



Ensemble Framework for Red Wine Quality Prediction

Rajni Mohana¹ · Parth Sharma¹ · Aman Sharma¹

Received: 16 May 2022 / Accepted: 29 July 2022 / Published online: 19 August 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

As the industrial revolution took place, civilisations and humans are evolving at a faster pace than ever seen before. To catch up with the increased supply demands, goods are being made artificially which are also increasing the profits. As a result, the quality of food is deteriorating which has led to the increased risk of severe health problems in human being. We now need a system to forecast the quality of the drink and meal we are consuming. In this research paper, we have focused on red wine quality. The dataset contains important features such as alcohol, residual sugar, density. Different measures were performed to evaluate our proposed framework such as precision and sensitivity. Our proposed framework attained an accuracy of 98.36% which outperformed previous literature work.

Keywords Red wine quality prediction · Ensemble learning · Skewness · Hyperparameter tuning · Stacking · Class imbalance

Introduction

According to the OIV, global wine consumption in 2020 would be 234 million hectolitres (Mhl). In comparison to 2019, this is a 3% (7 Mhl) decrease. Consumption has decreased for the third year in a row. It is at its lowest point since 2002 (Karlsson 2020). The USA, France, and Italy are the top three wine-consuming countries. Portugal, Italy, and France are the three countries with the highest per capita wine consumption (Karlsson 2020). A diet high in low-quality foods raises your risk of chronic diseases, whereas a diet high in high-quality foods protects you (Renee 2010). Certain types of cancer are even influenced by diet. According to the World Health Organization, a nutritious diet reduces the risk of malignancies such as colon, breast, and kidney (Diet, nutrition, and the prevention of chronic diseases: report of a Joint WHO, FAO Expert Consultation (World Health Organization 2003)). Every year, an estimated 600 million

individuals—about one in every ten people in the globe—become unwell after eating contaminated food, and 420,000 die, resulting in the loss of 33 million healthy life years (DALYs) (World. Food safety 2020). Alcohol usage at the age of 15 predicted weekly alcohol consumption and alcohol intake exceeding the prescribed level 4 years later. The increased alcohol intake of young teenagers was not a passing fad. It was a pattern that continued throughout young adulthood, putting the teenagers at a higher risk of becoming long-term, large-scale consumers (Andersen et al. 2003). At the age of 19, at least 80% drank alcohol monthly, and 24% of men and 11% of women used alcohol in excess of the prescribed national limits, i.e., 21 weekly units of alcohol for males and 14 for women. Use of alcoholic drinks at the age of 15 increased the likelihood of weekly alcohol consumption at the age of 19 (odds ratio [OR]—values ranging from 1.11 to 3.53). Drunkenness among 15-year-old boys and spirit use among 15-year-old girls showed the best predictive connection with excessive consumption at age 19 (OR = 2.44, confidence interval [CI]: 1.38–4.29, respectively, OR = 1.97, CI: 1.15–3.38) (Andersen et al. 2003). Excessive alcohol intake is associated to a number of undesirable outcomes, including being a risk factor for diseases and health effects, criminality, traffic accidents, and, in some cases, alcohol dependence. Each year, 2.8 million people die prematurely due to alcohol usage around the world (Ritchie and Roser 2018). For hundreds of years, red wine has been a component of

✉ Aman Sharma
amans.3008@gmail.com

Rajni Mohana
rajni.mohana@juitsolan.in

Parth Sharma
psharm28@stevens.edu

¹ Department of Computer Science & Engineering, Jaypee University of Information Technology, District Solan, Wahnaghat, Himachal Pradesh 173234, India

social, religious, and cultural gatherings. Monasteries in the Middle Ages believed that their monks lived longer because they drank wine on a regular, moderate basis. According to a report published in 2018 (Golan et al. 2019), drinking red wine in moderation has positive ties with trusted source; although there are no official guidelines around these advantages, drinking red wine in moderation has positive linkages with cardiovascular disease, atherosclerosis, hypertension, certain types of cancer, type 2 diabetes, neurological disorders, metabolic syndrome. Red wine, which is created from crushed black grapes, is a good source of resveratrol, a natural antioxidant found in grape skin (Abu-Amero et al. 2016). Antioxidants help the body fight oxidative stress. Many diseases, including cancer and heart disease, have been linked to oxidative stress. Fruits, nuts, and vegetables are just a few of the antioxidant-rich meals available. Whole grapes and berries have more resveratrol than red wine, and because of the health hazards associated with alcohol consumption, receiving antioxidants from food is likely to be healthier than drinking wine. To receive enough resveratrol, people may need to drink a lot of red wine, which may cause more harm than good. When it comes to alcoholic beverages, however, red wine may be more beneficial than others. Whole grapes and berries have more resveratrol than red wine, and because of the health hazards associated with alcohol consumption, receiving antioxidants from food is likely to be healthier than drinking wine. To receive enough resveratrol, people may need to drink a lot of red wine, which may cause more harm than good (Smith 2020).

The outline of the rest of the paper is as follows: In the second section, we have included the literature review in which we referred to various research works and explained the viability and performance of the different algorithms related to heart disease prediction. In Section 3, we have explained different machine learning algorithms. In Section 4, the proposed framework has been explained in details including model selection, parameter setting, experimental setup & proposed

methodology. In Section 5, performance metrics, comparison of proposed framework with existing machine learning (ML) models and with existing literature is explained. Results are shown with respect to existing models and literature in this section. Section 6 contains the conclusion and future scope.

Literature Review

This section explains the 5 different research work done previously and how they approached the problem and their methodologies. Every year, an estimated 600 million individuals—about one in every ten people in the globe—become unwell after eating contaminated food, and 420,000 die, resulting in the loss of 33 million healthy life years (DALYs) (World. Food safety 2020). As a result, researchers are coming up with different approaches. Few of them are discussed below. In (Kumar et al. 2020a) authors have applied algorithms such as random forest, support vector machine (SVM), and Naive Bayes. Alongside testing accuracy, the author tested training accuracy as well (Table 1).

Machine learning algorithms have revolutionised how data analytics and data mining works. Many researchers since the dataset was made available had used robust models and different metrics to achieve better results. Cortez et al. (2009) used simple multiple regression, support vector machine, and neural networks. On the other hand, Er and Atasoy (2016) used 4 different techniques to experiment with the results, but the models remained the same which included support vector machine, random forest, k-nearest neighbourhood. The first technique was cross-validation, followed by percentage split, cross-validation (after PCA), and percentage split after using PCA. Cross-validation after PCA resulted in highest accuracy among all the methods used. This influenced our research work. Gupta (2018) experimented by selecting few features and discarding few based on the correlation among the variables. This resulted

Table 1 Comparison of existing approaches for wine quality prediction

S. no	Author(s)	Approach	Dataset	Performance metrics
1	(Cortez et al. 2009)	Neural networks, support vector machine, multiple regression	UCI	Accuracy
2	(Appalasamy et al. 2012)	Naive Bayes, ID3	UCI	Accuracy
3	(Er and Atasoy 2016)	KNN, support vector machine	UCI	Accuracy
4	(Gupta 2018)	Neural networks, support vector machine	UCI	Accuracy
5	(Kumar et al. 2020b)	Support vector machine, random forest, Naive Bayes	UCI	Accuracy, recall, precision
6	(Ye et al. 2020)	XGBoost, LightGBM	UCI	Accuracy, recall, precision
7	(Tingwei 2021)	KNN, active learning	UCI	Accuracy
8	(Ahammed and Abedin 2018)	Linear discriminant analysis, multinomial logistic regression, random forest, support vector machine	UCI (Forina et al., 1991)	Accuracy, recall, precision
9	(Lee et al. 2015)	Decision tree	UCI	Accuracy, precision, recall
10	(Wei 2012)	LAGD hill climbing model	UCI	ROC-AUC

in the most robust model. Kumar et al. (2020b) used three models SVM, Naive Bayes, and random forest while taking all features in account. Ahammed and Abedin (2018) used linear discriminant analysis on red wines and got considerably high precision, recall values. Lee et al. (2015) saw the potential in decision tree as the first bagging method. Wie (2012) reporting was based on ROC-AUC scores. His study was solely based on decision trees. Our study is an advance in red wine quality prediction as we have taken into account the skewness and standardisation of data.

Our Contribution

- The proposed framework consists of stacking-based ensemble learning which adds diversity in the classifier.
- Skewness and Gaussian distribution and class imbalance are addressed.
- Hyperparameter tuning is used in order to select the best parameter for ML model training.
- The performance of the proposed framework is compared with existing literature on the basis of accuracy, precision, sensitivity, precision, and F1 score

Background and Preliminaries

This section explains the various machine learning classification methods that are used in the proposed framework. Before the final ensembling of top performing models, other classifier models were attempted. Ten different classifiers were trained on a training dataset. After the initial training, 4 models were selected based on their accuracy measure.

A. Random Forest

The random forest classifier is made up of a series of tree classifiers, each of which is constructed using a random vector sampled independently from the input vector, and each tree casts a unit vote for the most popular class to categorise an input vector (Breiman 1999). To grow a tree, the random forest classifier utilised in this study uses randomly selected characteristics or a mixture of features at each node. For each feature/feature combination chosen, bagging, a method of generating a training dataset by randomly drawing with replacement N samples, where N is the size of the original training set (Breiman 1996), was employed. Any examples (pixels) are categorised by selecting the class with the highest number of votes from all tree predictors in the forest (Breiman 1999).

B. K-Nearest Neighbour (KNN)

The k-nearest-neighbours (kNN) approach is a basic but effective non-parametric classification method (Hand et al.

2001). To classify a data record t , its k -nearest neighbours are collected, forming a neighbourhood of t . Majority vote among data records in the neighbourhood is commonly used to determine the classification for t , with or without taking distance-based weighting into account. However, in order to use kNN, we must first select an appropriate value for k , and the classification's success is heavily dependent on this number. In certain ways, the kNN approach is influenced by k .

III Support Vector Classifier

SVMs are based on statistical learning theory and try to determine the position of decision boundaries that result in the best class separation (Vapnik 1999). SVMs choose the one linear decision boundary that leaves the biggest margin between two classes in a two-class pattern recognition task when classes are linearly separable. The margin is defined as the total of the distances from the nearest points of the two classes to the hyperplane (Vapnik 1999). Using traditional Quadratic Programming (QP) optimization techniques, this problem of maximising the margin can be solved.

IV Naive Bayes Classifiers (Leung 2007)

Statistical classifiers are Bayesian classifiers. They are capable of predicting class membership probabilities, such as the likelihood that a given sample belongs to a specific class. The Bayesian classifier is based on the theorem of Bayes. The influence of an attribute value on a particular class is assumed to be independent of the values of the other attributes by Bayesian classifiers. This is known as class conditional independence. It is designed to simplify the computation and is hence termed "naive."

E. XGBoost (Smith 2020)

XGBoost is a scalable gradient boosting system that focuses on speed and performance. Intelligent tree penalization, proportionate leaf node reduction, and other randomisation settings make it apart from traditional gradient boosting algorithms.

F. Ensemble Learning (Lappalainen and Miskin 2000)

Ensemble learning learns by executing "base learner" multiple times. The final vote is casted on the hypothesis, and final weights are executed on "meta models." Various types of ensemble technique include bagging and boosting.

G. SMOTEENN (Prati et al. 2004)

This technique is very helpful in solving class imbalance as it generates synthetic data points by SMOTE using the ENN algorithm. Synthetic data points are very different from duplicate points.

H. Gradient Boosting Algorithm (Friedman 2001)

The gradient boosting approach may predict both continuous and categorical target variables. Mean square error (MSE) is the cost function when used as a regressor, and log loss is the cost function when used as a classifier.

Proposed Framework

In this section, we explained model selection criteria and parameter setting of different algorithms used in the building the framework. Experimental setup and proposed methodology have been explained further.

Model Selection

Figure 1 all-inclusive depicts proposed framework for red wine quality prediction. Firstly, we have taken data from UCI Machine Learning Repository (only red wine data), which is explained in Section 3.3. We deleted outliers after thoroughly analysing the data and discovering correlations among other parameters. Then, we split our data into two partitions: training data consisting of 80% of instances and testing data with 20% instances. As Chao Ye et al. (2020) used XGBoost and LightGBM that gave highest accuracy, we included these models after considering every predefined model, and these models helped to increase the overall accuracy. Considering red wine quality prediction literature review, most of the authors who achieved considerable accuracy (Kumar et al. 2020b), Cortez et al., (2009), Er and Atasoy, (2016), Gupta, 2018 (Gupta 2018 Jan)) have used SVM. Naturally, SVM was to be included in model selection section. Various bagging and boosting algorithms are proven to increase the accuracy considerably as we can see in Chao Ye et al. 2020 (Ye et al. 2020), which is why apart from LightGBM and XGBoost other algorithms including gradient boosting algorithm, decision tree, random forest were considered while selecting models for stacked ensemble-based classifier.

Apart from selecting our models using literature survey, hyperparameter tuning is also done to maximise the

accuracy which is discussed in Section 4.2. After our model was built, we used stacked classifier class to perform ensemble learning that uses meta classifiers on specified base learners chosen by us. The aim is to have a diverse set of learners together. Given various classifiers, we choose the one with higher accuracy to be used as one of the base learners.

Parameter Setting

In this section, we discussed various factors that were used to improve the accuracy of our stacked ensemble model. XGB-Classifier, random forest, SVM, gradient boosting classifier were with highest accuracy. These four were chosen as a base learner for the ensemble model. Hyperparameter tuning on the following model was done to further improve the accuracy which, as shown in Table 2, can in turn improve the accuracy of stacked classifiers. Random state is 42 throughout.

Experimental Setup

Dataset

The data was retrieved from UCI Machine Learning Repository (Cortez et al. 2009). It contains 11 input variables based on physicochemical tests which include fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulphur dioxide, total sulphur dioxide, density, pH, sulphates, alcohol. The output variable called quality contains variable indicating lowest quality wine with 3 going up to 8 which depicts good quality wine. This makes it a multiclass

Table 2 Hyperparameters used

Model	Learning rate	N estimators	C value	Kernel	Gamma
Gradient boosting	1	300	-	-	-
XGBoost	0.1	500	-	-	-
SVM	-	-	10	rbf	1

Fig. 1 The proposed ensemble framework for wine quality prediction

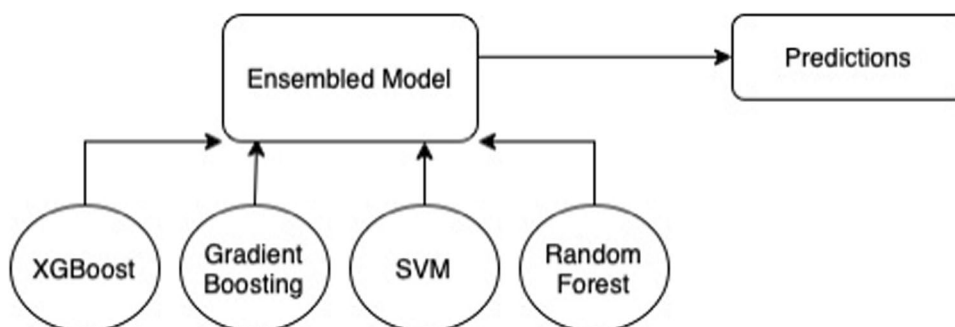


Table 3 Description of nominal attributes

Attribute	Description
Fixed acidity	Can be fixed or non-volatile 4.6 (least acidic) to 15.9 (most acidic)
Volatile acidity	Adds to the unlikable taste 0.12 to 1.58
Citric acid	Increases the originality of wine 0 to 1
Residual sugar	Most wines have dense accumulation of sugar 0.9 to 15.5
Chlorides	The salt content of the wine 0.01 to 0.61
Free sulphur dioxide	Helps shutting of fermentation 1 to 72
Total sulphur dioxide	Healthy in low amount but disagreeable if found high concentration 6 to 289
Density	It hangs on the quality of water 0.99 to 1
pH	Abundantly wines have 3–4 pH 2.74 to 4.01
Sulphates	Helps preventing uneasiness 0.33 to 2
Alcohol	The wine's alcohol content in percentage 8.4 to 14.9
Quality	Variable output (based on sensory data, score between 3 and 8) Class 1: 3 Class 2: 4 Class 3: 5 Class 4: 6 Class 5: 7 Class 6: 8

classification problem. Table 3 gives a detailed description of the attributes given in the dataset.

First of all, class imbalance is analysed in the dataset, as it is a multiclass classification problem, and we are predicting values for each class as essential to address the imbalance by SMOTEEEN. Further, the few features were highly skewed which could have resulted in partial results; as a result, columns whose skewness was greater than 0.75 were corrected by using Power Transformer library. Other than

quality which contains discrete values, every other column contains continuous values.

Data Visualisation

Multicollinearity can drastically affect the prediction of a machine learning algorithm. Multicollinearity affects the precision of computed coefficients, lowering your regression model's statistical power. You might not be able to rely on

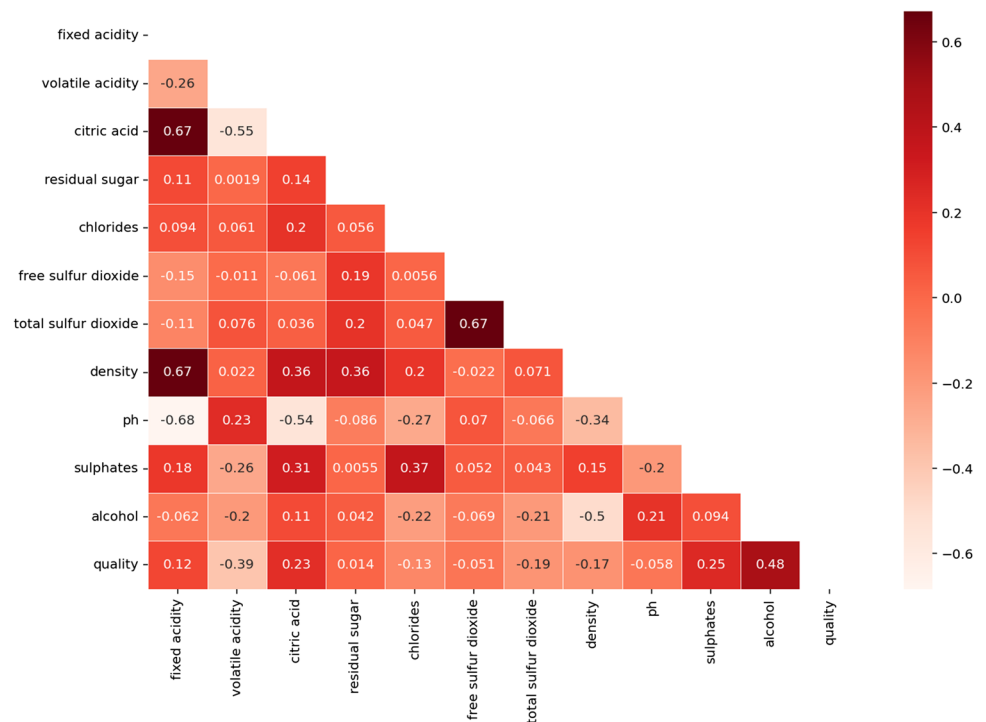
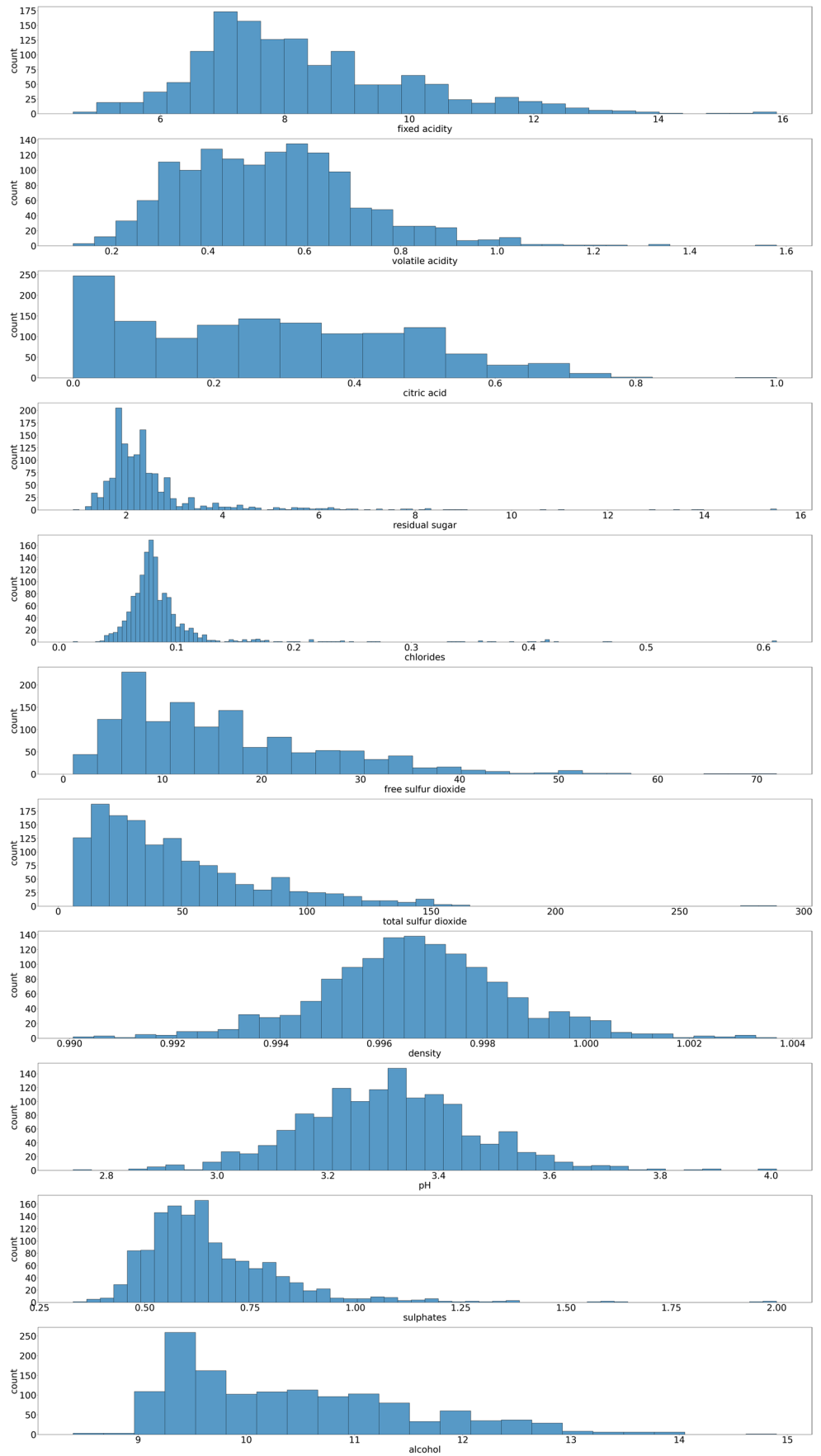
Fig. 2 Multicollinearity of attributes

Fig. 3 Class imbalance (pie chart represents distribution of each class, bar chart—X-axis: total count of target variable; Y-axis: target variable)



p values to detect statistically significant independent variables (Frost 2017). Figure 2 shows multicollinearity between different attributes. As we can see, this dataset is free from multicollinearity, but there is a heavy class imbalance as shown in Fig. 3; we have corrected the imbalance by using SMOTEENN library. As shown in Fig. 4, few attributes are highly skewed which can make the prediction wrong. To correct this, we identified highly skewed columns (whose skewness was greater by 0.75), which were corrected by using Power Transformer library. Another example of class imbalance can be seen in Fig. 5 where we can thoroughly examine the distribution of class among various attributes.

Proposed Methodology

Stacking is a type of ensemble learning where base learners predictions are done on the bases of meta-learner through union of various algorithms. In the proposed algorithm, top performing models are selected from all the algorithms and combined together to give even higher accuracy. The data is taken from UCI Machine Learning Repository, and the data is split into 80% training and 20% testing after correcting class imbalance through SMOTEENN, and skewness is corrected by using Power Transformer library.

The following attributes skewness is corrected because the value was high as 0.75: chlorides, total sulphur dioxide, residual sugar, free sulphur dioxide, sulphates, volatile acidity. Figure 6 shows our proposed methodology as a flow diagram.

Algorithm 1 below describes our proposed methodology which is divided into 3 different phases namely:

1. Phase I: Preprocessing phase
2. Phase II: Training phase

3. Phase III: Testing phase

Preprocessing phase is responsible for data mining. Data visualisation is done to find various insights from the data and make required changes. For example, duplicate rows are removed, and data skewness and class imbalance have been corrected in this phase. Training phase is responsible for finding highest performing models based on accuracy, ensembling them, and learning all the weights and parameters required to make accurate predictions from ensembled framework. The final phase which is the testing phase evaluates the parameters and weights learned on unseen data to analyse our proposed ensembled framework. Symbols used in Algorithm 1 are mentioned in Table 4 given below.

Final outcome of the above stated algorithm classifies the Red wine into 5 categories.

Categories = {3,4,5,6,7,8}.

Phase 1: Preprocessing Phase The dataset which is used in a form of matrix is highly skewed and contains imbalanced class distribution which can lead to inaccurate prediction hence low accuracy and precision. Also, the matrix contains many duplicate rows which can also contribute to skewed predictions. Initially, duplicate rows from the matrix are removed. To address the issue of class imbalance, we have used SMOTEENN technique which generates synthetic data points and Power Transformer library to correct the skewness of attributes. After preprocessing, the dataset is divided into 80% training and 20% testing dataset.

Pseudo Code: Phase I

- Step 1 : Dataset is chosen to be worked on (*Cortez et al. [14]*) (*Only Red wine quality dataset*)
- Step 2 : Duplicate entries from the dataset is removed to reduce the bias in predictions
- Step 3 : As the dataset is highly class imbalanced , SMOTE-EN is performed to correct the imbalance . If not corrected it will lead to low precision and low recall values signifying that the model is poor
- Step 4 : Numerical values are brought down between 0 and 1 to increase the performance and computation time of ML algorithms which will be used in Phase II
- Step 5 : Pre-processed dataset divided into Training and Testing datasets

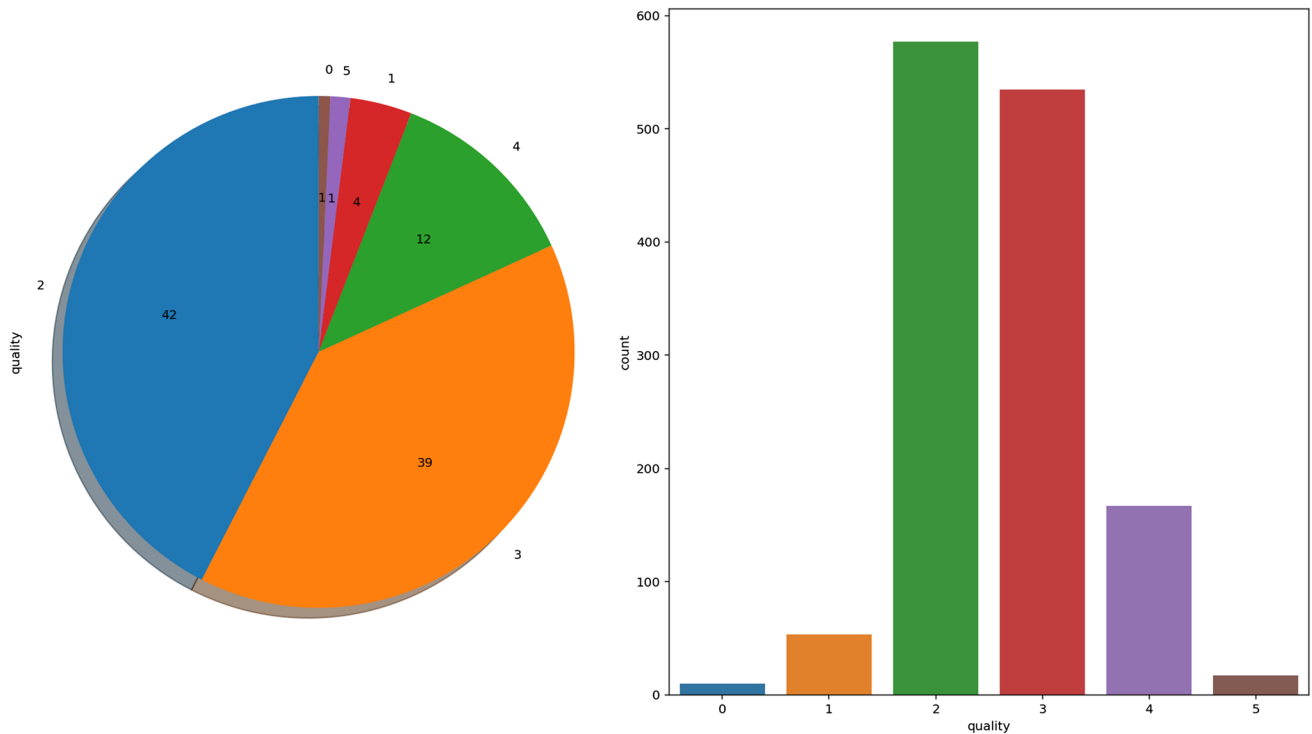


Fig. 4 Skewed attributes (X-axis: total occurrence of an attribute; Y-axis: attribute)

Training Phase After removing duplicates, addressing class imbalance and correcting skewness of the dataset follow. We can now run different machine learning algorithms on 80%

training dataset and judge the algorithms by taking accuracy as a metric. Top performing ML algorithms will be taken into account which can be stacked together.

Pseudo Code: Phase II

Step 1 : 8 different ML algorithms are trained on X and Y . Results are compared based on accuracy and 4 highest accuracy models are chosen for stacking purpose.

Step 2 : Highest accuracy model is chosen out of E to be used as meta-classifier

Step 3 : 4 highest accuracy models are stacked together with meta-classifier in order to perform stacking on X .Weights and parameters are returned from stacked classifier

Fig. 5 Analysing each column using barplot with quality

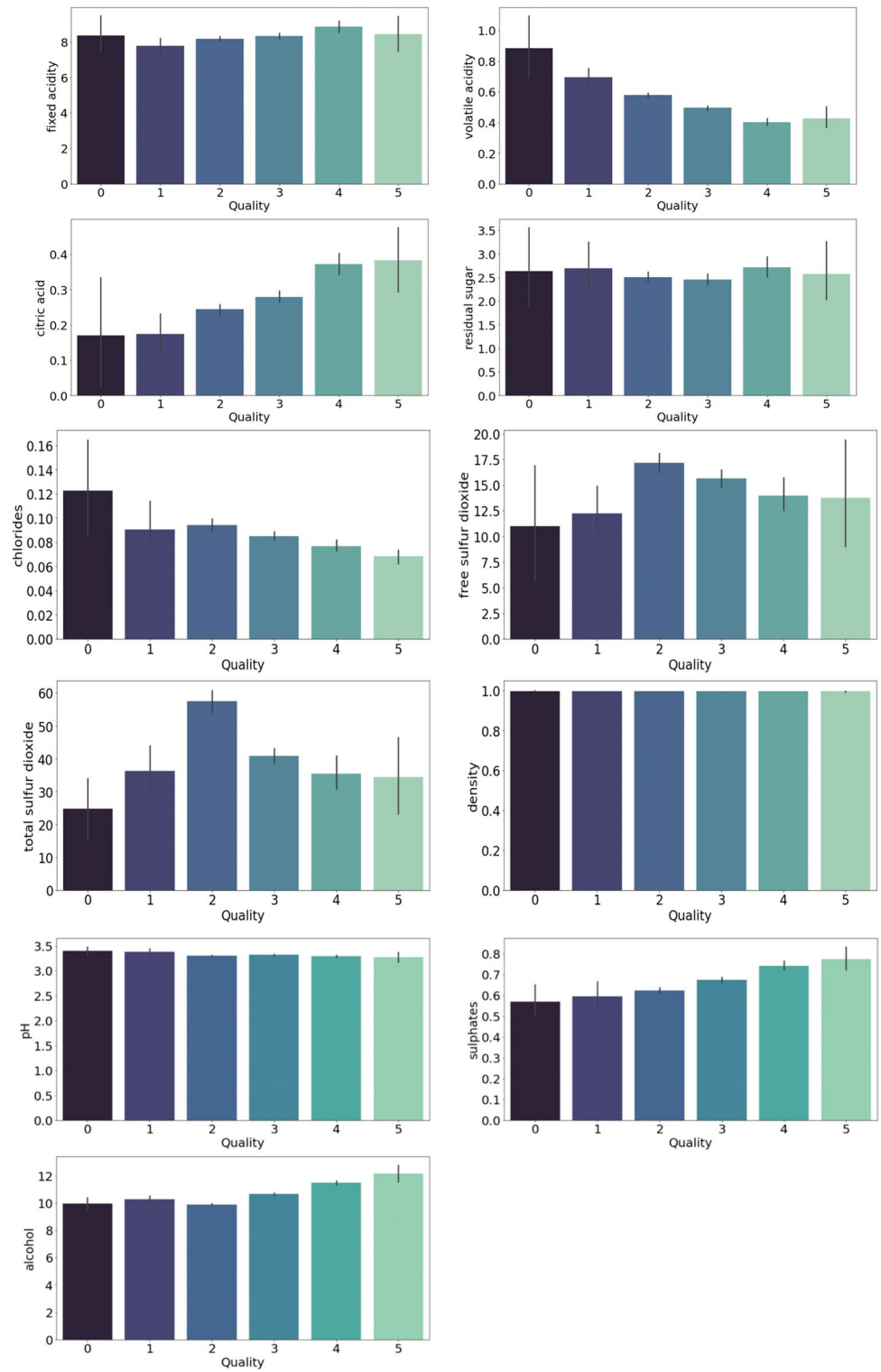
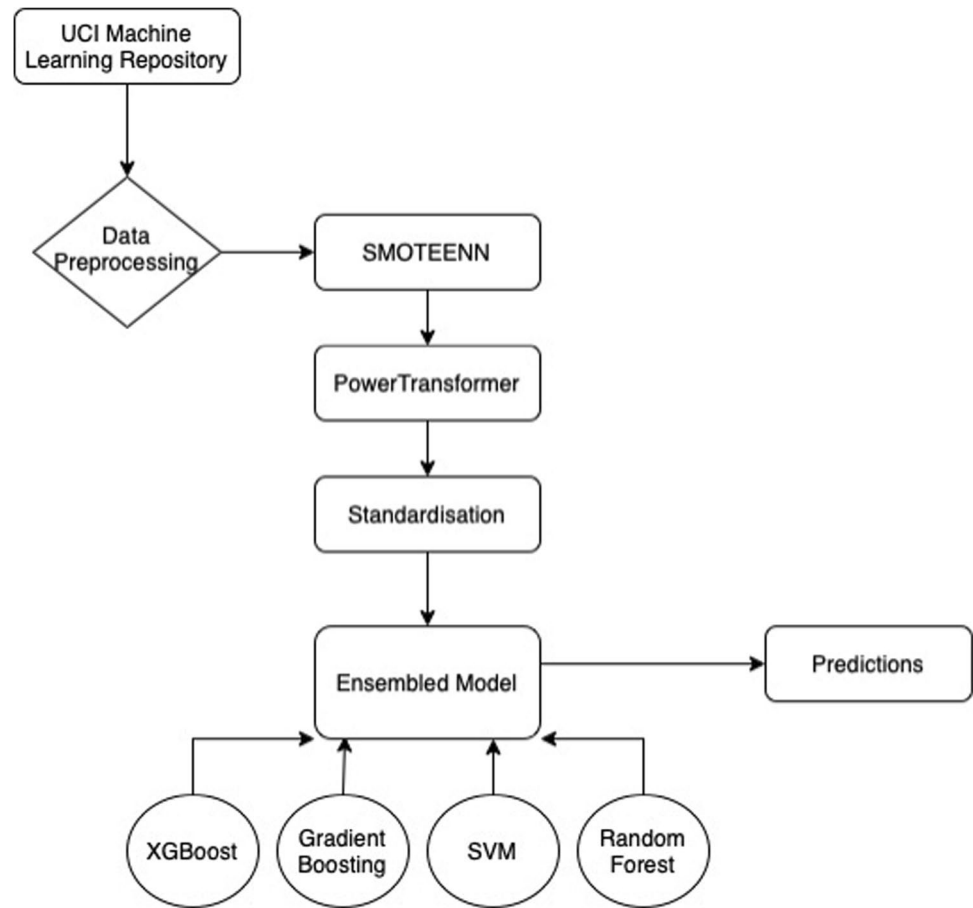


Fig. 6 Proposed methodology



Testing Phase In this phase, stacked model parameters are tested on the 20% testing dataset. Initially, the accuracy is tested; then, classification report is

generated. The accuracy of our proposed algorithm is 98.36% which outperforms the previous literature work.

Pseudo Code: Phase III

Step 1 : Weights and parameters are trained on Y (unseen dataset) and results are obtained based on accuracy

Step 2 : Apart from accuracy difference performance matrices(precision,recall,F1 score,AUC-ROC curve) are also used to compare our results with existing literature

Results and Discussion

In this section, we have discussed the results and analysis of our proposed framework. Different performance metrics have been used to evaluate the algorithms. Further, we

have compared our model with other existing models and its comparison with respect to accuracy, precision, sensitivity, precision, F1 score, ROC, and MCC. We have also discussed proposed models with different algorithms and models covered in Section 2.

Algorithm 1 This algorithm describes the phases to design proposed ensemble model

```

Input : Cortez et al. [14] (Only Red wine quality dataset)

Output : Variable output between 5 classes given in dataset ; Class = {3,4,5,6,7,8}

Begin

Phase I : Let  $S = \{s_1, s_2, s_3, s_4 \dots s_n\}$  be the given dataset

     $S_A = D(S)$  //D(S) removes duplicates and returns dataset with
    unique entries  $S_A$ 

     $S_B = C(S_A)$  //C( $S_A$ ) generates synthetic data points using
    SMOTEEN to combat class imbalance and
    returns balanced classes dataset  $S_B$ 

     $S_C = P(S_B)$  //P( $S_B$ ) implements Power Transformer to correct
    skewness in dataset and it returns corrected
    skewness dataset  $S_C$ 

     $S_D = N(S_C)$  //N( $S_C$ ) brings down the integer value in dataset
    between 0 and 1 to make computations faster thus
    Normalising the dataset and returns Normalised
    dataset  $S_D$ 

    X=80% training dataset  $X \in S_D$  //Preprocessed dataset ( $S_D$ ) is divided into
    training dataset X

    Y=20% testing dataset  $Y \in S_D$  //Preprocessed dataset ( $S_D$ ) is divided into
    testing dataset Y

Phase II : Let  $R = \{r_1, r_2, r_3 \dots r_n\}$  be different ML algorithms

    E= B(R) // B(R) trains the dataset on 8 ML algorithms
    on X then test the weights and parameters on
    unseen testing data Y and returns E which are
    the top 4 algorithms based on their accuracies
    alongside their saved weights and parameters

    L= meta classifier //Highest accuracy model out of E is chosen as
    meta classifier L.It is mandatory to choose a
    meta classifier to perform stacking

    T= X[E(L)] // highest four accuracy achieving models i.e E
    alongside meta classifier L are stacked together
    to generate predictions on known training
    dataset X returning the saved weights and
    parameters T

Phase III : Begin

    P'=T(Y) //Saved weights and parameters runned on
    Testing dataset to check the performance of
    proposed framework and returns predictions
    i.e accuracy P'

    End

End

```

Table 4 Symbols used in Algorithm 1

S. no	Symbols	Meaning
1	S	Attributes of dataset
2	D	Remove duplicate rows
3	C	SMOTEENN
4	P	Power Transformer
5	Z	Normalisation
6	X	Training set
7	Y	Testing set
8	R	ML models
9	B	Selecting top 4 models
10	L	meta classifier
11	T	Stacked model on X
12	P'	Predictions

Table 5 Comparison of ML algorithm and their respective accuracies

S. no	Algorithm	Accuracy
1	XGBoost classifier	97.54%
2	Gradient boosting classifier	96.99%
3	Random forest	95.90%
4	SVM	94.26%
5	K-neighbors classifier	94.26%
6	Decision tree classifier	91.80%
7	Logistic regression	76.50%
8	Naive Bayes	56.27%

Performance Metrics

The following five parameters are used to assess the performance of the proposed framework:

1.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.

$$Accuracy = \frac{(True\ Positive + True\ Negative)}{(True\ Positive + False\ Positive + False\ Negative + True\ Negative)} \tag{1}$$

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

$$Precision = \frac{True\ Positive}{(True\ Positive + False\ Positive)} \tag{2}$$

3.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.

$$Recall = \frac{True\ Positive}{(True\ Positive + False\ Negative)} \tag{3}$$

4.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.

$$F1score = \frac{2 * (Recall * Precision)}{(Recall + Precision)} \tag{4}$$

Comparison with ML Models

We created several baseline models, but the models with highest accuracy were chosen for stacking purposes as ensemble modelling adds more diversity to the predictions. We have chosen accuracy as the metrics to judge models initially. Accuracy of all the models used is shown in Table 5. Top 4 best performing models which include XGBClassifier, random forest, SVM, gradient boosting classifier are stacked together to give even better

Fig. 7 Comparison of accuracy of proposed framework with different ML models

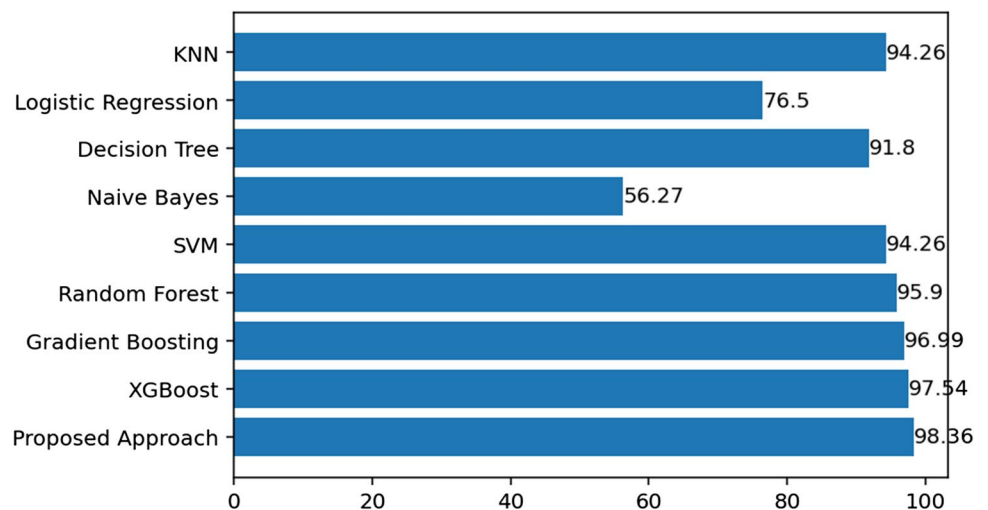
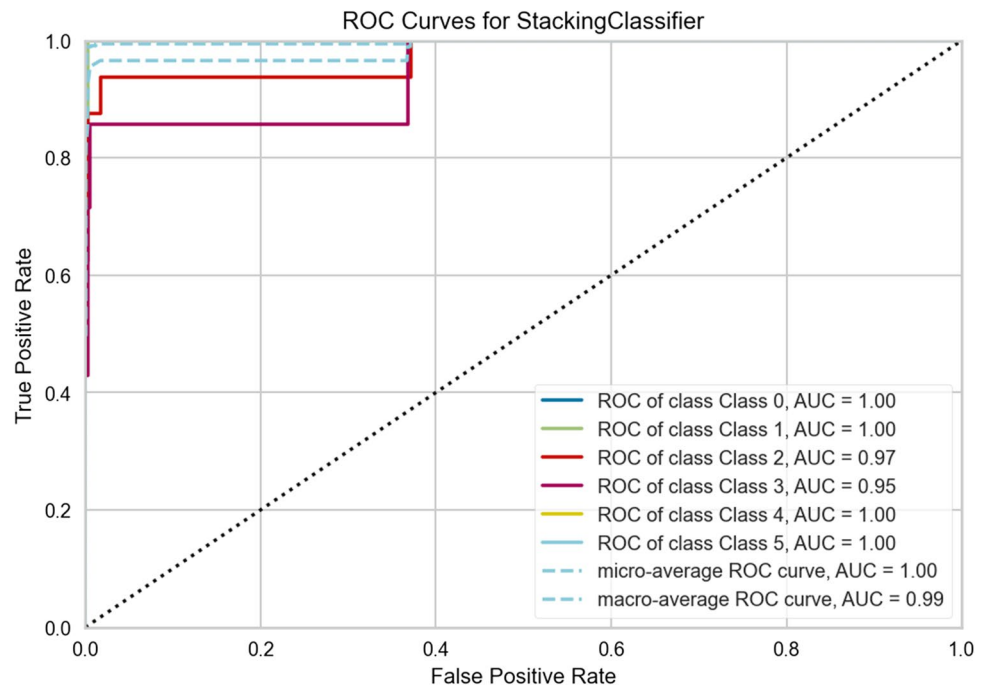


Fig. 8 ROC-AUC comparison for each class



accuracy. We also tested out proposed algorithm on various factors including accuracy, precision, recall, F1 score. Figure 7 graphically represents the comparison of different machine learning algorithms with our proposed algorithm. As we can see, our proposed algorithm outperforms the existing algorithm and previous literature work as shown in Fig. 8. Hence, our work is an advance in red wine classification. We stacked these classifiers to get accuracy of 98.36%. As it is a multiclass classification problem, we got an average of 98.0% precision and 98% recall. As shown in Table 6, besides accuracy we have calculated precision, recall, and F1 score for comparison with other algorithms.

Additionally, to analyse our models, we have built ROC (receiver operating characteristic curve) because it shows the

Table 6 Comparison of proposed framework with existing ML models

Models	Precision	Recall	F1	Accuracy
Random forest	0.96	0.96	0.95	95.90%
Gradient boosting	0.97	0.97	0.96	96.72%
XGBoost	0.98	0.98	0.97	97.54%
Logistic regression	0.75	0.77	0.75	76.50%
KNN	0.94	0.94	0.94	94.26%
SVM	0.94	0.94	0.94	94.26%
Decision tree	0.92	0.92	0.92	92.34%
Chao Ye et al. (2020) (Ye et al. 2020)	88.15	88.67	88.41	91.04%
Proposed methodology	0.98	0.98	0.98	98.36%

tradeoff between specificity and sensitivity for every combination of tests. As we can see in Fig. 9, our proposed algorithm ROC curve is approximately perfect. The better the model, the closer the area of ROC is to 1. As our problem is multiclass classification, ROC curve has taken macro average into account. Figure 8 depicts multiclass classification more comprehensively by plotting the curve for each class through our proposed framework. Ensembled model helped us to add diversity and multiplicity in our model. Further stacked based models add assortment which means if an individual model gives a wrong prediction about a certain feature, another model used in a stacked-based ensemble may have a chance to correctly identify the same feature.

Our work greatly contributes towards food/wine analytics as we are able to classify good and worst quality of wine while outperforming the literature review. This can greatly impact the future research work, which can nearly perfectly predict the correct quality of the food item.

Comparison with Existing Literature

Our proposed algorithm shows a perfect ROC curve and good accuracy, and it can be considered an advance in red wine quality prediction in return, an advance in classifying quality of any other food item. Chao Ye et al. (Ye et al. 2020) used XGBoost which influenced our work. Further, every author used SVM and random forest which heavily influenced our work. As shown in Fig. 10, our proposed methodology outperforms previous literature work done on the dataset. This can be used further in biomedical research

Fig. 9 ROC curve comparison

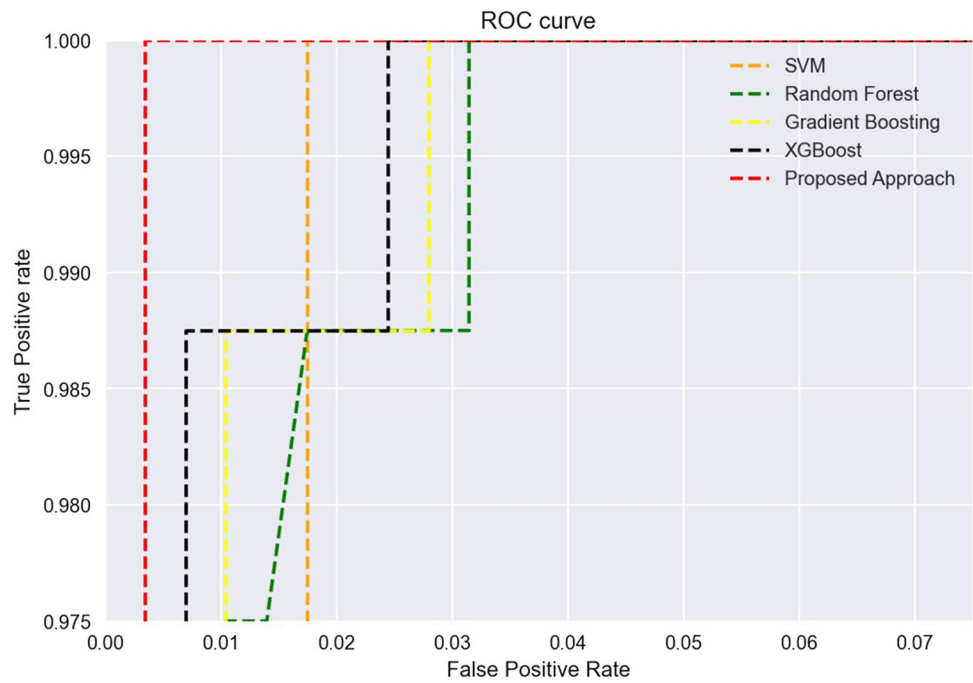
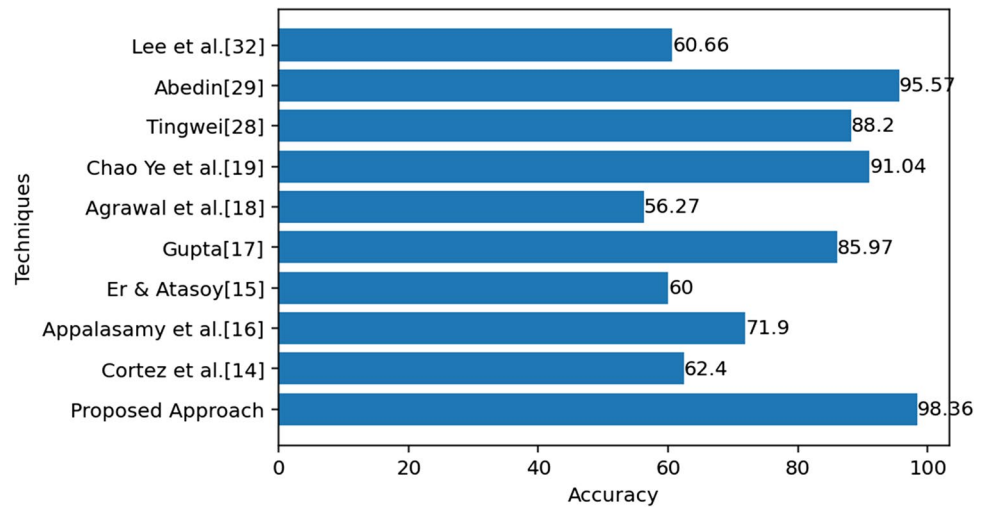


Fig. 10 Comparing proposed model with existing literature



work relating to food/water quality predictions. Applying ensemble-based stacking can benefit further research work as it provides diversity to classifiers and increases other parameters like accuracy and precision.

Conclusion and Future Work

In this paper, we offer a machine learning-based computational framework for predicting red wine quality. The proposed framework successfully sorted red wines into various classifications. The proposed framework’s key

contribution is handling skewed and imbalanced data utilising a power transformer and the SMOTEENN technique. Furthermore, ensemble learning helps to increase variety among the base learners, which increases prediction accuracy. Precision, specificity, recall, and F1 score are used to evaluate the performance of all approaches. On a single benchmark dataset, all algorithms are trained and tested. Most present strategies do not take into account the uneven nature of data and skewed data when creating red wine quality prediction tools. The proposed framework addresses the challenges of class imbalance and data skewness.

Author Contribution RM: Helped in proposed methodology and implementation. PS: Helped in the implementation and writing of the manuscript. AS: Helped in proposed methodology, implementation, and manuscript writing.

Data Availability The data used to support the findings of this study is included within this article.

Declarations

Competing interests The authors declare no competing interests.

Conflict of Interest Rajni Mohana declares that she has no conflict of interest. Parth Sharma declares that he has no conflict of interest. Aman Sharma declares that he has no conflict of interest.

References

- Abu-Amero KK, Kondkar AA, Chalam KV (2016) Resveratrol and ophthalmic diseases. *Nutrients*. 8(4):200. <https://doi.org/10.3390/nu8040200>
- Ahammed B, Abedin M (2018) Predicting wine types with different classification techniques. *Model Assist Stat Appl* 13(1):85–93
- Andersen A, Due P, Holstein BE, Iversen L (2003) Tracking drinking behaviour from age 15–19 years. *Addiction* 98(11):1505–1511
- Appalasaamy P, Mustapha A, Rizal ND, Johari F, Mansor AF (2012) Classification-based data mining approach for quality control in wine production. *J Appl Sci* 12(6):598–601
- Breiman L (1996) Bagging predictors. *Mach Learn* 24(2):123–140
- Breiman L (1999) Random forests—random features. Technical Report 567, Statistics Department. University of California, Berkeley. <ftp://ftp.stat.berkeley.edu/pub/users/breiman>
- Cortez P, Cerdeira A, Almeida F, Matos T, Reis J (2009) Modeling wine preferences by data mining from physicochemical properties. *Decis Support Syst* 47(4):547–553
- Er Y, Atasoy A (2016) The classification of white wine and red wine according to their physicochemical qualities. *Int J Intell Syst App Eng* 1(4(Special Issue-1)):23–6
- Friedman, JH (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat* 1189–1232
- Frost J (2017) Multicollinearity in regression analysis: problems, detection, and solutions. *Statistics By Jim*. Published April 2, 2017. Accessed April 12, 2022. <https://statisticsbyjim.com/regression/multicollinearity-in-regression-analysis/>
- Golan R, Gepner Y, Shai I (2019) Wine and health—new evidence. *Eur J Clin Nutr* 72:55–59. <https://doi.org/10.1038/s41430-018-0309-5>
- Gupta Y (2018) Selection of important features and predicting wine quality using machine learning techniques. *Procedia Comput Sci* 1(125):305–312
- Hand, D Mannila, H Smyth, P (2001) Principles of data mining. The MIT Press
- Karlsson B (2022) Wine consumption in the world 2020 in decline, a detailed look. *Forbes*. <https://www.forbes.com/sites/karlsson/2021/12/31/wine-consumption-in-the-world-2020-in-decline-a-detailed-look/?sh=3c0dcc0e3f71>. Published January 16, 2022. Accessed April 9, 2022
- Kumar S, Agrawal K, Mandan N (2020a) Red wine quality prediction using machine learning techniques. (ICCCI) 2020:1–6. <https://doi.org/10.1109/ICCCI48352.2020.9104095>
- Kumar S, Agrawal K, Mandan N (2020b) Red wine quality prediction using machine learning techniques. In: 2020b International Conference on Computer Communication and Informatics (ICCCI) 2020b Jan 22 (pp. 1–6). IEEE
- Lappalainen H, Miskin JW (2000) Ensemble learning. In *Advances in independent component analysis* Springer, London, pp 75–92
- Lee S, Park J, Kang, K (2015) Assessing wine quality using a decision tree. In 2015 IEEE International Symposium on Systems Engineering (ISSE) 176–178 IEEE
- Leung KM. Naive Bayesian classifier. Polytechnic University Department of Computer Science/Finance and Risk Engineering. 2007 Nov;2007:123–56
- Prati, RC Batista, GE and Monard, MC (2004) Learning with class skews and small disjuncts. In *Brazilian Symposium on Artificial Intelligence*. Springer, Berlin, Heidelberg 296–306
- Renee J. Consequences of poor quality in food. *Healthfully*. Published March 7, 2010. Accessed April 12, 2022. <https://healthfully.com/90778-consequences-poor-quality.html>
- Ritchie H, Roser M (2018) Alcohol consumption. *Our world in data*. Published April 16, 2018. Accessed April 10, 2022. <https://ourworldindata.org/alcohol-consumption>
- Smith, J (2020) Is red wine good for you? *Medicalnewstoday.com*. Published April 21, 2020. Accessed April 11, 2022. <https://www.medicalnewstoday.com/articles/265635#can-wine-improve-health>
- Tingwei, Z (2021) Red wine quality prediction through active learning. In *Journal of Physics: Conference Series* 1966(1):012021 IOP Publishing
- Vapnik VN (1999) The nature of statistical learning theory. *Information science and statistics*, 2nd edn. Springer, New York. <https://doi.org/10.1007/978-1-4757-3264-1>
- Wei CC (2012) Receiver operating characteristic for diagnosis of wine quality by Bayesian network classifiers. *AMR* 591–593:1168–1173. <https://doi.org/10.4028/www.scientific.net/amr.591-593.1168>
- World Health Organization (2003) Diet, nutrition, and the prevention of chronic diseases: report of a Joint WHO, FAO Expert Consultation [Geneva, 28 January - 1 2003 Diet, nutrition, and the prevention of chronic diseases: report of a Joint WHO/FAO Expert Consultation [Geneva, 28 January - 1 February 2002]. Accessed May 12, 2022. <https://www.who.int/publications/i/item/924120916X>
- World. Food safety. *Who.int*. Published April 30, 2020. Accessed May 12, 2022. <https://www.who.int/news-room/fact-sheets/detail/food-safety>
- Ye C, Li K, Jia, GZ (2020) A new red wine prediction framework using machine learning. In *Journal of Physics: Conference Series* 1684(1):012067 IOP Publishing

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.