

AHP and MACHINE LEARNING TECHNIQUES for Wine Recommendation

Kunal Thakkar^{#1}, Jay Shah^{#2}, Raghav Prabhakar^{#3}, Anant Narayan^{#4}, Anurag Joshi^{#5}

[#]CE/IT Department, SVKM's NMIMS University
Shirpur, Maharashtra, India

Abstract— In today's modern competitive world it is very necessary and critical for a company to keep its customers loyal and similarly equally important to provide better recommendations to its customers. Hence automated computer tools and in particular machine learning is used to facilitate and enhance recommendations as it is a promising and is assumed to be a modern world trending aspect.

A variety of these techniques, including Support Vector Machines(SVM's) and Random Forest(RF's) have been widely applied in recommending products for the development of predictive models, resulting in effective and accurate decision making.

Our paper aims to use modern and effective techniques like SVM and RF which groups together the dataset and provides deep learning. 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 on the basis of certain attributes.

Keywords— Analytical Hierarchy Process, Wine, wine attributes, Eigen values and Eigen vectors, MATLAB, PCA (Principal Component Analysis), Random Forrest, R Language, Decision Tree.

I. INTRODUCTION

A. Decision Tree

Decision tree [1][2] is a graphical method to show different choices made by individuals or machines in form of tree. The nodes of graph depict the occurrence of an event and edge shows decision paradigms. This approach is used in machine learning and data mining [3], using R packages. Work is done on training data. Then a set of validation data is used for verification of the result thus obtained. Examples of use of decision trees is – predicting the chances of rain at a particular time of day, predicting the chances of road accidents or predicting health degradation on consuming alcoholic beverages.

B. SVM

SVM (Support Vector Machine) [8] is a machine learning algorithm which is used for both classification and regression techniques. Through SVM [9] we plot data items sample space of n-dimensions, where n is features taken into consideration.

The value of each feature gives the value of a particular component. Then classification is done by finding the hyper plane that differentiates two classes very well.

C. Random Forest

Random Forest [4] is a group of items that are studied based on classification and regression technique.

In a normal decision tree one decision tree is built. In a random forest algorithm a number of decision trees are built during the process.

A data frame is classified based on two aspects: -

- Observation
- Variables

Large number of decision trees are made and used so it's called a forest. It's an efficient method of classification of data and estimating missing data. In this method, prototypes are used to give relation between variables and classification. Thus, it's an experimental method which aims at communication between variables. It's a faster technique as compared to other algorithms.

D. AHP

The Analytical Hierarchy Process (AHP) [5][6] is used for arranging and examining complicated problems by mathematical calculations. It's a process that defines multiple attribute issue to advise a particular commodity to an individual. Combinations of separate performance indicators with one of main performance indicator are done so as to allocate different weights to each attribute. The process of AHP [7] is mainly used to calculate weights. The inputs for AHP [10] are alternatives and attributes.

In this paper, we have taken red wine dataset and have allotted weights to them. The result thus obtained after analysis of the data is used for recommending a wine to individuals.

II. METHODOLOGY

We have used various Machine Learning techniques to recommend a wine and have calculated its accuracy as shown below:-

A. AHP(Analytical Hierarchy Process)

In this section we have calculated various attributes of AHP in different ways and have explained the details:-

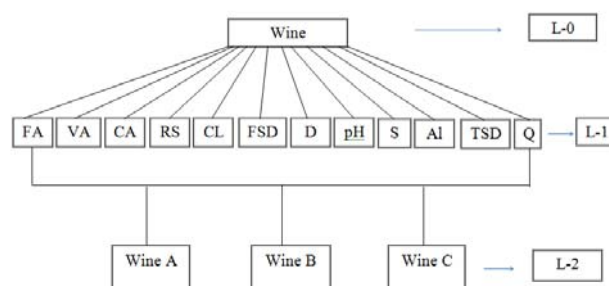


Fig. 1 Hierarchical levels for attributes of wines

Here, FA - fixed acidity, S – sulphates, VA - volatile acidity, A – alcohol, CA - citric acid, TSD - total sulphur dioxide, RS - residual sugar, Q – quality, CL – chlorides, FSD - free sulphur dioxide, D – density, pH - pH value

Criteria	A	RS	D
A	1	3	5
RS	1/3	1	3
D	1/5	1/3	1

Table-1

The Eigen vector values are calculated from the above table 1, where sum of each column is considered and it is divided by individual element followed by addition of rows to get further results.

0.640
0.257
0.103

Table-2

The values calculated above show the weights for Alcohol, Residual Sulphur and Density which are 64 %, 25.7 % and 10.3 % respectively.

The Eigen values (λ) can be calculated for different attributes mentioned in above matrix by using following formula:-

$$\lambda_{\max} = \text{sum of 1}^{\text{st}} \text{ Column} * \text{weight of 1st column} + \text{sum of 2}^{\text{nd}} \text{ Column} * \text{weight of 2nd column} + \text{sum of 3}^{\text{rd}} \text{ Column} * \text{weight of 3rd column} \quad -(1)$$

$$\lambda_{\max} = 1.53*0.640 + 4.33*0.257 + 9*0.103$$

$$\lambda_{\max} = 3.039$$

The Consistency Index is equal to the value of largest Eigen value. It can be obtained by the calculations shown further. Then, CI can be calculated using the formula given below:-

$$CI = (\lambda_{\max} - n) / n-1 \quad -(2)$$

$$CI = 3.039-3 / 3-1$$

$$CI = 0.0584$$

After calculating CI this index can be used by comparing with its appropriate value. The appropriate CI is known as Random Consistency Index (RI).

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49

Table-3

For n (size of the matrix) = 3 and RI = 0.58 (obtained from table 1).

The Consistency Ratio (CR) is calculated as

$$CR = CI / RI \quad -(3)$$

$$= 0.0195/0.58$$

$$= 0.0336$$

$$CR = 3.36\%$$

The CR ratio should be under 10% so that the chosen attribute is a good one. The obtained values of CR justify this condition.

In the second step of our calculations we show the matrix for calculating weights for the wine type alternatives namely Wine A, B and C. The matrix for selecting the best alternative based on the user attributes is as follows:-

Alternative	A	B	C
A	1	2	4
B	1/2	1	5
C	1/4	1/5	1

Table-4

The Eigen vectors obtained from the above matrix are:-

0.54
0.36
0.1

Table-5

The above obtained values show weights for wine types A, B and C which are 54 %, 36 % and 10 % respectively.

The Eigen value (λ_{\max}) can be calculated by using equation (1) for table 2 as given below:-

$$\lambda_{\max} = \text{sum of 1}^{\text{st}} \text{ Column} * \text{weight of 1st column} + \text{sum of 2}^{\text{nd}} \text{ Column} * \text{weight of 2nd column} + \text{sum of 3}^{\text{rd}} \text{ Column} * \text{weight of 3rd column.}$$

$$\lambda_{\max} = 1.75*0.543 + 3.2*0.363 + 10*0.1$$

$$\lambda_{\max} = 3.094$$

These values can also be verified through MATLAB as shown in figure below:-

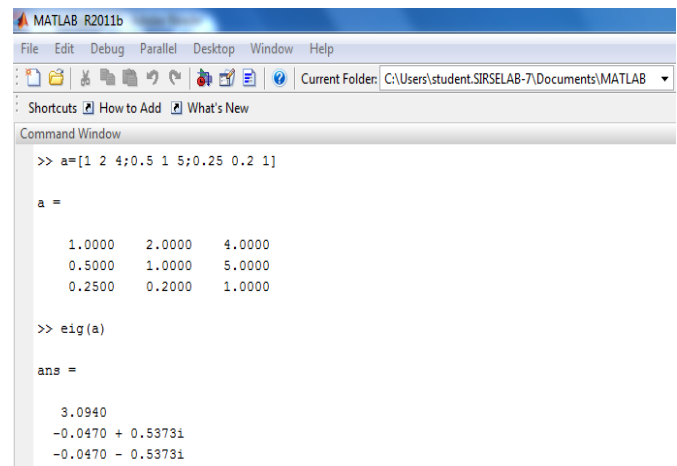


Fig. 2 Use of MATLAB to verify AHP results for Wine Types

B. SVM (Support Vector Machine)

The results obtained through the use of SVM technique[10] are shown below:-

```

Console C:/Users/jayshah/Desktop/
number of classes: 8
Levels:
3 4 5 6 7 8
10-fold cross-validation on training data:
Total Accuracy: 66.54159
Single Accuracies:
74.21384 63.75 63.125 63.75 66.25 69.375 66.875 66.875 62.5 68.75

> model1$tot.accuracy
[1] 66.54159
    
```

Fig. 3 Total Accuracy of SVM using R

Total accuracy = 0.6654

The next step shows the graph between the attributes:-

1) Volatile Acidity vs. Fixed Acidity:

SVM classification plot

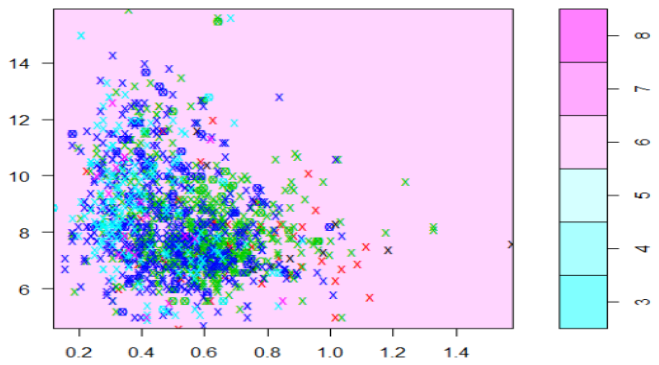


Fig.4 SVM plot for VA vs. FA

2) Residual Sugar vs. Citric Acid:

SVM classification plot

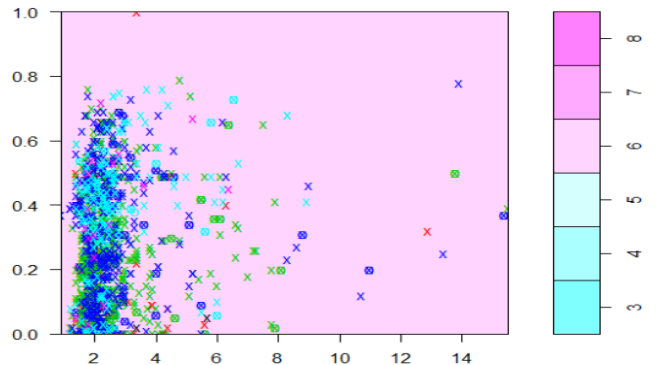


Fig. 5 SVM plot for RS vs. CA

3) Sulphur Dioxide vs. Chlorides:

SVM classification plot

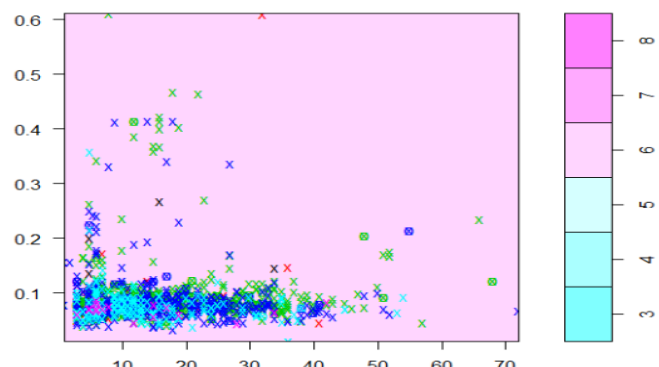


Fig. 6 SVM plot for SD vs. Cl

4) PH vs. Density:

SVM classification plot

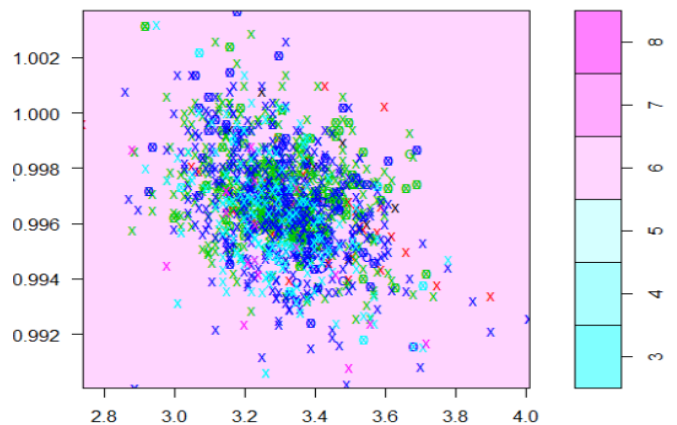


Fig.7 SVM plot for pH vs. Density

5) Alcohol vs. Sulphates:

SVM classification plot

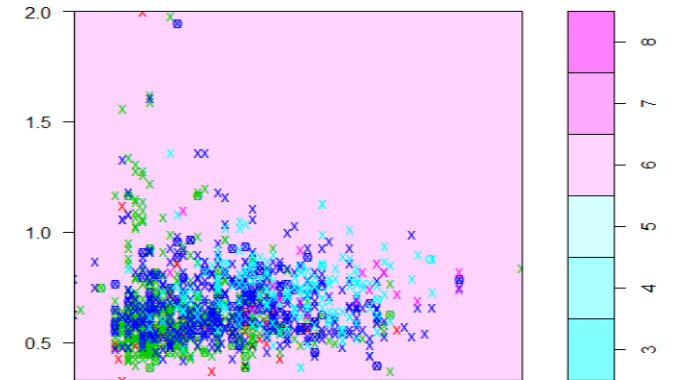


Fig. 8 SVM plot for A vs. S

C. RF (Random Forest)

The Random Forest method gives a total accuracy as shown in the fig. below:-

prediction	3	4	5	6	7	8
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	1	6	132	34	5	0
6	1	3	45	115	20	2
7	0	0	2	7	21	2
8	0	0	0	0	3	1

[1] 0.6725

prediction	3	4	5	6	7	8
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	4	6	140	28	0	0
6	1	1	34	117	12	4
7	0	0	4	13	34	1
8	0	0	0	0	0	1

[1] 0.73

prediction	3	4	5	6	7	8
3	0	0	0	0	0	0
4	2	0	2	0	0	0
5	1	11	151	33	1	0
6	0	6	28	111	19	1
7	0	0	2	8	21	2
8	0	0	0	0	1	0

[1] 0.7075

Fig. 9 Accuracy using confusion matrix

$$\text{Total accuracy} = (0.6725 + 0.73 + 0.7075) / 3 = 0.7033$$

III. RESULT ANALYSIS

In this section we explain the results obtained by performing AHP, SVM and RF respectively. The AHP process yields the following results:-

Types of Wines

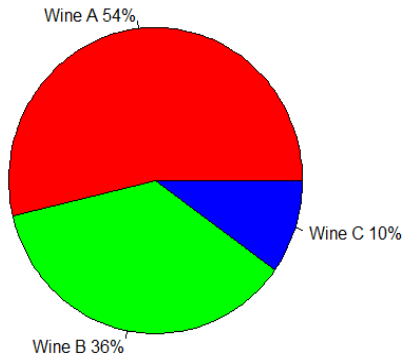


Fig. 10 Wine selection using AHP

The various Machine Learning languages yield a different output of accuracies. SVM and RF technique give a better accuracy depending on the data set. Here RF gives a better accuracy because it works well on larger data sets compared to SVM. The two techniques can be differentiated on the bases of following criteria: -

Criteria	SVM	RF
Result	Solution is unique	Solution is random
Accuracy	0.6654	0.7033
Working	Works well in case of small data sets	Works well in case of large data sets
Over fitting	Reduces over fitting with the use of regularization	Reduces over fitting with the use of optimal number of trees

Table-6

IV. CONCLUSIONS

The Analytical Hierarchy process is a well known classification algorithm. We have used this technique in this paper to recommend wine on the basis of its

components. The results obtained show the Wine selection on the basis of its attributes. The Machine Learning Techniques used here help in finding the component accuracy of wine attributes. The obtained results show that accuracy of RF is better as compared to SVM which is approximately 70.33% while that of SVM is 66.54%. Hence proper usage of Machine Learning techniques will help in finding the proofs in order to recommend a particular wine keeping in mind its contents.

V. FUTURE WORK

In order to make it more concrete we are going to use effective technique called Ensemble as well as Deep learning in our upcoming work.

REFERENCES

- [1] Shahruckh Teli and Prashasti Kanikar, *A Survey on Decision Tree Based Approaches in Data Mining*, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 4, 2015.
- [2] J.R. Quinlan, *Induction of Decision Trees*. Centre for Advanced Computing Sciences, New South Wales Institute of Technology, Sydney 2007, Australia.
- [3] Brijendra Singh and Hemant Kumar Singh, *WEB DATA MINING RESEARCH: A SURVEY*, *IEEE International Conference on Computational Intelligence and Computing Research, IEEE International Conference on Computational Intelligence and Computing Research, 2010*.
- [4] Leo Breiman, *Random Forests* Article DOI: 10.1023/A:1010933404324, Volume 45, Issue 1, pp 5–32, *October 2001*.
- [5] Lan Gan, Xuehu Wang and Rong Li, *Notice of Retraction Research and Implementation of AHP-Based Method Base - Model Base Application of Hierarchical Model*, Knowledge Engineering and Software Engineering, 2009. KESE '09. Pacific-Asia Conference, 19-20 Dec. 2009.
- [6] Thomas L. Saaty, *Group Decision Making and the AHP*, Graduate School of Business, University of Pittsburgh, Pittsburgh, Pennsylvania, Publisher : Springer Berlin Heidelberg
- [7] Nathasit Gedsri and Dundar F. Kocaoglu, *Applying the Analytic Hierarchy Process (AHP) to build a strategic framework for technology road mapping*, Published in *Journal Mathematical and Computer Modelling: An International journal*, Volume 46 Issue 7-8, Elsevier Science Publishers B.V. Amsterdam, The Netherlands, The Netherlands, October 2007.
- [8] DURGESH K. SRIVASTAVA and LEKHA BHAMBHU, *DATA CLASSIFICATION USING SUPPORT VECTOR MACHINE*, *Journal of Theoretical and Applied Information Technology*, 2010.
- [9] Emir Kremic and Abdulhamit Subasi, *Performance of Random Forest and SVM in Face Recognition*, *The International Arab Journal of Information Technology* VOL. 13, NO. 2, March 2016.
- [10] Suhas Machhindra Gaikwad, Dr. Preeti Mulay, Rahul Raghvendra Joshi, *Analytical Hierarchy Process to Recommend an Ice Cream to a Diabetic Patient based on Sugar Content in it*, 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15), 2015