the key elements in this context and can be used for certification. Such type of quality certification helps to assure wine quality in market. Wine has various characteristics like density, pH value, alcohol and other acids.

The use of Artificial Neural Networks (ANNs), as a non-linear statistical data modelling tool was chosen to select wine samples, grape variety, wine barrel (oak type), wine type (red or white) and production year with total Phenolic content, total and volatile acidity, and alcohol content. The aim of this modelling was to: (1) successfully predict the overall poly phenolic content of a wine sample and (2) to determine how parameters such as alcohol content, pH etc. affect the overall Poly phenolic content in wine. This type of information should enable winemakers to better optimise the concentration

OBJECTIVES OF RESEARCH:

Our objective is to provide proper quality based on some parameters like fixed acidity, critic acid etc by which quality is predicted.

As our parameters has some specific range by excluding pH range may cause severe problems to health.

The quality of wine grapes may be evaluated over a broad base, notably the following aspects:

* The balance between vigour and canopy capacity.
* The evenness of ripening.
* The vigour of the vineyard.
* The exposure of the grapes to sunburn.
* The incidence of diseases.
* The berry characteristics such as size , colour of the seeds and taste.
* The sugar content of the grapes.
* The grape colour composition of red grapes.

PROBLEM STATEMENT:

To predict the quality of wine.

REVIEW OF LITERATURE :

From the details and data associated with the production cycle in the wine production industry, a life cycle shall provide information on relevant impacts to environment, human health and more of wine production. The contributions of each sub process is of great importance in order to classify wine and know the areas which create the most problems, to be later compared and discussed about with further context to a new type of wine, ecological wine or organic wine.

DATA COLLECTION:

Data are collected on 12 different properties of the wines one of which is Quality, based on sensory data, and the rest are on chemical properties of the winesincluding density, acidity, alcohol content etc.

* **Fixed acidity**: Acidity is a fundamental property of wine, imparting sourness and resistance to microbial infection. Fixed acidity is the no of grams of tartaric acid per dimension cube.
* **Volatile acidity**: Wine spoilage is legally defined by volatile acidity which is calculated as no of grams of acetic acid per dimension cube of wine.
* **Critic acid**: It is the no of grams of critic acid per dimension cube of wine.
* **Residual sugar**: Residual sugar refers to the sugar remaining after fermentation stops. Given as no of grams per dimension cube.
* **Chlorides**: It is the no of grams of sodium chloride per dimension cube of wine.
* **Free sulphur dioxide**: It is the no of grams free sulphites per dimension cube of wine.
* **Total sulphur dioxide**: It is the no of grams of total sulphites in per dimension cube of the wine.
* **Density**: It gives the density of the wine in grams per dimension cube.
* **pH**: It gives the pH of the wines . pH is used to measure ripeness in relation to acidity.
* **Sulphates**: It gives the no of grams of potassium sulphate per dimension cube of wine.
* **Alcohol**: It gives the volume of alcohol in percentage.
* **Quality**: This is the target variable. Here the wine is rated from 1 to 10 based on the quality.

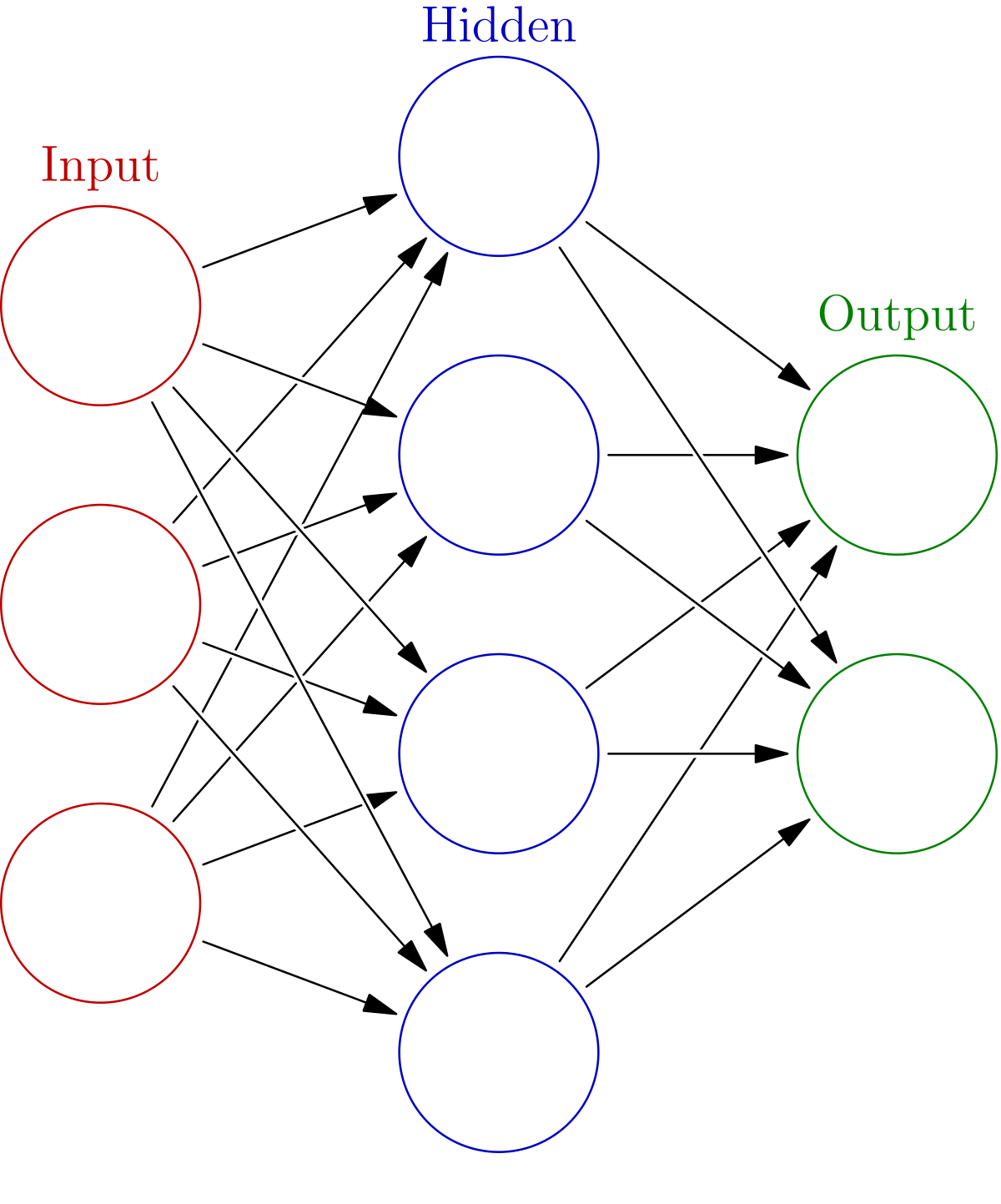
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Fixed acidity | Volatileacidity | citricacid | Residualsugar | chlorides | freesulfurdioxide | totalsulfurdioxide | density | pH | sulphates | alcohol | quality | id |  |
| 6.7 | 0.15 | 0.38 | 1.7 | 0.037 | 20 | 84 | 0.99046 | 3.09 | 0.53 | 11.4 | 6 | 981 |  |
| 6.7 | 0.44 | 0.31 | 1.9 | 0.03 | 41 | 104 | 0.99 | 3.29 | 0.62 | 12.6 | 7 | 982 |  |
| 6.1 | 0.17 | 0.21 | 1.9 | 0.09 | 44 | 130 | 0.99255 | 3.07 | 0.41 | 9.7 | 5 | 983 |  |
| 6.6 | 0.39 | 0.22 | 4 | 0.038 | 17 | 98 | 0.99018 | 3.25 | 0.53 | 13 | 7 | 984 |  |
| 6.8 | 0.32 | 0.34 | 6 | 0.05 | 5 | 129 | 0.9953 | 3.19 | 0.4 | 9.1 | 5 | 985 |  |
| 8.3 | 0.28 | 0.27 | 17.5 | 0.045 | 48 | 253 | 1.00014 | 3.02 | 0.56 | 9.1 | 6 | 986 |  |
| 6.8 | 0.14 | 0.35 | 1.5 | 0.047 | 40 | 117 | 0.99111 | 3.07 | 0.72 | 11.1 | 6 | 987 |  |
| 6.3 | 0.21 | 0.31 | 1.2 | 0.043 | 30 | 117 | 0.99158 | 3.49 | 0.68 | 11 | 6 | 988 |  |
| 6.7 | 0.31 | 0.08 | 1.3 | 0.038 | 58 | 147 | 0.9922 | 3.18 | 0.46 | 10 | 5 | 989 |  |
| 5.9 | 0.27 | 0.29 | 11.4 | 0.036 | 31 | 115 | 0.9949 | 3.35 | 0.48 | 10.5 | 8 | 990 |  |
| 5.6 | 0.2 | 0.22 | 1.3 | 0.049 | 25 | 155 | 0.99296 | 3.74 | 0.43 | 10 | 5 | 991 |  |
| 6.7 | 0.24 | 0.46 | 2.2 | 0.033 | 19 | 111 | 0.99045 | 3.1 | 0.62 | 11.9 | 6 | 992 |  |
| 6.8 | 0.19 | 0.4 | 9.85 | 0.055 | 41 | 103 | 0.99532 | 2.98 | 0.56 | 10.5 | 6 | 993 |  |
| 8 | 0.27 | 0.33 | 1.2 | 0.05 | 41 | 103 | 0.99002 | 3 | 0.45 | 12.4 | 6 | 994 |  |
| 6.8 | 0.21 | 0.27 | 18.15 | 0.042 | 41 | 146 | 1.0001 | 3.3 | 0.36 | 8.7 | 5 | 995 |  |
| 7.4 | 0.24 | 0.42 | 14 | 0.066 | 48 | 198 | 0.9979 | 2.89 | 0.42 | 8.9 | 6 | 996 |  |
| 8.1 | 0.17 | 0.21 | 1.6 | 0.036 | 24 | 119 | 0.99396 | 3.18 | 0.52 | 10.1 | 6 | 997 |  |
| 7.5 | 0.13 | 0.38 | 1.1 | 0.023 | 42 | 104 | 0.99112 | 3.28 | 0.53 | 11.8 | 6 | 998 |  |
| 7.2 | 0.37 | 0.4 | 11.6 | 0.032 | 34 | 214 | 0.9963 | 3.1 | 0.51 | 9.8 | 6 | 999 |  |
| 7.1 | 0.46 | 0.23 | 13.7 | 0.045 | 44 | 192 | 0.9981 | 3.11 | 0.53 | 9.4 | 5 | 1000 |  |
| 6.6 | 0.42 | 0.33 | 2.8 | 0.034 | 15 | 85 | 0.99 | 3.28 | 0.51 | 13.4 | 6 | 1001 |  |
| 6.9 | 0.28 | 0.37 | 9.1 | 0.037 | 16 | 76 | 0.9948 | 3.05 | 0.54 | 11.1 | 5 | 1002 |  |
| 6.2 | 0.21 | 0.38 | 6.8 | 0.036 | 64 | 245 | 0.9951 | 3.06 | 0.36 | 9.3 | 6 | 1003 |  |
| 7.8 | 0.24 | 0.38 | 2.1 | 0.058 | 14 | 167 | 0.994 | 3.21 | 0.55 | 9.9 | 5 | 1004 |  |
| 6 | 0.31 | 0.32 | 7.4 | 0.175 | 47 | 159 | 0.9952 | 3.19 | 0.5 | 9.4 | 6 | 1005 |  |
| 6.6 | 0.25 | 0.25 | 1.3 | 0.04 | 28 | 85 | 0.98984 | 2.87 | 0.48 | 11.2 | 6 | 1006 |  |
| 6 | 0.19 | 0.71 | 1.5 | 0.152 | 9 | 55 | 0.9927 | 3.12 | 0.46 | 9.8 | 6 | 1007 |  |
| 7.7 | 0.34 | 0.28 | 11 | 0.04 | 31 | 117 | 0.99815 | 3.27 | 0.29 | 9.2 | 6 | 1008 |  |
| 5.8 | 0.23 | 0.21 | 1.5 | 0.044 | 21 | 110 | 0.99138 | 3.3 | 0.57 | 11 | 6 | 1009 |  |
| 7.7 | 0.35 | 0.49 | 8.65 | 0.033 | 42 | 186 | 0.9931 | 3.14 | 0.38 | 12.4 | 8 | 1010 |  |
| 6.8 | 0.3 | 0.27 | 11.6 | 0.028 | 22 | 97 | 0.99314 | 2.96 | 0.38 | 11.7 | 6 | 1011 |  |
| 7.2 | 0.2 | 0.3 | 2 | 0.039 | 43 | 188 | 0.9911 | 3.3 | 0.41 | 12 | 6 | 1012 |  |
| 7 | 0.21 | 0.34 | 8.5 | 0.033 | 31 | 253 | 0.9953 | 3.22 | 0.56 | 10.5 | 6 | 1013 |  |
| 7.5 | 0.18 | 0.45 | 4.6 | 0.041 | 67 | 158 | 0.9927 | 3.01 | 0.38 | 10.6 | 6 | 1014 |  |
| 7.2 | 0.17 | 0.28 | 17.55 | 0.05 | 33 | 154 | 0.99971 | 2.94 | 0.43 | 9 | 7 | 1015 |  |
| 6.1 | 0.38 | 0.42 | 5 | 0.016 | 31 | 113 | 0.99007 | 3.15 | 0.31 | 12.4 | 7 | 1016 |  |
| 7.1 | 0.33 | 0.64 | 13.2 | 0.056 | 12 | 105 | 0.9972 | 3.05 | 0.39 | 9.2 | 5 | 1017 |  |
| 7.5 | 0.23 | 0.35 | 17.8 | 0.058 | 128 | 212 | 1.00241 | 3.44 | 0.43 | 8.9 | 5 | 1018 |  |
| 6 | 0.14 | 0.37 | 1.2 | 0.032 | 63 | 148 | 0.99185 | 3.32 | 0.44 | 11.2 | 5 | 1019 |  |
| 5.6 | 0.18 | 0.58 | 1.25 | 0.034 | 29 | 129 | 0.98984 | 3.51 | 0.6 | 12 | 7 | 1020 |  |
| 8 | 0.28 | 0.32 | 7.6 | 0.045 | 61 | 204 | 0.99543 | 3.1 | 0.55 | 10.1 | 6 | 1021 |  |
| 7.6 | 0.23 | 0.4 | 5.2 | 0.066 | 14 | 91 | 0.99488 | 3.17 | 0.8 | 9.7 | 5 | 1022 |  |
| 7.2 | 0.18 | 0.41 | 1.2 | 0.048 | 41 | 97 | 0.9919 | 3.14 | 0.45 | 10.4 | 5 | 1023 |  |
| 5.2 | 0.17 | 0.27 | 0.7 | 0.03 | 11 | 68 | 0.99218 | 3.3 | 0.41 | 9.8 | 5 | 1024 |  |
| 6 | 0.34 | 0.66 | 15.9 | 0.046 | 26 | 164 | 0.9979 | 3.14 | 0.5 | 8.8 | 6 | 1025 |  |
| 8.1 | 0.3 | 0.31 | 1.1 | 0.041 | 49 | 123 | 0.9914 | 2.99 | 0.45 | 11.1 | 6 | 1026 |  |
| 6.2 | 0.36 | 0.38 | 3.2 | 0.031 | 20 | 89 | 0.98956 | 3.06 | 0.33 | 12 | 7 | 1027 |  |
| 5.2 | 0.155 | 0.33 | 1.6 | 0.028 | 13 | 59 | 0.98975 | 3.3 | 0.84 | 11.9 | 8 | 1028 |  |
| 7.8 | 0.3 | 0.29 | 16.85 | 0.054 | 23 | 135 | 0.9998 | 3.16 | 0.38 | 9 | 6 | 1029 |  |

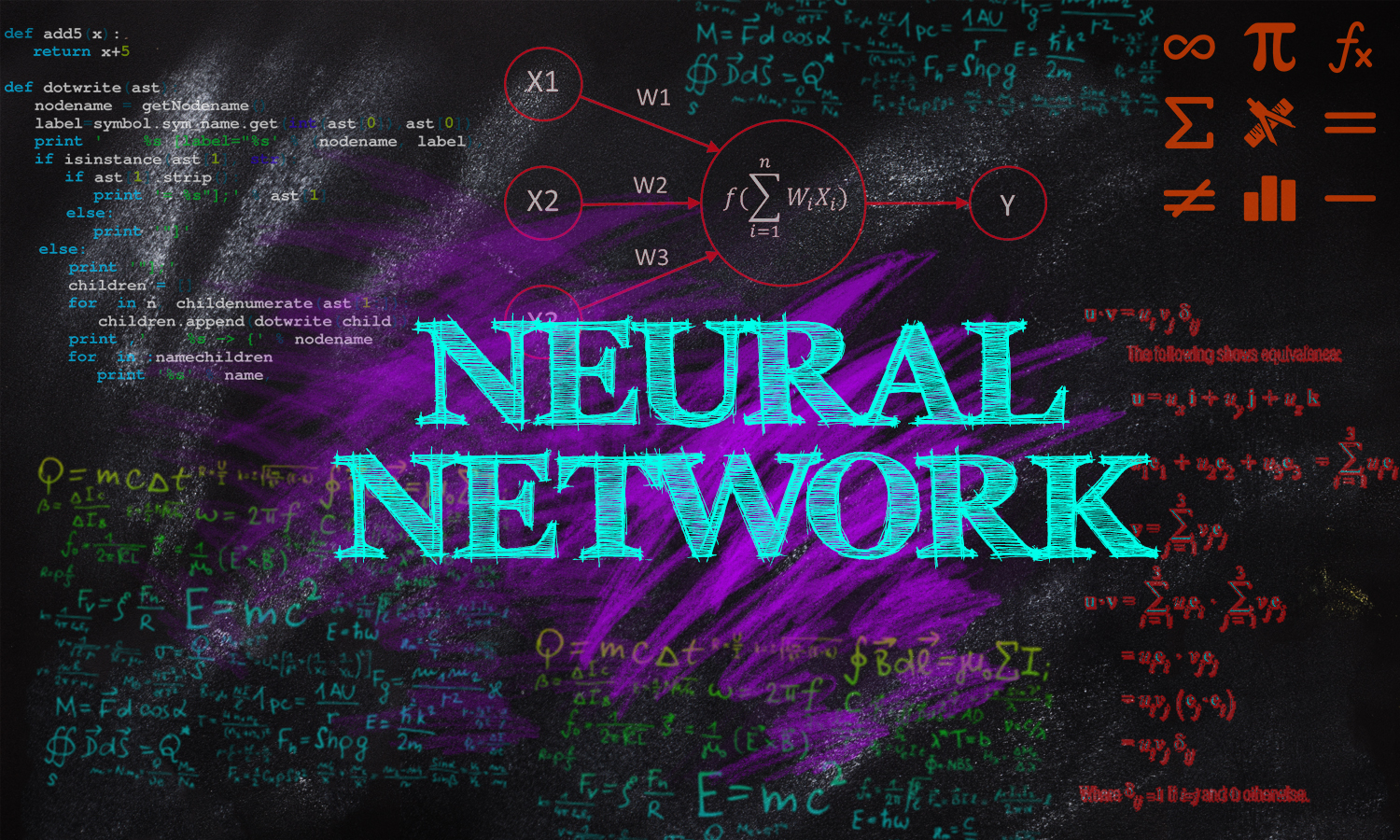
METHODOLOGY:

We have used Artificial Neural Networks for the prediction of quality of wine.

**ARTIFICIAL NEURAL NETWORKS:**

Artificial neural networks consist of a large number of independent computational units (so-called neurons) that are able to influence the computations of each other. A neuron has several inputs, and one output.

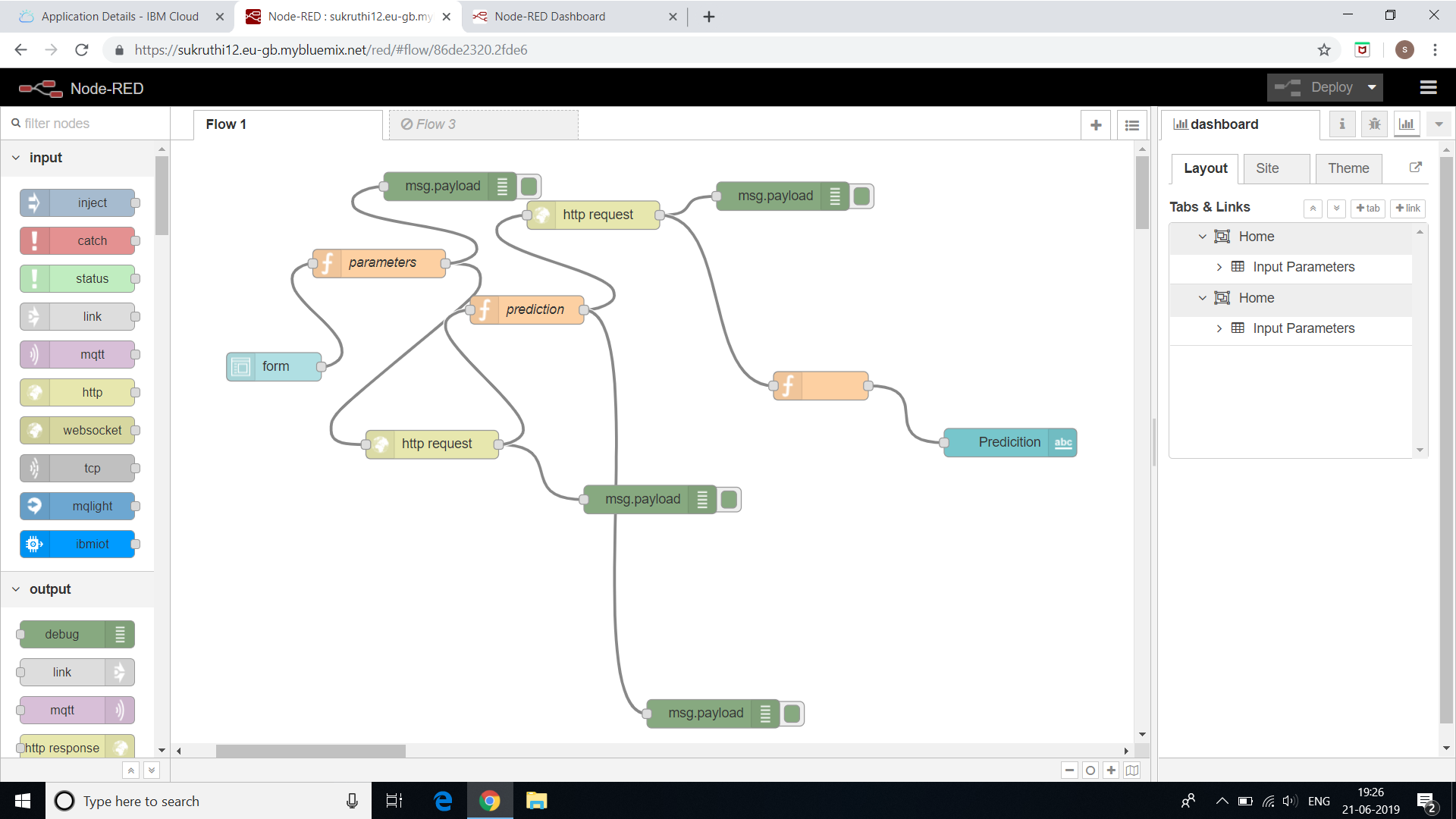
In recent years, neural networks have turned out as a powerful method for numerous practical applications in a wide variety of disciplines. In more practical terms neural networks are one of nonlinear statistical data modelling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data. In wine technology artificial neural networks (ANNs) are useful for predict the quality analyses, predicting chemical, functional and sensory properties of various wine products during processing and distribution. In wine technology, ANNs have been used for classification and for predicting wine process condition.

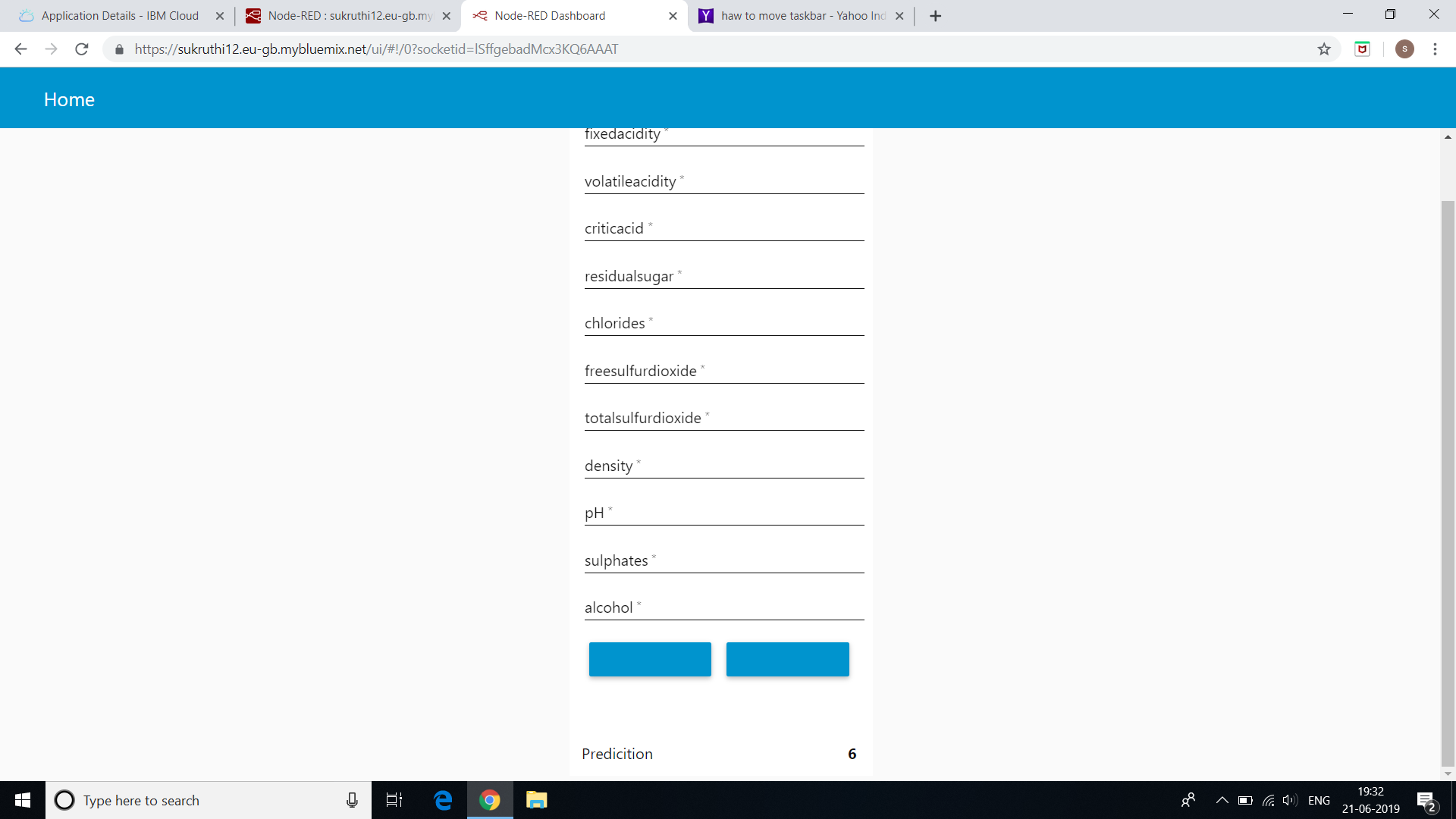


Artificial neural network is a self learning model which learns from its mistakes and give out the right answer at the end of the computation.

DATA MODELLING USING SUPERVISED ML TECHNIQUES:

* In the above dataset first we have done the data pre processing. In that 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.
* from sklearn.model\_selection import train\_test\_split is used to **split** our dataset into training and testing data, more of which will be covered later. The next import, from sklearn import preprocessing is used to **preprocess** the data before fitting into predictor, or converting it to a range of -1,1, which is easy to understand for the machine learning algorithms.
* We use pd.read\_csv() function in pandas to import the data by giving the dataset url of the repository. Notice that ‘;’ (semi-colon) has been used as the separator to obtain the csv in a more structured format.
* So, if we analyse this dataset, since we have to predict the wine quality, the attribute quality will become our dependent variable(output) and the rest of the attributes will become the independent variable(input).
* We just stored quality in y, which is the common symbol used to represent the dependent variable machine learning and dropped quality and stored the remaining features in X .
* Next, we have to split our dataset into test and train data, we will be using the train data to train our model for predicting the quality. The next part, that is the test data will be used to verify the predicted values by the model.
* We have used, train\_test\_split() function that we imported from sklearn to split the data. Notice we have used test\_size=0.2 to make the test data 20% of the original data. The rest 80% is used for training.





FINDINGS AND SUGGESTIONS:

* We found that Artificial neural network is more accurate for predicting the quality of wine.
* According to our analysis we have decided to use the aritificial neural network to predict the quality of wine.

**CONCLUSION:**

* Target variable:Quality
* Parameters associated:Alcohol,pH,Acidity,Volatile acidity etc.
* The following quality can be achieved

|  |  |  |
| --- | --- | --- |
| S.NO | Quality(1-10) | Grade |
| 1 | 8 | best |
| 2 | 7 | better |
| 3 | 6 | good |