Original Research Paper

Smart Agri Wine: An Artificial Intelligence Approach to Predict Wine Quality

Gaurang S Patkar and D. Balaganesh

Department of Computer Science and Multimedia, Lincoln University College, Kuala Lumpur, Malaysia

Abstract: Lately the item quality has been one of the critical parts of each and every industry. The customary strategies for surveying the item quality are exceptionally tedious and furthermore not having the ideal outcome with the resultant in unique Technology development. Through the ideas of Artificial Intelligence (AI) and Information Science (IS) it is more productive to evaluate or to foresee any sort of thing effectively. In this study, the investigation of wine information is done on University of California Irvine (UCI) Machine Learning (ML) dataset. The fundamental motivation behind this examination is to foresee wine quality dependent on physicochemical information using AI.

Keywords: Artificial Intelligence, Machine Learning, University of California Irvine, Information Science

Introduction

Wine is the most ordinarily utilized refreshment internationally and its qualities are viewed as significant in the public eye. The nature of the wine is consistently significant for its customers and essentially for makers in the present cutthroat market to raise income. Generally, wine quality used to be controlled by testing toward the finish of the creation; to arrive at that level, one as of now invests bunches of energy and cash. With the rise of ML techniques and their success in the past decade, there have been various efforts in determining wine quality by using the available data (Li *et al*., 2017; Shanmuganathan, 2016; Riul Jr *et al*., 2004). The different modelling techniques to improve the quality of food using AI show how AI is expanding sanitation and quality activities in food industry (Sahni *et al*., 2021). The new system which is developed works on Real Time with a merged function of RFID and GPS to sustain the detailed information about farmers and E-Rickshaw (Tyagi *et al*., 2021). On the off chance that the quality isn't acceptable, the different strategy should be carried out all along, which is exorbitant. Each individual has their own assessment on taste, so distinguishing a quality dependent on an individual's taste is testing. With the improvement of innovation, the makers began to depend on different gadgets for testing being developed stages. In this way, they can have a superior thought regarding wine quality, which, obviously, sets aside heaps of cash and time. Likewise, this aided in gathering heaps of information with different boundaries, for example, the amount of

various synthetic substances and temperature utilized during the creation and the nature of the wine delivered. The analysis of the basic parameters that determine the wine quality is very important. Human experts may introduce error while predicting the quality of wine while tasting because it is very subjective and depends on person's choice. In addition to humanitarian efforts, ML can be an alternative to identify the most important parameters that control the wine quality. In this study, it is shown how ML can be used to identify the best parameter on which the wine quality depends and in turn predict wine quality.

Literature Review

Today, fluctuated purchasers appreciate wine to an ever increasing extent. The wine business is investigating new advances for both wine making and offering measures to back up this development (Cortez *et al*., 2009). Physicochemical and tangible tests are utilized for assessing wine accreditation (Ebeler *et al*., 1999). The segregation of wines is certainly not a simple cycle inferable from the intricacy and heterogeneity of its headspace. The arrangement of wines is vital in view of various reasons. These reasons are the financial worth of wine items, to ensure and guarantee the nature of wines, to disallow contaminated of wines and to control refreshment preparing (Preedy *et al*., 2016). This study shows the viability of group AI dependent on a democratic technique (Moritani and Takefuji, 2018). Three AI calculations were exclusively assessed utilizing the

Red-wine dataset. The democratic strategy arranges wine quality by casting a ballot weighted calculations. Exploratory outcomes showed that the proposed casting a ballot strategy is superior to any individual AI techniques for the red wine grouping issue. Dahal *et al.* (2021) demonstrated how statistical analysis can be used to identify the components that mainly control the wine quality prior to production.

In this study, the creators have utilized nonlinear classifiers just as probabilistic classifiers to characterize various wines by accomplishing great grouping accuracies (Aich *et al.*, 2019). The creators proposed another structure that joined MF-DCCA with XG Boost and Light GBM. The proposed new structure depends on target tests and along these lines it tends to be coordinated into a choice emotionally supportive network, improving the speed and nature of the oenologist execution (Ye *et al.*, 2020). In this exploration, a half breed model that comprises of two classifiers in any event, for example the irregular woodland, support vector machine, is proposed for wine quality expectation (Dua D. and Graff C., 2019).

Experimental Setup

Initially the wine data is collected from Kaggle (Cortez *et al.* 2009). In this data there are 1599 records. Each record contains 11 different features of wine. The next step is to check for null values, duplicate values etc. After data cleaning, data visualization helps in clearly explaining each feature in wine dataset. Random forest classifier is used to identify patterns and relationships in features. Lastly we train and test the model. The algorithm for wine quality prediction is as follows:

Step1: Import libraries (numpy and pandas)
Step2: Load kaggle data set
Step3: Separate dataset into features and labels
Step4: Split data for testing (20% and training 80%)
Step5: Perform normalization
Step6: Train decision tree classifier
Step7: Check efficiency

Splitting the dataset into test and train data can vary. The train data is used to train the model for predicting wine quality. In every Machine Learning program, there are two things, features and labels. Features are the part of a dataset which are used to predict the label.

Materials and Methods

Figure 1 shows the block diagram of wine quality prediction using Artificial Intelligence, where X is the number of features, F is the machine learning model, f (XS) is the output prediction and XS is one of the input features. To build the machine leaning model, the data is split into training and testing data using train-test split ratio. The training data is fed to Random Forest Regressor. The first step is to import the python libraries. Load the wine data and add headers. Check for missing data in dataset. This is very important to get correct results. The last step is to find out the quality of wine.

To begin with, verify the dissemination of the quality variable in the dataset. Make sure that enough great quality wines are available in the dataset. Then, the relationships between the factors are checked. This permits improving comprehension of the connections between the factors in a speedy impression. Quickly, it is seen that there are a few factors that are emphatically connected to quality. Almost certainly, these factors are additionally the main highlights in the AI model. The primary thing to do is normalize the information Normalizing the information implies that it will change the information so its dissemination will have a mean of 0 and a standard deviation of 1. It's essential to normalize the information to even out the scope of the information.

The UCI Machine Learning vault is utilized as a dataset for trial purposes. This dataset was utilized for research by (Cortez *et al.*, 2009). It has two classes of wine information. One class contains data on red wines and the other contains information of white wines. To begin with, required libraries are imported for information examination. The detail of information in the dataset is displayed in Table1.

Table1 shows few rows and 12 unique credits of wine in the dataset. The last section gives the nature of the wine. The quality characteristic takes esteems between 1 to 10, 1 method inferior quality wine and 10 are for acceptable quality wine. The dataset may contain some missing qualities. The examination of information in the dataset is done to check for missing qualities. The yield has shown that the information is legitimate and no missing qualities in the dataset. The python library is utilized for execution. To comprehend the informational collection appropriately more extra data is needed for itemized examination. After starting examination and point by point study, it is seen that all wines in the dataset with quality value more than 7 will be of good quality, wine with quality values between the range of 5 and 6 are of average quality and qualities fewer than 5 are of bad quality. This data will be exceptionally helpful for the dynamic process. The quality of wine information is displayed in Table 2. Table 3 shows the statistics of wine data. The total number of records is 1599.

Figure 2 shows graph of quality v/s total number of data in dataset. It is clearly seen that most wines fall under mid-range quality.

There are fewer wines of top caliber and enormous tasting and not very many wines that are of inferior quality. Further investigation of the information is fundamental for additional assessment of value.

Table 1: Different features of wine dataset

Sr.no.	Fixed acidity	Volatile Acidity	Citric aid	Residual sugar	Chlorides	Free sulfur dioxide	Total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9.8	5
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9.8	5
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9.8	6
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.

Table 2: Quality of wines

Quality values	Total number
>7	217
between 5 and 6	1319
<5	63
% of wine with >7	13.57%

Table 3: Statistics of wine data

|	Fixed acidity	Volatile acidity	Citric acid	Residual sugar	chlorides	Free sulfur dioxide	Density	pH	Sulphates	Alcohol	Quality
Count	1599.0	1599. 0	1599.0	1599.0	1599.0	1599.0	1599.0	1599.0	1599.0	1599.0	1599.0
Mean	8.0	1.0	0.0	3.0	0.0	16.0	1.0	3.0	1.0	10.0	6.0
Std	2.0	0.0	0.0	1.0	0.0	10.0	0.0	0.0	0.0	1.0	1.0
Min	5.0	0.0	0.0	1.0	0.0	1.0	1.0	3.0	0.0	8.0	3.0
25%	7.0	0.0	0.0	2.0	0.0	7.0	1.0	3.0	1.0	10.0	5.0
50%	8.0	1.0	0.0	2.0	0.0	14.0	1.0	3.0	1.0	10.0	6.0
75%	9.0	1.0	0.0	3.0	0.0	21.0	1.0	3.0	1.0	11.0	6.0
Max	16.0	2.0	1.0	16.0	1.0	72.0	1.0	4.0	2.0	15.0	8.0
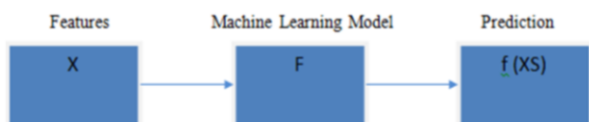
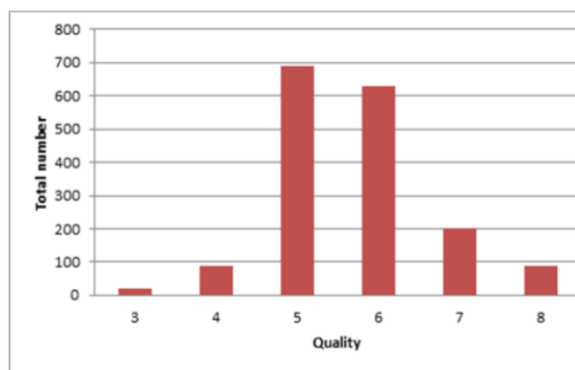**Fig. 1:** Block diagram of wine quality prediction using artificial intelligence

Fig. 2: Quality v/s total number

Python pandas portray technique is utilized to get accommodating measurements, like mean, middle and standard deviation of attributes in the information.

Some helpful measurements that are used:

1. Mean (Average): Here every one of the qualities for a given component are added, by then segment it by the quantity of tests
2. Median: Arrange all the example esteems in mathematical request, in a rundown. The center number in this rundown will be the median
3. Mode: The value that happens the most in a rundown of samples
4. Range: The difference between the most noteworthy worth and the least qualities in a list
5. Standard Deviation: First ascertain the mean, at that point take away each number in the rundown with the mean and square the outcome. At that point ascertain the mean of those squared contrasts; lastly compute its square root

Presently, the accompanying stage is to think about the features in our enlightening assortment in more detail. The idea of wine depends on a ton of compound properties that impact its taste, smell and flavor. In spite of the way that winemaking is seen as workmanship, it's in actuality wonderful sensible taking a gander at the circumstance impartially. Wines contain fluctuating degrees of organic acids, alcohol, sugars, salts from mineral and organic acids, pigments, phenolic compounds, nitrogenous substances, gums, pectins, volatile aromatic compounds, mucilage, salts, vitamins and sulfur dioxide.

Results

In this section, the connections between highlights are investigated. Presently that some area information about wine is known, it's an ideal opportunity to investigate more. The dataset which is used contains a lot of highlights, for example, liquor levels, measure of remaining sugar and pH esteem.

A portion of these highlights may be subject to different highlights, some may not. Some of them may influence our quality appraisals as well. In information science or Artificial Intelligence, it's very critical to consider the highlights that make up the information and notice if there are any co-relations between them.

Physiochemical Parameters

The physiochemical parameters incorporate the assessment of sugar content, pH, acridity content and the alcoholic substance.

pH vs. Fixed Acidity: Here the connection amongst pH and fixed acidity is investigated. For this make information outline containing pH and fixed acidity columns to break down the relationship. The seaborn library is utilized to instate the joint framework with the information outline. Ultimately, show the outcomes as regression plot and appropriation plot in a similar grid. This is displayed in Fig. 3.

Figure 3 it is observed that pH change with changing fixed acidity levels. The increase in fixed acidity levels, there is decrease in pH levels. Figure 2 it is also observed that when pH level is low there is high acidity in the wine.

Volatile Acidity vs Quality

Figure 4 shows the relation between quality and volatile acidity. When wine has less volatile acidity value that means the wine is of very good quality. Maintaining less volatile acidity is very important as it affects the taste and aroma of wine. If volatile acidity is more, then the wine will have low quality.

Alcohol vs. Quality

Figure 5 shows the graph of quality v/s alcohol content. Alcohol content is very important is wine. Figure 4 it is observed that if alcohol content in wine is more, than the wine is of very good quality. Alcohol content in wine also affects the taste, flavor harmony and the body of wine. Wines of high alcohol content are from warm climate.

Discussion

The result of the wine prediction is also tested with human experts and shown in Table 4. It is observed that the accuracy of automated wine quality prediction is more than 90% accurate and also it is in agreement with human expert 90%. To test it more rigorously, more test data will be generated by producing different varieties of homemade local fruit and flower wines.
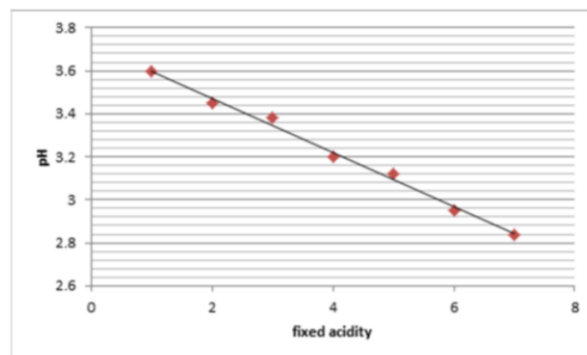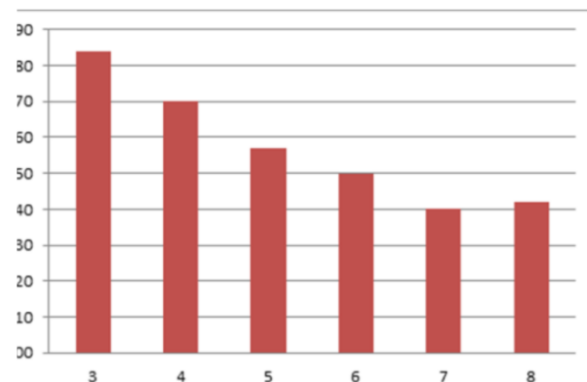

Fig. 3: pH v/s fixed acidity
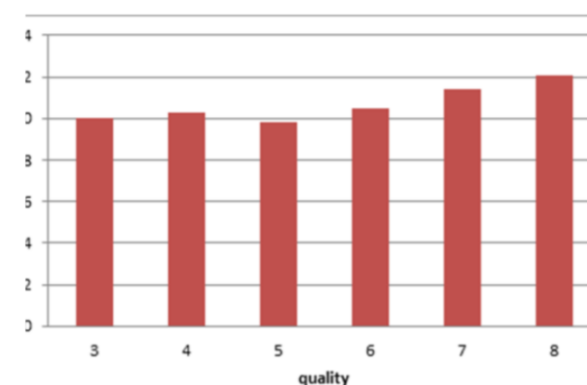

Fig. 4: Quality v/s volatile acidity

Fig. 5: Quality v/s alcohol

Table 4: Test results

Number of wine sample data	Automated prediction	Remark (between manual and auto system)
1599	90%	In agreement 90%

Conclusion

It is believed that many people commonly like wines with high liquor content and are excessively repetitive and eager. Great is normally connected with low degrees of unpredictable causticity. This implies that unpredictable acridity is an indication of a spoiler and can cause an upsetting scent. From this examination, it is presumed that when the wine has less unstable causticity esteem, it implies the wine is of excellent quality. Additionally, when the liquor content in wine is more, the wine is of generally excellent quality. The accuracy of the automated wine quality prediction is more than 90%.

Acknowledgement

This work is based on wine quality prediction using Artificial Intelligence.

Author's Contributions

All authors equally contributed in this work.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Aich, S., Sain, M., & Yoon, J. H. (2019). Prediction of Different Types of Wine Using Nonlinear and Probabilistic Classifiers. In Integrated Intelligent Computing, Communication and Security (pp. 11-19). Springer, Singapore. doi.org/10.1007/978-981-10-8797-4_2

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. Decision support systems, 47(4), 547-553. doi.org/10.1016/j.dss.2009.05.016

Dahal, K. R., Dahal, J. N., Banjade, H., & Gaire, S. (2021). Prediction of Wine Quality Using Machine Learning Algorithms. Open Journal of Statistics, 11(2), 278-289. doi.org/10.4236/ojs.2021.112015

Dua, D., & Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

Ebeler, S. E. (1999). Linking flavor chemistry to sensory analysis of wine. In Flavor Chemistry (pp. 409-421). Springer, Boston, MA. doi.org/10.1007/978-1-4615-4693-1_35

Li, H., Zhang, Z., & Liu, Z. (2017). Application of artificial neural networks for catalysis: A review. Catalysts, 7(10), 306. doi.org/10.3390/catal7100306

Moritani, M., & Takefuji, Y. (2018). Ensemble machine learning based on a voting method using red wine dataset for business education, Hawaii int. conf. on education, Honolulu.

Preedy, V., & Mendez, M. L. R. (2016). Wine Applications with Electronic Noses, Electronic Noses and Tongues in Food Science, Cambridge, MA, USA: Academic Press, (137-151).

Riul Jr, A., de Sousa, H. C., Malmegrim, R. R., dos Santos Jr, D. S., Carvalho, A. C., Fonseca, F. J., ... & Mattoso, L. H. (2004). Wine classification by taste sensors made from ultra-thin films and using neural networks. Sensors and Actuators B: Chemical, 98(1), 77-82. doi.org/10.1016/j.snb.2003.09.025

Sahni, V., Srivastava, S., & Khan, R. (2021). Modelling Techniques to Improve the Quality of Food Using Artificial Intelligence. Journal of Food Quality, 2021. doi.org/10.1155/2021/2140010

Shanmuganathan, S. (2016). Artificial neural network modelling: An introduction. In Artificial neural network modelling (pp. 1-14). Springer, Cham. doi.org/10.1007/978-3-319-28495-8_1

Tyagi, N., Khan, R., Chauhan, N., Singhal, A., & Ojha, J. (2021). E-Rickshaws Management for Small Scale Farmers using Big Data-Apache Spark. In IOP Conference Series: Materials Science and Engineering (Vol. 1022, No. 1, p. 012023). IOP Publishing. https://iopscience.iop.org/article/10.1088/1757-899X/1022/1/012023/meta

Ye, C., Li, K., & Jia, G. Z. (2020, November). A new red wine prediction framework using machine learning. In Journal of Physics: Conference Series (Vol. 1684, No. 1, p. 012067). IOP Publishing. https://iopscience.iop.org/article/10.1088/1742-6596/1684/1/012067/meta