2020 International Conference on Computer Communication and Informatics (ICCCI -2020), Jan. 22-24, 2020, Coimbatore, INDIA

Red Wine Quality Prediction Using Machine Learning Techniques

Sunny Kumar
Department of Computer Science
Roorkee Institute of Technology
Roorkee, Uttarakhand, India
sunnykumar825288@gmail.com

Kanika Agrawal
Department of Computer Science
Roorkee Institute of Technology
Roorkee, Uttarakhand, India
ka841995@gmail.com

Nelshan Mandan
Department of Computer Science
Roorkee Institute of Technology
Roorkee, Uttarakhand, India
nelsonmishra007@gmail.com

***Abstract*—Nowadays people try to lead a luxurious life. They tend to use the things either for show off or for their daily basis. These days the consumption of red wine is very common to all. So it became important 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. Dataset is taken from the sources and the techniques such as Random Forest, Support Vector Machine and Naïve Bayes are applied. Various measures are calculated and the results are compared among training set and testing set and accordingly the best out of the three techniques depending on the training set results is predicted. Better results can be observed if the best features out from other techniques are extracted and merged with one another to improve the accuracy and efficiency.**

Keywords— processes; data extraction; Naïve Bayes; SVM; Random Forest; quality.

I. Introduction

Data mining is the path toward discovering new examples to separate the quality information from immense storehouse. It incorporates various kinds of measurements, machine learning and arrangement of databases. The fundamental target is to isolate significant information from tremendous database and after those changes over the important substance into a meaningful substance for future research. Knowledge Discovery in Databases (KDD) generally incorporates data mining as its critical investigation step. Aside from the analysis, it likewise incorporates intricacy contemplations, large data house analysis, pre analysis and post analysis of the information and finally finds the interesting data and then updates it. Information analysis regularly tests the speculations and models on the information, paying little mind to the substance of information. Data mining is a blend of factual models and machine learning. The term data mining deals with the extraction of learning and models from huge dataset. Data mining undertaking is the programmed strategy of extracting patterns from the huge proportion of data, finds the inconsistencies and then finally detects the required result. For better results various data mining algorithms and their best features are stitched together in such a way that they produce efficient and accurate results with less of errors. Different terms like data fishing data dredging, and data snooping refers to the creation of new theories out from the bigger data collection.

II. Literature Review

Today, various customers appreciate wine to an ever increasing extent. Wine industry is looking into new advances for both wine making and offering structures in order to back up this development [1]. Physicochemical and tactile tests are utilized for assessing wine confirmation [2]. The segregation of wines isn't a simple procedure inferable from the intricacy and heterogeneity of its headspace. The arrangement of wines is significant in light of the fact that of various reasons. These reasons are financial estimation of wine items, to secure and guarantee the nature of wines, to preclude corruption of wines, and to control refreshment preparing [3]. Data mining innovations have been applied to plan wine quality. The point of machine learning techniques like various applications is to make models from information to anticipate wine quality. In 1991, a "Wine" informational index which contains 178 occurrences with estimations of 13 distinctive synthetic constituents, such as, alcohol, magnesium was given into UCI store to order three cultivars from Italy [4]. For new information mining classifiers this data has been significantly utilized as a benchmark since it is exceptionally simple to separate. For wine characterization as indicated by geological area; Principal Component Analysis (PCA) was done and announced [5]. The information they utilized in their examination incorporates 33 Greek wines with physicochemical factors. Another work of wine grouping relied upon the physicochemical data. This data associated with wine smell chromatograms as estimated with a Fast GC Analyser [6]. In the last investigation, three portrayal methods, for example, Naïve Bayes, Random Forest and Support Vector Machines (SVM) are contrasted agreeing and their exhibition in a two-organized architecture. Some have proposed a couple of uses of data mining frameworks to wine quality appraisal. Cortez et al. [1] proposed a taste desire framework. In their taste expectation framework, a Support Vector Machine, Naïve Bayes, and a Random Forest were applied to engineer examination of wines. Shanmuganathans procedure was about forecast the effects of season and climate on wine yields and wine quality [7]. The Wine informatics framework as shown by Chen et al. [8] depicted the flavour and traits of wine from typical language audits. They used association rules and progressive clustering. In research article [9], the authors have compared different machine learning algorithms such as Naïve Bayes, Decision Tree and Support Vector Machines on Cardiotocography data to predict the best algorithm out of them. In research article [10], authors showed

978-1-7281-4514-3/20/$31.00 ©2020 IEEE

the different techniques, applications and challenges faced by text analysis.

III. Research Methodoloy And Experiment Design

The data is extracted from UCI machine learning repository [11] to do the research. The dataset contains 1599 instances with 12 variables for red wine data. The data evaluation is based on the inputs taken and then finally concludes with the prediction of red wine quality. For this dataset qualities are predicted between the range 3-8, where '3' predicts poor quality of red wine and '8' predicts excellent quality of red wine.

The highlights include fixed acidity, citrus acid, volatile acidity, residual sugar, chlorides, thickness, free sulphur dioxide, absolute sulphur dioxide, pH, alcohol and sulphates .The value of pH depicts the acidity and basicity of the wine. Consumable wines have their pH scale between 3-4. The amount of salt depicts the chloride content in the wine. The goal of the information file is to anticipate the rating that master will accommodate a wine test, utilizing an extent of physicochemical properties, for instance, acidity and liquor properties. As a result of security and strategic issues, simply physicochemical (inputs) and output factors are available.

In the field of machine learning, a confusion matrix is a table that is frequently used to depict the presentation of a grouping model on a lot of test information for which the genuine qualities are known. It permits the perception of the presentation of a calculation. This research basically uses the red wine data set and then calculates the confusion matrix, relevant performance measures and finally compares the different machine learning algorithms on the basis of accuracy predicted on this dataset.

A. Measures to Calculate Performance in Research

Performance measures are the measures that are used in the research so as calculate and evaluate the techniques to detect the effectiveness and efficiency of the techniques. Some of them are listed below:

- Accuracy: The value predicted when the sum of True Positive and True Negative is divided by the sum of True Positive, False positive, False Negative and True Negative values of a confusion matrix.

Accuracy=(True Positive + True Negative) / (True Positive + False Positive + False Negative + True Negative)

- Precision: The value obtained when True Positive is divided by the sum of True Positive and False Positive values of a confusion matrix.

Precision = True Positive / (True Positive + False Positive)

- Recall: Sensitivity sometimes also known as Recall. It is the value obtained when True Positive is divided by the sum of True Positive and False Negative values of a confusion matrix.

Recall= True Positive / (True Positive + False Negative)

- Specificity: It is opposite of Recall.

Specificity = True Negative / (True Negativity + False Positive)

- F-Measure: F1 Score is obtained by multiplying Recall and Precision divided by sum of Recall and precision of a confusion matrix. Result is then multiplied by two.

*F1 Score = 2 * (Recall * Precision) / (Recall + Precision)*

- Misclassification Error: It is obtained by subtracting accuracy from one and gives the error in the calculations done.

Misclassification Error = 1-Accuracy

B. Techniques Involved in Research

Techniques used are as given below. These are:

- Naive Bayes Algorithm: Naive Bayes algorithm relies upon bayes speculation. To find whether a particular part has a spot with a particular class it utilizes the possibility of likelihood. Naive Bayes classifiers are profoundly versatile, requiring various parameters straight in the quantity of factors in a learning problem.
- Support Vector Machine: This technique was taken from factual learning theory by Vapnik and Chervonenkis. It was first exhibited in 1992 by Boser Guyon and Vapnik. This technique is utilized for the characterization of both nonlinear and linear information. It utilizes a nonlinear mapping to change the primary preparing information into a higher estimation. It scans for the linear optimal isolating hyperplane in this new estimation. A hyperplane can isolate information from two classes, with a reasonable nonlinear mapping to adequately high estimation. The SVM uses support vectors and edges to find this hyperplane [12]. A SVM model is a portrayal of the models as point in space, mapped with the goal that instances of the different classes are isolated by a gap that is as wide as would be prudent. SVM can play out a nonlinear type of classification.
- Random Forest: This technique utilizes a blend of tree indicators; each individual tree depends upon an random vector. This arbitrary vector has indistinguishable and a similar circulation for all trees in the forest. It was portrayed by Breiman in 2001[13]. Random forest helps in predicting the important variables in classification and regression problems in a simple way.

IV. Implementation

An analysis is done on the redwine.csv dataset extracted from huge database [11] that contains the details of Red Wine. The datasets contain 1599 observation and have 12 attributes such as fixed acidity, sugar, sulphates, chlorides, volatile acidity, citrus acid, residual, free sulphur dioxide, absolute sulphur dioxide, thickness, pH and alcohol. All these attributes are used from the dataset. The dataset is seperated into sets of training and testing with the probabilities 0.7 & 0.3 respectively. Libraries such as naïve bayes, pysch, dplyr, knitr, ggplot2, random forest and e1701 are imported. After importing the libraries, summary of the model is calculated using Naïve Bayes, Random Forest and support Vector Machine algorithms. After calculating the summaries, the confusion matrix of 6*6, depending on the dataset observations and the quality, is calculated. Variable ‘matrix’ is used to denote the confusion matrix.

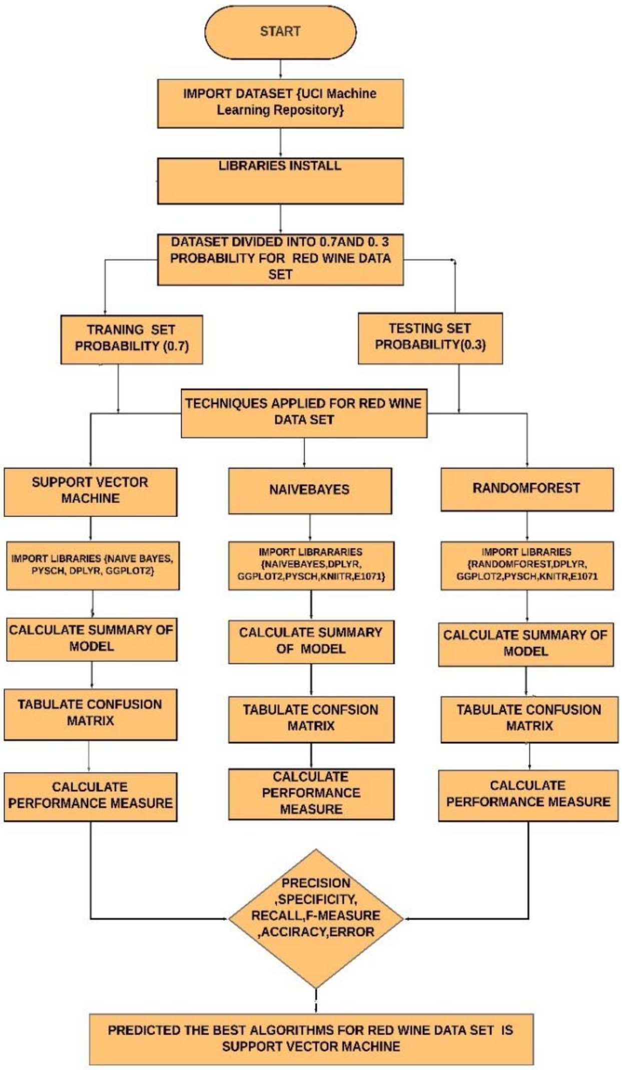

Fig. 1. Flow chart showing steps used in research to predict red wine quality

Further various performance measures such as specificity, precision, f-score, recall, accuracy, and misclassification error are calculated using the algorithms. Results were predicted on the basis of these measures. This research finally shows that the best accuracy is shown by the Support Vector Machine algorithm on red wine dataset extracted from UCI, then Random Forest algorithm and last comes the algorithm that depends on probability. “Fig. 1” shows the steps used in the research and hence detect the quality of red wine using data mining techniques. “Fig. 2” shows 1599 observations and 12 variables of red wine dataset. “Fig. 3” shows the deviation from standard values and the corresponding average values for different attributes of training set using Naïve Bayes algorithms. “Fig. 4” shows the deviation from standard values and the corresponding average values for different attributes of testing set using Naïve Bayes algorithms. “Fig. 5’ shows the error matrix of red wine dataset for training set using Naïve Bayes algorithm. “Fig. 6” shows the error matrix of red wine dataset for testing set using Naïve Bayes algorithm. “Fig. 7” shows the error matrix of red wine dataset for training set using Support Vector Machine algorithm. “Fig. 8” shows the error matrix of red wine dataset for Testing set using Support Vector Machine algorithm. “Fig. 9” shows the error matrix of dataset of red wine for training set using Random Forest Algorithm. “Fig. 10” shows the table of red wine dataset for testing set using Random Forest algorithm.

```
'data.frame':   1599 obs. of  12 variables:
 $ fixed.acidity       : num  7.4 7.8 7.8 11.2 7.4 7.4 7.9 7.3 7.8 7.5 ...
 $ volatile.acidity    : num  0.7 0.88 0.76 0.28 0.7 0.66 0.6 0.65 0.58 0.5
 ...
 $ citric.acid         : num  0 0 0.04 0.56 0 0 0.06 0 0.02 0.36 ...
 $ residual.sugar      : num  1.9 2.6 2.3 1.9 1.9 1.8 1.6 1.2 2 6.1 ...
 $ chlorides           : num  0.076 0.098 0.092 0.075 0.076 0.075 0.069 0.065
0.073 0.071 ...
 $ free.sulfur.dioxide : num  11 25 15 17 11 13 15 15 9 17 ...
 $ total.sulfur.dioxide: num  34 67 54 60 34 40 59 21 18 102 ...
 $ density             : num  0.998 0.997 0.997 0.998 0.998 ...
 $ pH                  : num  3.51 3.2 3.26 3.16 3.51 3.51 3.3 3.39 3.36 3.35
 ...
 $ sulphates           : num  0.56 0.68 0.65 0.58 0.56 0.56 0.46 0.47 0.57 0.8
 ...
 $ alcohol             : num  9.4 9.8 9.8 9.8 9.4 9.4 9.4 10 9.5 10.5 ...
 $ quality             : int  5 5 5 6 5 5 5 7 7 5 ...
```

Fig. 2. Snapshot showing 1599 observations and 12 variables of red wine dataset

```
=============================== Naive Bayes ===============================

 Call:
naive_bayes.formula(formula = quality ~ ., data = train)

---------------------------------------------------------------------------

Laplace smoothing: 0

---------------------------------------------------------------------------

 A priori probabilities:

          3           4           5           6           7           8
0.005258545 0.033304119 0.429447853 0.394390885 0.125328659 0.012269939

---------------------------------------------------------------------------

 Tables:

---------------------------------------------------------------------------
 ::: fixed.acidity (Gaussian)
---------------------------------------------------------------------------

fixed.acidity        3        4        5        6        7        8
         mean 8.466667 7.828947 8.106327 8.314222 8.746853 8.628571
         sd   2.015606 1.641836 1.485782 1.802083 1.843825 1.870182

---------------------------------------------------------------------------
 ::: volatile.acidity (Gaussian)
---------------------------------------------------------------------------

volatile.acidity         3         4         5         6         7
            mean 0.9066667 0.6922368 0.5769082 0.4998444 0.4006294
            sd   0.3641108 0.2098477 0.1662908 0.1604368 0.1466034
```

```
volatile.acidity         8
            mean 0.4385714
            sd   0.1598488

-----------------------------------------------------------------------------
 ::: citric.acid (Gaussian)
-----------------------------------------------------------------------------

citric.acid         3         4         5         6         7         8
       mean 0.2083333 0.1936842 0.2402653 0.2744444 0.3585315 0.3828571
       sd   0.2899253 0.2150603 0.1798765 0.1924691 0.1851454 0.2040147

-----------------------------------------------------------------------------
 ::: residual.sugar (Gaussian)
-----------------------------------------------------------------------------

residual.sugar        3        4        5        6        7        8
          mean 2.566667 2.607895 2.571122 2.533333 2.573077 2.764286
          sd   1.579451 1.172980 1.460379 1.613063 1.166338 1.418888

-----------------------------------------------------------------------------
 ::: chlorides (Gaussian)
-----------------------------------------------------------------------------

chlorides          3          4          5          6          7          8
     mean 0.13683333 0.09294737 0.09303265 0.08409778 0.07695105 0.06935714
     sd   0.08202296 0.08779705 0.05378670 0.03925412 0.03052644 0.01291158

-----------------------------------------------------------------------------

# ... and 7 more tables

-----------------------------------------------------------------------------
```

Fig.3. Snapshot showing the mean and standard deviation values for different attributes of training set using Naïve Bayes algorithms

```
================================ Naive Bayes ================================

 Call:
naive_bayes.formula(formula = quality ~ ., data = test)

-----------------------------------------------------------------------------

Laplace smoothing: 0

-----------------------------------------------------------------------------

 A priori probabilities:

          3           4           5           6           7           8
0.008733624 0.032751092 0.417030568 0.410480349 0.122270742 0.008733624

-----------------------------------------------------------------------------

 Tables:

-----------------------------------------------------------------------------
 ::: fixed.acidity (Gaussian)
-----------------------------------------------------------------------------

fixed.acidity        3        4        5        6        7        8
         mean 8.200000 7.653333 8.323560 8.426064 9.192857 8.350000
         sd   1.606238 1.637012 1.743173 1.789993 2.317119 3.197395

-----------------------------------------------------------------------------
 ::: volatile.acidity (Gaussian)
-----------------------------------------------------------------------------

volatile.acidity          3          4          5          6          7
            mean 0.85125000 0.69833333 0.57738220 0.49183511 0.41232143
            sd   0.32522748 0.25207472 0.16134698 0.16250367 0.14259913


volatile.acidity          8
            mean 0.37000000
            sd   0.05830952

-----------------------------------------------------------------------------
 ::: citric.acid (Gaussian)
-----------------------------------------------------------------------------

citric.acid         3         4         5         6         7         8
       mean 0.1150000 0.1246667 0.2524607 0.2723404 0.4176786 0.4200000
       sd   0.2035518 0.1556033 0.1805014 0.2018016 0.2122091 0.2092845

-----------------------------------------------------------------------------
 ::: residual.sugar (Gaussian)
-----------------------------------------------------------------------------

residual.sugar         3         4         5         6         7         8
          mean 2.7375000 2.9133333 2.4204188 2.3428191 3.0973214 1.9250000
          sd   1.3085456 2.8610354 1.0544526 0.8976434 1.7488547 0.2217356

-----------------------------------------------------------------------------
 ::: chlorides (Gaussian)
-----------------------------------------------------------------------------

chlorides           3           4           5           6           7
     mean 0.101000000 0.084933333 0.091973822 0.087010638 0.075660714
     sd   0.030386400 0.033775449 0.053636931 0.040325148 0.026764479

chlorides           8
     mean 0.065250000
     sd   0.005737305

-----------------------------------------------------------------------------

# ... and 7 more tables

-----------------------------------------------------------------------------
```

Fig. 4. Snapshot showing the mean and standard deviation values for different attributes of testing set using Naïve Bayes algorithm.

> matrix

p1	3	4	5	6	7	8
3	6	0	0	0	0	0
4	0	38	9	5	1	0
5	0	0	476	0	2	0
6	0	0	5	445	0	0
7	0	0	0	0	140	0
8	0	0	0	0	0	14

Fig. 5. Error matrix of red wine dataset for training set using Naïve Bayes algorithm.

> matrix

p1	3	4	5	6	7	8
3	4	0	0	0	0	0
4	0	14	1	1	0	0
5	0	1	185	0	0	0
6	0	0	0	183	1	0
7	0	0	4	2	55	0
8	0	0	1	2	0	4

Fig. 6. Error matrix of red wine dataset for testing set using Naïve Bayes algorithm.

> matrix

	3	4	5	6	7	8
3	0	0	0	0	0	0
4	2	1	1	1	0	0
5	4	25	377	116	4	0
6	1	10	102	290	71	5
7	0	1	2	34	74	5
8	0	0	0	1	0	0

Fig. 7. Error matrix of red wine dataset for training set using Support Vector Machine algorithm.

> matrix

	3	4	5	6	7	8
3	0	0	0	0	0	0
4	0	0	1	0	0	0
5	3	8	155	50	4	0
6	0	8	42	132	23	4
7	0	0	1	13	22	4
8	0	0	0	1	1	0

Fig. 8. Error matrix of red wine dataset for testing set using Support Vector Machine algorithm.

> matrix

	3	4	5	6	7	8
3	0	0	0	0	0	0
4	0	1	0	0	0	0
5	5	27	385	115	6	0
6	2	9	95	314	85	6
7	0	0	2	13	58	4
8	0	0	0	0	0	0

Fig. 9. Error matrix of red wine dataset for training set using Random Forest Algorithm

```
> matrix

     3   4   5   6   7   8
3    0   0   0   0   0   0
4    0   0   0   0   0   0
5    2  10 165  47   4   0
6    1   6  34 149  36   4
7    0   0   0   0  10   4
8    0   0   0   0   0   0
```

Fig. 10. Error matrix of red wine dataset for testing set using Random Forest algorithm.

V. Results And Discussions

Nowadays people used to consume red wine either as a necessity or for show off. All this results in health loss. Hence to preserve human health it became essential to predict the red wine quality before its consumption. So, in this research the dataset taken contains the information related to red wine extracted from the database [11] which is used to predict the wine quality. Different machine learning algorithms are executed on the dataset in RStudio software. Accuracy is calculated and the best algorithm is predicted for a given dataset. During the usage, the data is separated into testing set and training set each with probability of 0.3 and 0.7 respectively. The result shows that, accuracy obtained for training set and testing set using Naïve Bayes algorithm are 55.91% and 55.89% respectively , using SVM algorithm are 67.25% and 68.64% respectively and using Random Forest algorithm are 65.83% and 65.46% respectively. As there is high probability of division for training set hence taking the accuracy of training sets for examination shows that Support Vector Machine algorithm gives the more accuracy then Random Forest algorithm and last comes Naïve Bayes algorithm. Results can also be modified if the research uses the technique that involves the amalgamation of all the above three algorithms. If the proper adjustment for the hyperplane of the svm algorithm and the accurate balanced tree and appropriate probability are taken then more efficient results can be obtained..

TABLE I. Performance measures of training set of red wine dataset using Naïve Bayes

	Training set					
	Wine Quality 3	Wine Quality 4	Wine Quality 5	Wine Quality 6	Wine Quality 7	Wine Quality 8
Precision	0.3333333	0.5	0.9939577	0.9813953	0.956044	03333333
Recall	0.5	0.4545455	0.9648094	0.9906103	0.9775281	0.6666667
Specificity	0.00175901	0.9953618	0.9982159	0.4761905	0.9094608	0.991684
F-measure	0.4	0.4761905	0.9762611	0.9859813	0.337037	0.4444444
Accuracy(%)	0.5591586					
Error (%)	0.4408414					

. Table I below shows the performance measure values for training set of red wine dataset using Naïve Bayes algorithm.

Table II below shows that performance measure values for testing set of red wine dataset using Naïve Bayes algorithm. Table III below shows the performance measure values for training set of red wine dataset using Support Vector Machine algorithm. Table IV below shows the performance measure values for testing set of red wine dataset using Support Vector Machine algorithm. Table V below shows the performance measure values for training set of red wine dataset using Random Forest algorithm. Table VI below shows that performance measure values for testing set of red wine dataset using Random Forest algorithm.

TABLE II. Performance measures of testing set of red wine dataset using Naïve Bayes algorithm.

	Testing set					
	Wine Quality 3	Wine Quality 4	Wine Quality 5	Wine Quality 6	Wine Quality 7	Wine Quality 8
Precision	0	0.3333333	1	0,9767442	0.9393939	0
Recall	0	0.25	0.972028	0.9882353	0.96875	0.5
Specificity	0.00220264	0.9953704	1	0.4304636	0.9192708	0.9900498
F-measure	0	0.2857143	0.9858156	0.9824561	0	0
Accuracy(%)	0.558952					
Error (%)	0.441048					

TABLE III. Performance measures of training set of red wine dataset using Support Vector Machine algorithm.

	Training set					
	Wine Quality 3	Wine Quality 4	Wine Quality 5	Wine Quality 6	Wine Quality 7	Wine Quality 8
Precision	0	0.5	0.9446547	0.9781931	0.8923077	0
Recall	0	0.5	0.9846547	1	1	0
Specificity	0	0.9990834	0.9946572	0.6796117	0.9429695	0.9859719
F-measure	0	0.5	0.9796438	0.9889764	0	0
Accuracy(%)	0.6725821					
Error(%)	0.3274179					

TABLE IV. Performance measures of testing set of red wine dataset using Support Vector Machine algorithm.

	Testing set					
	Wine Quality 3	Wine Quality 4	Wine Quality 5	Wine Quality 6	Wine Quality 7	Wine Quality 8
Precision	0	0	0.9763314	0.9802632	0.7692308	0
Recall	0	0	1	1	1	0
Specificity	0	1	0.9537572	0.785124	0.9766355	0.9859155
F-measure	0	0	0.991453	0.9900332	0	0
Accuracy(%)	0.6864407					
Error(%)	0.3135593					

TABLE VI. Performance measures of training set of red wine dataset using Random Forest algorithm.

	Training set					
	Wine Quality 3	Wine Quality 4	Wine Quality 5	Wine Quality 6	Wine Quality 7	Wine Quality 8
Precision	0	0.3333333	0.986911	0.9830508	0.9367089	0
Recall	0	0.25	0.9973545	1	1	0
Specificity	0	0.9981618	0.9955397	0.6219136	0.9267327	0.99002
F-measure	0	0.2857143	0.9804941	0.991453	0	0
Accuracy (%)	0.6583851					
Error (%)	0.3416149					

TABLE VI. Performance measures of testing set of red wine dataset using Random Forest algorithm.

	Testing set					
	Wine Quality 3	Wine Quality 4	Wine Quality 5	Wine Quality 6	Wine Quality 7	Wine Quality 8
Precision	0	0	0.9748428	0.977778	0.88	0
Recall	0	0	0.9872611	1	1	0
Specificity	0	1	0.9914163	0.6920152	0.9461358	0.985782
F-measure	0	0	0.9627329	0.988764	0	0
Accuracy(%)	0.654661					
Error(%)	0.3415339					

VI. CONCLUSIONS AND FUTURE SCOPE

Data mining nowadays is most important technique which is utilized for investigation of the archives. It looks at the information and produces the required yield. With the headway in the innovation it helps in playing the sound test in the market thus benefits the client. As a result of its property of investigating the information it is utilized in the examination to process diverse execution appraisals utilizing different calculations. In this exploration accuracy, precision, misclassification error, F-score, recall and specificity are resolved. Since the training dataset contains about 70% of the data from the original dataset, thus the results demonstrates the Support Vector Machine as the best algorithm giving an accuracy of 67.25% implemented on red wine quality prediction on RStudio software, then comes Random Forest giving an accuracy 65.83% and last comes the Naïve Bayes algorithm giving an accuracy of 55.91%. In future, better algorithms can be developed which involves the combination of best features of all other data mining techniques. If certain adjustments in the hyperplane, and balanced tree technique along with the appropriate probability are used then much better accuracy can be observed.

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 Electronic Noses,” 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 modelling 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, 2019.

[11] UCI Machine Learning Repository, Wine quality data set, [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Wine+Quality.

[12] J. Han, M. Kamber, and J. Pei, “Classification: Advanced Methods,” in Data Mining Concepts and Techniques, 3rd ed., Waltham, MA, USA: Morgan Kaufmann, 2012, pp. 393-443.

[13] W. L. Martinez, A. R. Martinez, “Supervised Learning” in Computational Statistics Handbook with MATLAB, 2nd ed., Boca Raton, FL, USA: Chapman & Hall/CRC, 2007, pp. 363-431.