Wine Quality Classification Implementing Support Vector Machine

Aadishesh Sharma
Department of Computer Science and Engineering
DAV Institute of Engineering and Technology
Punjab, India

Arshpreet Kaur
Department of Computer Science and Engineering
DAV Institute of Engineering and Technology
Punjab, India

Abstract —Quality of wine is evaluated by its physical aspects such as taste, odour, aroma, mouth feel and after taste sensations it leaves with the taster. This is a subjective matter changing from people with different taste. This conspicuously cannot be considered as an accurate quality assessment standard. On the contrary, chemical properties do offer more stable and conclusive standards for the quality assessment. In this paper, chemical properties having much influence have been found and worked upon. The pH is found responsible for acidity in wine and other properties such as alcohol concentration which affects the taste of the wine can be assimilated to predict the quality of wine. For the assessment, discriminative model approach is applied used in Machine Learning on the wine data set using Support Vector Machines(SVMs). The results gathered can be used for providing guidance to the wine manufacturers regarding quality of wine and helps in quoting price for the same. Thus, eliminating the inconsistency of wine tasters in quality assessment of wine.

Keywords—***Discriminative model, Support Vector Machine, Wine Quality, Machine Learning, Classification.***

I. INTRODUCTION

Wine from the ancient times has been a symbol of affluence and luxury across different cultures. Wine has evolved from a source of nutrition to a food complementary and reflection of a healthy lifestyle. Chemical properties of wine are a fix standard and can offer better results for quality assessment of wine. Classification can be understood as a problem of identifying a set of groups a new observation belongs to, on account of a training set of data. In classification technique, data are divided into more or equal to two classes, and the learner must produce a model that assigns unobserved data to one or more of these classes. For the classification of data, Discriminative model approach is applied using Support Vector Machines(SVM's).Support Vector Machines[1] can be understood as a learning model that analyses the data and produces results which are used for classification and regression purposes[13]. Given a set of training data, Support Vector Machine training algorithm builds a model that allocates new unseen data to one category or the other, thus making it a binary classifier. SVM can be further classified into Linear SVM and Non-Linear SVM which are used for the classification of data linearly separable and non-linearly separable respectively.

In our work, SVM is applied for the classification of wine data according to its quality categorized into Good, Bad and Average with respect to their respective features. The quality of wine is measured on properties like alcohol and pH which significantly affects the quality of wine [3] and thus, are used to predict the quality.

II. LITERATURE REVIEW

A. Classification

Classification in machine learning can be understood as a problem of identifying set of categories to which data belongs to, on the basis of previously trained data whose categories are known. Classification in machine leaning can be done in supervised, unsupervised and by reinforcement technique.

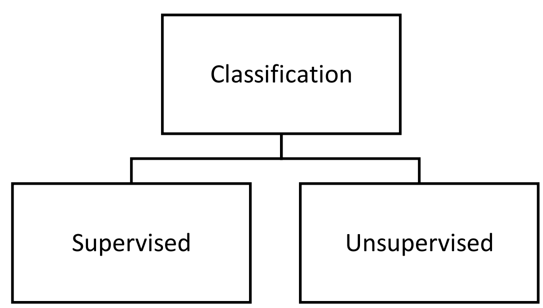

Fig 1 .Classification types

Supervised learning is the machine learning approach of identifying and predicting new unseen examples of data from the labelled training dataset. In supervised learning, we are provided with a data set and we can relate easily what would be our correct output, having the abstraction that there is some connection among the input and the output. Supervised learning problems can be further grouped into regression and classification problems.

Unsupervised learning is a machine learning approach which is used to deduce information from datasets comprising of input data without any labelled responses. Unsupervised learning algorithms permit us to tackle with the difficult problems without having the appropriate knowledge about the results. We can make out structure from the given data where we don't need to know what would be the outcome of the variables. With unsupervised learning models, it is not possible to evaluate the results obtained as we don't have prior knowledge regarding the given data set.

B. Support Vector Machine:

SVM's are supervised learning technique that analyse the data and often discern the related patterns used for classification and regression dissection. It classifies the data into categories that is binary classification and on the basis of the SVM's training algorithm which constructs the model that allocates the new examples in the dataset into one of the two respective classes.

SVM constructs a hyperplane or set of hyperplanes which classify the data into respective category. Preferably, classification or separation is considered to be good if the hyperplane have the maximum possible distance between the nearest training data points of the two classes.

Let the data points are :

$$(z1,y1),(z2,y2),(z3,y3),(z4,y4)\ldots\ldots\ldots.(z_n, y_n).$$

where y_n=1 / -1 , simply signify the respective class of the $z_n$and n is the no. of sampling. All z_n are supposed to be the vector. The hyperplane which separates the data is the decision rule:

$$w . z + b = 0 \qquad \text{----- (1)}$$

where w is a vector perpendicular to the separating hyperplane and b is scalar. The offset parameter b provides us to increase the margin and moreover maximum margin is required for the better classification of the data which is known as widest street approach. As the main objective is to maximize the distance between the hyperplanes, we are concerned with SVM and the parallel hyperplanes. Parallel hyperplanes are represented by the following equations:

$$w.z + b = 1$$

$$w.z + b = -1$$

If the training data are decoded as linearly separable, then hyperplanes will be chosen in such a way that the distance between them is viably the widest and moreover there should be no points between the respective hyperplanes [15]. By geometry, the distance between the hyperplane is found to be 2/|w|. So in order to maximize the distance, |w| is minimize. To excite data points, the following condition is considered for each i either.

$$w.\ zi - b \geq 1 \quad \text{or} \quad w.\ zi - b \leq 1$$

Above equation is expressed as:

$$yi\ (\ w.\ zi - b) \geq 1 \quad ,\ 1 \leq i \leq n \qquad \text{---- (2)}$$

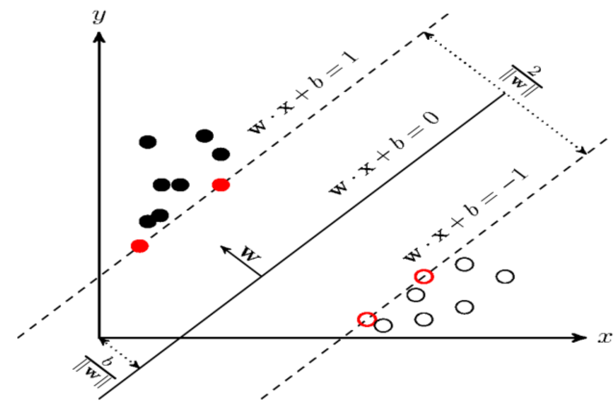

Fig. 2 Parallel hyperplanes having the maxium margin –Widest street approach for the svm.

Data points lying on the separating hyperplanes are known as support vectors and are abbreviated as -SV's and these SVs should satisfies the following equation which ensures that the distance between the hyperplanes is maximum[7].

$$y_j\ [wT\ .\ z_j + b] = 1 \qquad \text{-----(3)}$$

The another important aspect in SVM approach is to consider the most effective separating hyperplane which ensures that the margin is maximum, For the fulfilment of this purpose the concept of Lagrange has been used to minimize the |w| , further this will maximize the distance between the hyperplanes that is the widest street approach. The above optimization case is cracked by the saddle points of Lagrange's Function is shown below [5].

$$L(\alpha) = \sum \alpha_i\ - 1/2 \sum \alpha_i \alpha_j y_i\ y_j z_i T.\ z_j \text{------(4)}$$

α_i - is a Lagrange's Multiplier.
y_i and y_j - these are constants having value 1 and -1.
z_iT. z_j - it is called as vector pair or data sample pair.

This above equation shows that optimization depends on the dot product of the pairs of sample (vector pair)[6].This forms the basis for the support vector machine algorithm which is very important to note.

In situations where data is linearly inseparable, Kernel functions are used [8]. Instinctively, a kernel is just a function that allows the data to be processed or treated more easily. In many datasets, we often have no separating maximum-margin hyperplanes. So, we have to move to a non-linear solution known as Kernel trick which allows to create non-linear classifiers where the algorithm remains same except the vector pairs or data sample pairs are substituted by an appropriate non-linear kernel function. This permits the algorithm to map the data into a feature space in which we can form a separating maximum-margin hyperplane [12].

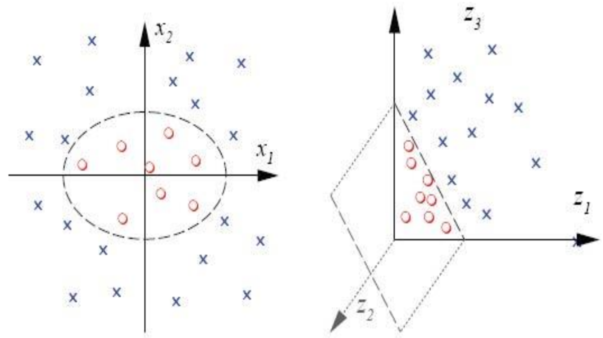

Fig. 3 Separating data in higher dimensional space

Mathematical definition: K(x, y) = < $\phi(x).\phi(y)$>. Here K is the Kernel function, x, y are n dimensional inputs where ϕ is a map from a-dimension to b-dimension space. <x,y> denotes the dot product. Usually b is much larger than a.

After introduction of kernel function, the optimisation problem is changed into:

$$L(\alpha) = \sum \alpha_i - 1/2 \sum \alpha_i \alpha_j \; y_i \, y_j \, K(z_i,z_j) \text{-----}(5)$$

Where $K(x_i,x_j)$ is the kernel function.

C. Winedata Analysis:

A total of 2037 samples were taken and studied from the Machine Learning dataset repository from the fields of Portugal [2].Data are collected on 12 different properties of the wine are : fixed acidity, volatile acidity, citric acid, residual sugar, pH, Alcohol, Chlorides, Free Sulphur Dioxide, Total Sulphur Dioxide, Density, Sulphates, Quality etc. Out of the 12 properties Quality is based on sensory data, and the rest are based on chemical properties of the wines including density, acidity, alcohol content etc.

The wine is classified using two of the 12 chemical properties given in the selected dataset namely pH and alcohol because both the features highly influence the wine properties as compared to other features present in the wine dataset. Sensory recognition like aroma, flavour, mouth feel and bitterness greatly depends on the concentration of the alcohol present in the wine composition [11] [4].In addition to alcohol, pH affects the colour, carbon dioxide absorption, stability, ageability and fermentation rate of the wine [9].

III. METHODOLOGY

For selection and evaluation of features following methodology is applied:

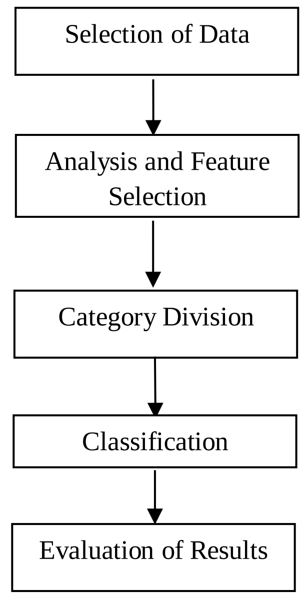

Fig 4 .Methodology

Selection of Data: Data has been selected from the Machine Learning dataset repository which consists of 2037 varieties of data. Only white wine data is chosen and further it has been analysed.

Analysis and Feature Selection: In literature review, it is observed that both pH and alcohol significantly affects the quality of wine.

Category Division: The quality property of wine is rated from 1-10 where minimum quality is rated 3 and maximum as 9 and categorised into three: Average, Good and Bad where the quality measuring 6 is considered to be Average, and quality less than or equal to 5 is taken as Bad and above 6 considered as Good.

Classification: Support Vector Machine (SVM) algorithm is implemented on the selected wine data using Gaussian Radial Basis Function [14] as the data are linearly inseparable.

Evaluation of Results: Results obtained were evaluated on the basis of performance measures such as Accuracy, Precision and Recall.

IV. RESULTS AND DISCUSSION

For evaluating a classifier, a number of performance parameters like Accuracy, Recall and Precision are defined [10].

In our proposed approach, classification is done using Support Vector Machines. For this method, 5-fold cross-validation is used. For evaluating classification results, confusion matrix was obtained for each classification. The results were obtained by classification of three different binary classes i.e. Average-Bad Quality, Good-Bad Quality and Average-Good Quality on the two selected features and subsequently the average of results obtained from all three binary classifier is taken into consideration.

In our work, an accuracy of 72.3% has been achieved with Precision and Recall of 74.3% and 82.6% respectively as shown in **fig. 5**.

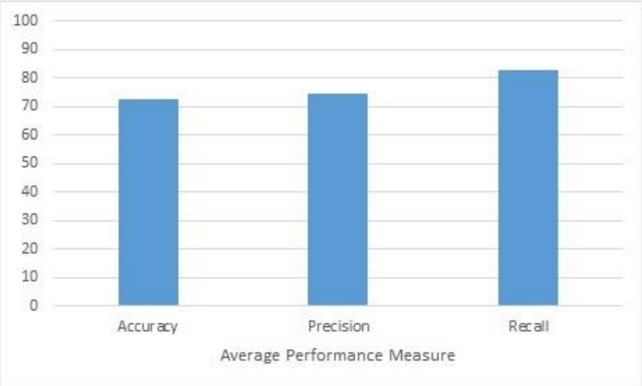

Fig. 5 Average Performance Measure of all classifications on selected features

A brief performance comparison of other algorithms applied on the same dataset has been compared in the **Table I** below

TABLE I: Proposed approach with other techniques

Techniques Applied	Accuracy
Application of Tree Based Models[2]	50%
Proposed Approach	**72.3%**
Application of Random Forest[2]	67.7%
Application of Nearest Neighbour[2]	55%

As it can be distinctly seen in the **Table I**, our approach surpasses the performance parameter like Accuracy compared to other techniques applied. Since the feature selection in our approach is not as complex as in other techniques, our approach also provides real time performance in comparison to other applied techniques.

V. CONCLUSION

In our work, classification of wine quality is done using Support Vector Machines applied on selected features obtained by studying chemical properties of the wine. Among all other applied techniques for wine quality classification, our approach performs better and thus provides an accuracy of 72.3% with next best to an accuracy of 67.7%.

REFEERENCES

[1] C. Cortes and V. Vapnik, "Support-vector networks", Machine Learning, vol. 20, no. 3, pp. 273-297, 1995.

[2] https://onlinecourses.science.psu.edu/stat857/node/223

[3] Kodur, S. (2011). Effects of juice pH and potassium on juice and wine quality, and regulation of potassium in grapevines through rootstocks (Vitis): A short review. Vitis - Journal of Grapevine Research..

[4] Fischer, U.; Noble, A.C. The effect of ethanol, catechin concentration, and pH on sourness and bitterness of wine. Am. J. Enol. Vitic. **1994**, 45, 6–10.

[5] E. Osuna, R. Freund, and F. Girosi, “Training support vector machines:Application to face detection,” in Proc. Computer Vision and PatternRecognition, Puerto Rico, 1997, pp. 130–136.

[6] T. Joachims, “Transductive inference for text classification using support vector machines,” presented at the Int. Conf. Machine Learning,Slovenia, June 1999.

[7] C. J. Burges, “A tutorial on support vector machines for pattern recognition,”Knowledge Discovery and Data Mining, vol. 2, pp. 121–167, June1998.

[8] Liu, L., Shen, B., & Wang, X. (2013). Research on Kernel Function of Support Vector Machine

[9] Boulton, R.B.; Singleton, V.L.; Bisson, L.F.; Kunkee, R.E. (1996) Principles and practices of winemaking. New York: Chapman & Hall: 521– 253.

[10] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks", Information Processing & Management, vol.45,no.4,pp.427-437,2009.

[11] Fontoin, H.; Saucier, C.; Teissedre, P.L.; Glories, Y. Effect of pH, ethanol and acidity on astringency and bitterness of grape seed tannin oligomers in model wine solution. *Food Qual.Preference* **2008**, *19*, 286–291.

[12] B. Boser, I. Guyon and V. Vapnik, "A training algorithm for optimal margin classifiers", Proceedings of the fifth annual workshop on Computational learning theory–COLT' 92, 1992.

[13] Drucker, Harris; Burges, Christopher J. C.; Kaufman, Linda; Smola, Alexander J.; and Vapnik, Vladimir N. (1997); "Support Vector Regression Machines", in Advances in Neural Information Processing Systems 9, NIPS 1996, 155–161, MIT Press.

[14] M. Orr. Introduction to radial basis function networks. Technical report, Institute for Adaptive and Neural Computation, Edinburgh University, 1996.

[15] Bishop, C. M., Pattern Recognition and Machine Learning, Springer, ISBN-13: 978-0-387-31073-2, 2006.