# A Hybrid Wine Classification Model for Quality Prediction

Terry Hui-Ye Chiu, Chien-Wen Wu, and Chun-Hao Chen(✉)

Department of Information and Finance Management, National Taipei University Technology, Taipei 106, Taiwan
terry.h.chiu@gmail.com, {xcwwu,chchen}@ntut.edu.tw

**Abstract.** "Wine is bottled poetry" a quote from Robert Louis Stevenson shows the wine is an exciting and complex product with distinctive qualities that make it different from other products. Therefore, the testing approach to determine the quality of the wine is complex and diverse. The opinion of a wine expert is influential, but it is also costly and subjective. Hence, many algorithms based on machine learning techniques have been proposed for predicting wine quality. However, most of them focus on analyzing different classifiers to figure out what the best classifier for wine quality prediction is. Instead of focusing on a particular classifier, it motivates us to find a more effective classifier. In this paper, a hybrid model that consists of two classifiers at least, e.g. the random forest, support vector machine, is proposed for wine quality prediction. To evaluate the performance of the proposed hybrid model, experiments also made on the wine datasets to show the merits of the hybrid model.

**Keywords:** Machine learning · Decision tree · Random forest · Support vector machine · Wine quality prediction

## 1 Introduction

Wine has always been an essential part of the dinning culture in western countries. From the manufacturer point of aspect, understanding the quality of the wine and creating a steady production is an important goal for the industry. However, testing the quality of the wine is complex and diverse. The wine quality is evaluated in terms of subtlety and complexity, ageing potential, stylistic purity, varietal expression, ranking by experts, or consumer acceptance. By excluding the controllable object measures, the views of experts are very subjective because it may cause the most considerable influence on both winemakers and consumers [1].

Recording the steps of wine production procedure is to preserve the quality and knowledge of the whole winemaking process, and the collected information is the best tool to guarantee the wine quality. Currently, the wine industry has established the protected designation of origin (PDO) system [2] with the support of analytical chemistry and chemometric tools to obtain information related to a specific wine. With the improvement of technology both in software and hardware, winemakers started to use

the collected data to improve the winemaking technique. Due to the high cost and lack of technological resources, it was difficult for most of the wine industries to classify the wines based on the chemical components. Many algorithms based on machine learning to assess the quality of wine have then been gained much attention for the wine industry to determine what attributes make a "good" wine that the consumers can satisfy with them. For instance, Yeo et al. focused on predicting the wine price using a machine learning technique by using past historical wine price data [3]. For wine production, Ribeiro et al. utilized the linear regression, neuron network and decision tree for predicting the wine vilification [4].

In 2009, Cortez et al. collected a wine quality dataset which consists of significant larger instances [6]. Then, three machine learning models, including multiple regression, support vector machine (SVM) and neuron network (NN), are trained using the collected wine dataset. It shows that SVM outperforms the other two methods, and indicates the importance of the correct setting of hyperparameters. Over the years, the wine dataset has been adopted in several studies with various methods such as SVM [7–10] , random forest (RF) [11–14], decision-tree-based algorithms [12,14], and NN [4,7,8] to predict the quality of the wine based on physiochemical characteristics in the wine.

However, the past literature mostly focused on using or comparing different machine learning models that can provide the best prediction result for specific datasets. To get a more effective classifier, in this paper, a hybrid model that consists of two classifiers at least, e.g. the random forest, support vector machine, is proposed for wine quality prediction. To evaluate the performance of the proposed hybrid model, experiments also made on the wine dataset to show the merits of the hybrid model.

## 2 Background Knowledge

Over the years, several different machines learning models are used to predict wine quality. The literature suggested that the LR and SVM provide better results than other models. In this section, the two commonly used classifiers are described.

### 2.1 The SVM Classifier

The support vector machine (SVM) [19] is a supervised machine learning model for solving a classification problem. The central concept of SVM is utilized the kernel function to find the hyperplane that can separate instances into categories. As mentioned earlier, the SVM has proven to be an effective classifier for wine quality prediction [7–10]. There are three hyperparameters in SVM that include penalty factor C, parameter gamma $\gamma$ and kernel function in SVM. The parameter C is a regularization parameter that controls the trade-off between maximizing the margin and minimizing the training error. The gamma parameter defines how far the influence of a single training example reaches. The final parameter is a kernel; there are three different main types of kernels linear, poly and rbf, which may fit best with the different dataset. Hyperparameter tuning relies more on experimental results than theory, and therefore the best method to determine the optimal settings is by trial and error.
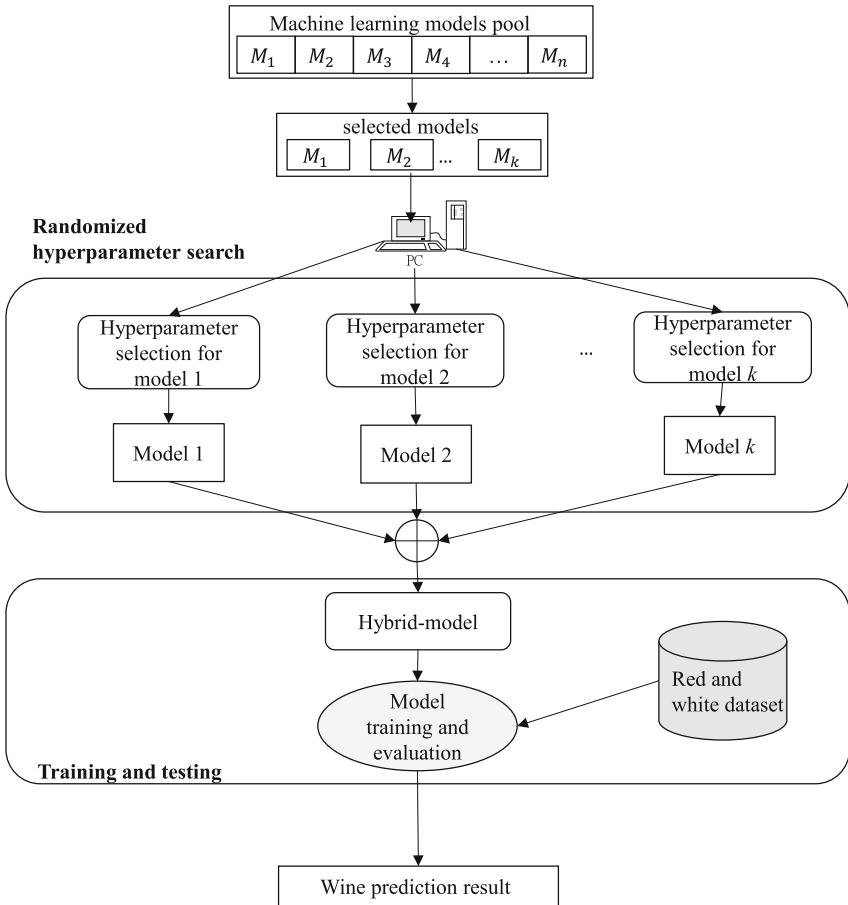
## 2.2   Random Forest

Random forest (RF) [20] is a supervised learning algorithm, but different from SVM used on both classification as well as regression. Several studies have shown that using random forest can provide a good prediction accuracy, and one study showed an extremely high accuracy rate [14]. In this study, we used for classification problems. Based on the name, RF constructed from different trees and more trees means more robust forest. In general, the RF algorithm creates different decision trees on random data samples and then gets the prediction from each of the trees. Next step is to use a voting technique to select the best solution. It has an advantage over other methods because it reduces the over-fitting by averaging the result. In this study, we will tune the RF model to provide the best result. The random forest has six hyperparameters: (1) No. of estimators (2) Maximum features (3) Maximum depth (4) Minimum samples split (5) Minimum samples leaf and (6) bootstrap. Tuning hyperparameters can improve the accuracy of the model. However, if evaluating the model only on the training set can lead to overfitting.

# 3   Proposed Hybrid Wine Classification Model

In this paper, the goal is to predict the quality of wine using the hybrid wine classification model that composes of multiple classifiers. The flowchart of the proposed model is shown in Fig. 1.

From Fig. 1, it first selects the classifiers from the machine learning models pool. The hyperparameter of selected classifiers is them determine by the randomized search method. The selected models are gathered as a hybrid classification model. Then, the input red and white data sets are used to train and test the model. This paper attempts to provide a hybrid wine classification model that produces the best performance. The pseudo-code of the proposed model is illustrated in Table 1(Algorithm 1).

In Table 1, the algorithm firstly selected at least two models (line 2). When the selection process is complete, the initial ranges of the hyperparameters associated with each model are set (line 3). For example, for SVM model ($M_0$), the hyperparameters as $p^m = \{$C: [1, 100, 10000], gamma: [0.1 0.01, 0.0001] and kernel: ['rbf', 'linear', 'poly']$\}$ are initially set. The hyperparameters for each model is different, and there is no specific appropriate range of value for any particular model. Therefore, the trial and error strategy is used in the algorithm in order to find the proper range of values that can provide a better model performance. For example, the hyperparameters are fitted to the model, and the model is then evaluated using the predefined criteria (mainly accuracy). Base on the performance, it decides whether the range values should be added or removed from the initial setting (line 6). After the modification process, we compare the performance of new setting against past settings (line 7). If the new setting indeed finds "optimal" solution, then the process of trial and error can be interrupted. Otherwise, it will continue the process until reaching the required iteration setting (lines 4–9). Continue using SVM as an example, the hyperparameter setting after the trial and error process is $p^m = \{$C: [500, 1000, 10000], gamma: [0.01 0.001, 0.0001] and kernel: ['rbf']$\}$. This range set of hyperparameters are passed to the next procedure. That is to find the best set of hyperparameter for SVM by using the random grid search method (line 11). After the tuning procedure, the selected models are merged to form a hybrid model $M^{new}$ (line 16).

Machine learning models pool

$M_1$ | $M_2$ | $M_3$ | $M_4$ | ... | $M_n$

selected models

$M_1$ | $M_2$ | ... | $M_k$

**Randomized hyperparameter search**

PC

Hyperparameter selection for model 1

Hyperparameter selection for model 2

...

Hyperparameter selection for model $k$

Model 1

Model 2

Model $k$

Hybrid-model

Model training and evaluation

Red and white dataset

**Training and testing**

Wine prediction result

**Fig. 1.** Flowchart of the hybrid wine classification model

The model $M^{new}$ is then trained and tested for $n$ times with different training and testing data for each iteration (line 18). The criteria used to evaluate the models are accuracy, precision, recall and f1-score. From past studies, prediction techniques like SVM and RF are commonly used and have better results. Therefore, in this paper, we use those two models as the selected models and the dataset collected by [6] as our testing dataset. Experimental results are shown in the next section.

## 4   Experimental Evaluation

### 4.1   Dataset Description

The wine dataset from the UCI database [6] that consists of two sets of wine data (red and white) is used in this paper. The red wine contains 1599 instances, and white wine contains 4898 instances. Both datasets contain 11 physiochemical variables, including

**Table 1.** Pseudo-code of the hybrid wine classification model.

| Algorithm 1: The hybrid wine quality prediction model |
| --- |
| Input: Dataset $D$, Machine learning model $M$ (for illustration purpose use assume $M_0$ as SVM and $M_1$ as random forest) |
| Parameters: Hyperparameter for model $M$ as $P_M$, trial and error iteration $j$, training iteration $n$ |
| Output: The hybrid model $M^T$ |

```
1.  Procedure hybrid_algorithm(){
2.  {M₀, M₁, …} ← Select at least 2 models from M
3.  for m = 0 to (i - 1){
4.      pᵐ ← pickInitParameters(Mₘ)
5.      for iter = 0 to j {
6.        pᵐ ← addRemoveModifyParameters(Mₘ, pᵐ)
7.         bool goal ← testModel(Mₘ, pᵐ)
8.         if goal
9.            break
10.     }
11.     pᵐ: {p₀, p₁, …, pₘ} ← RandomizedGridSearch(Mₘ)
12.   }
13.   for i = 0 to n {
14.     Separate D into Dₜᵣₐᵢₙ and Dₜₑₛₜ
15.     for m = 0 to (i - 1){
16.       Mⁿᵉʷ ← MergeModel(Mᵢ, Mⁿᵉʷ)
17.     }
18.     Mᵢᵀ ← trainAndTestModel(Mⁿᵉʷ, Dₜᵣₐᵢₙ, Dₜₑₛₜ)
19.   }
20.   Return Mᵀ
21. }
```

fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, Sulphates, and alcohol. The output data (sensory data) is a quality rating from 0 (very bad) to 10 (excellent). We carefully examined the data to make sure no anomaly exists. As a result, for red wine, there are 24 duplicate instances, and 937 duplicate instances for white wine. However, those are not classified as an anomaly because all features and label values are precisely the same.

### 4.2 Performance Measure Metrics

The four criteria, including accuracy, precision, recall and f1-score, used to evaluate the performance of the model, are described in the following. The accuracy is to measure how often the classifier correctly classifies instances. The formulation is defined as follows:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{1}$$

where TP (true positive) means the samples in the data which classified as belonging to the correct class, TN (true negative) means the samples in the data which classified as not belonging to the expected class correctly, FP (false positive) means the samples in the data which classified as belonging to the expected class incorrectly, and FN (false negative) means the samples in the data which classified as not belonging to the expected class, incorrectly.

In a multi-class condition, the micro and averaged accuracy, precision, and recall are always the same. Also based on past works, it is obvious the wine data is imbalanced, hence only using accuracy may not provide a clear picture. Therefore, the macro-averaging measurement for precision (macro-precision) and recall (macro-recall) are employed for a more detailed comparison.

The precision measures the percentage of the relevant results. When precision is one, it means the algorithm's prediction is perfect. The macro-precision will be lower than average precision for the model performing poorly on the rare classes even it performs well on specific classes. Since the measurement value can tell another story, hence it can be a complementary metric. Macro-averaging precision is performed by first computing the precision of each class, and then taking the average of all precisions.

$$Macro - PrecisionMP_{Classifier} = \frac{\sum_c P_c}{number of classes} \tag{2}$$

The recall measures the percentage of total relevant results correctly classified by the algorithm. When the recall is one, it means that all truly positive samples were predicted as the positive class. Similar to micro-precision, the value will be lower if rare classed performed poorly.

$$Macro - RecallMR_{classifier} = \frac{\sum_c R_c}{number of classes} \tag{3}$$

Accuracy is useful when the class distribution in the dataset is even, but F1-score is a better metric when the dataset has imbalanced classes. Hence, we also used it as a criterion to evaluate the performance of the model, and F1-score is defined as follows.

$$Macro - F1 F1_{classifier} = \frac{\sum_c F1_c}{number of classes} \tag{4}$$

### 4.3   Experimental Analysis

Since most of the past works mainly focus on accuracy, therefore, we compared the accuracy of the proposed model against others. In addition, because most works set the training and testing datasets ratio to 80/20, we also set the same ratio for comparison. For comparison, we included the work from Cortez et al. [6] and Apalasamy et al. [17] for performance comparison. The accuracy of each model is shown in Table 2.

In Table 2, the accuracies of the proposed model for red and white wines are 0.66 and 0.67. The results reveal that the proposed model performed slightly better than other models. Also, like most models, the performance of white wine is slightly higher than

**Table 2.** Comparison of different models in terms of accuracy.

| Models | Accuracy | |
|---|---|---|
| | Red wine | White wine |
| Cortez et al. [6] | 0.45 | 0.51 |
| Apalasamy et al. [17] | 0.62 | 0.65 |
| Proposed model | **0.66** | **0.67** |

**Table 3.** Result for different testing data ratio

| Testing dataset size percentage | Red wine | | | White wine | | |
|---|---|---|---|---|---|---|
| | *10%* | *20%* | *30%* | *10%* | *20%* | *30%* |
| Accuracy | 0.69 | 0.71 | 0.66 | 0.70 | 0.68 | 0.67 |
| Macro-precision | 0.42 | 0.43 | 0.34 | 0.42 | 0.37 | 0.47 |
| Macro-recall | 0.41 | 0.38 | 0.32 | 0.41 | 0.34 | 0.35 |
| Macro-F1 score | 0.41 | 0.39 | 0.32 | 0.41 | 0.34 | 0.36 |

red wine. The experiments were then made to examine further the performance of the proposed model under different training and testing data ratio. The results of different testing size for red and white wine of the proposed model are shown in Table 3.

For red wine, the accuracy was at the highest when testing dataset ratio set at 20%. The macro-precision and recall are low across different ratio. It means the amount of false positive is very close or equal to a false negative. The F1 score for red wine gradually decreases with the increase in ratio. The white wine shows a different result, where the precision is always higher than recall. When the ratio was set at 10%, the accuracy and F1 score are at the highest. Also, the macro-precision and recall are at the closest. The low macro-F1 score for red and white wine indicates the data is highly skewed on some classes.

## 5   Conclusions and Future Work

This paper has proposed a hybrid wine classification model for quality prediction, which is unlike most past works focusing on which machine learning models provide the best performance in predicting wine quality. The proposed algorithm first selects n models from the given model pool. Then, the hyperparameters are then searched by the randomized search method. The models with acceptable performances are merged as the hybrid model. Experiments were done on the real dataset that contains red and white wines, and the results indicated the proposed hybrid model is effective in terms of accuracies when compared to other existing approaches. In the future, we will continue to design an algorithm that can be used to obtain both the hybrid models and hyperparameters for any wine dataset based on evolutionary algorithms.

# References

1. Cardebat, J.-M., Livat, F.: Wine expert rating: a matter of taste? Int. J. Wine Bus. Res. **28**, 43–58 (2016)
2. Canizo, B.V., Escudero, L.B., Pellerano, R.G., Wuilloud, R.G.: 10 – Quality monitoring and authenticity assessment of wines: analytical and chemometric methods. In: Quality Control in the Beverage Industry, Grumezescu, A.M., Holban, A.M., (eds.), pp. 335–384. Academic Press (2019)
3. Yeo, M., Fletcher, T., Shawe-Taylor, J.: Machine learning in fine wine price prediction. J. Wine Econ. **10**(2), 151–172 (2015)
4. Ribeiro, J., Neves, J., Sanchez, J., Delgado, M., Machado, J., Novais, P.: Wine vinification prediction using data mining tools. In: International Conference on European Computing Conference, Tbilisi, Georgia (2009)
5. Andonie, R., Johansen, A.M., Mumma, A.L., Pinkart, H.C., Vajda, S.: Cost efficient prediction of Cabernet Sauvignon wine quality. In: IEEE Symposium Series on Computational Intelligence, pp. 1–8 (2016)
6. Cortez, P., Cerdeira, A., Almeida, F., Matos, T., Reis, J.: Modeling wine preferences by data mining from physicochemical properties. Decis. Support Syst. **47**(4), 547–553 (2009)
7. Gupta, Y.: Selection of important features and predicting wine quality using machine learning techniques. Procedia Comput. Sci. **125**, 305–312 (2018)
8. Lingfeng, Z., Feng, F., Heng, H.: Wine quality identification based on data mining research. Int. Conf. Comput. Sci. Educ. 358–361 (2017)
9. Bhattacharjee, S., Chaudhuri, M.R.: Understanding quality of wine products using support vector machine in data mining. Prestige Int. J. Manag. IT-Sanchayan **5**(1), 67–80 (2016)
10. Er, Y., Atasoy, A.: The classification of white wine and red wine according to their physicochemical qualities. Int. J. Intell. Syst. Appl. Eng. 23 (2016)
11. Trivedi, A., Sehrawat, R.: Wine quality detection through machine learning algorithms. In: International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering, pp. 1756–1760 (2018)
12. Shaw, B., Suman, A.K., Chakraborty, B.: Wine quality analysis using machine learning. In: Mandal, J.K., Bhattacharya, D. (eds.) Emerging Technology in Modelling and Graphics. AISC, vol. 937, pp. 239–247. Springer, Singapore (2020). https://doi.org/10.1007/978-981-13-7403-6_23
13. Hu, G., Xi, T., Mohammed, F., Miao, H.: Classification of wine quality with imbalanced data. In: IEEE International Conference on Industrial Technology, pp. 1712–1217 (2016)
14. Aich, S., Al-Absi, A.A., Hui, K.L., Lee, J.T., Sain, M.: A classification approach with different feature sets to predict the quality of different types of wine using machine learning techniques. In: International Conference on Advanced Communication Technology, pp. 1–2 (2018)
15. Kumar, S., Agrawal, K., Mandan, N.: Red wine quality prediction using machine learning techniques. In: International Conference on Computer Communication and Informatics, pp. 1–6 (2020)
16. Mahima, G.U., Patidar Y., Agarwal, A., Singh, K.P.: Wine quality analysis using machine learning algorithms. In: The Micro-Electronics and Telecommunication Engineering, Lecture Notes in Networks and Systems (2020). https://doi.org/10.1007/978-981-15-2329-8_2
17. Appalasamy, P., Mustapha, A., Rizal, N., Johari, F., Mansor, A.: Classification-based data mining approach for quality control in wine production. J. Appl. Sci. **12**, 598–601 (2012)
18. Petropoulos, S., Karavas, C.S., Balafoutis, A.T., Paraskevopoulos, I., Kallithraka, S., Kotseridis, Y.: Fuzzy logic tool for wine quality classification. Comput. Electron. Agri. **142**, 552–562 (2017)

19. Vapnik, V.N.: An overview of statistical learning theory. IEEE Trans. Neural Netw. **10**(5), 988–999 (1999)
20. Liaw, A., Wiener, M.: Classification and regression by random forest (2007)
21. Bergstra, J., Bengio, Y.: Random search for hyper-parameter optimization. J. Mach. Learn. Res. **13**, 281–305 (2012)