Original Article

Check for updates

Available online at
www.heca-analitika.com/ijds

Infolitika Journal of Data Science

Vol. 1, No. 2, 2023

Enhancing the Red Wine Quality Classification Using Ensemble Voting Classifiers

Deny Joefakri Iwa Supriatna [1], Huzair Saputra [1,*] and Khaidir Hasan [2]

[1] Department of Informatics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; denyjoefakri@gmail.com (D.J.I.S.); huzairsaputra@gmail.com (H.S.)
[2] Aceh Communications, Informatics and Encryption Service, Banda Aceh 23243, Indonesia; khaidirhasan@gmail.com (K.H.)

* Correspondence: huzairsaputra@gmail.com

Article History

Received 5 September 2023
Revised 13 October 2023
Accepted 21 October 2023
Available Online 26 October 2023

Keywords:
Quality assessment
Machine learning
Class imbalance
Hard voting
XGBoost

Abstract

This study introduces an ensemble voting classifier for red wine quality classification using machine learning algorithms. Wine quality assessment, traditionally reliant on subjective expert evaluations, is addressed through data-driven methodologies. The dataset comprises physicochemical attributes and quality ratings of red wines. Results reveal individual models with accuracy ranging from 0.816 to 0.873, while the ensemble approach significantly enhances accuracy. The combination of Random Forest and XGBoost achieves an accuracy of 0.885, demonstrating its potential in red wine quality assessment. In conclusion, this study showcases the potential of machine learning in enhancing the classification of red wine quality, offering a more objective and precise alternative to traditional sensory evaluation. The ensemble voting classifier, especially when combining Random Forest and XGBoost, provides a robust solution for this task, improving the accuracy of wine quality assessments.

1. Introduction

Wine is a complex and multifaceted beverage with a rich history dating back thousands of years [1]. It is crafted from fermented grapes or other fruits, and its quality is influenced by many factors, including the grape variety, terroir (the environment in which the grapes are grown), weather conditions, and the winemaking process [2]. The diversity in wine styles, from light and crisp whites to bold and robust reds, makes wine a subject of fascination for both novices and experts.

Evaluating wine quality, however, is no simple task. Traditionally, it has been assessed through sensory evaluation, involving carefully examining attributes such as appearance, aroma, taste, and mouthfeel [3]. Wine experts, known as sommeliers, rely on their trained palates to subjectively judge the quality of a wine. While their expertise is invaluable, it is susceptible to bias, subjectivity, and variability [4].

Recent developments in machine learning, chemometrics and data analysis methods have created new opportunities for automating quality evaluation procedures [5–9]. It is now possible to examine enormous databases of wine samples, integrating both chemical composition measurements and sensory assessment ratings, and construct models that can precisely classify wine quality.

Various studies have been conducted to assess wine quality using machine learning [10–12]. These studies have explored a range of algorithms, including but not limited to decision trees, random forests, support vector machines, neural networks, and k-nearest neighbors, to develop more accurate and reliable models for evaluating wine quality. However, challenges such as class imbalance and overfitting have been persistent issues in this field, indicating the need for more robust and balanced solutions.

In this study, we proposed an ensemble voting classifier as a potential solution to further enhance wine quality classification. Ensemble methods combine multiple machine learning models to improve prediction accuracy and robustness [13–15]. By integrating the outputs of different algorithms, the ensemble classifier can take advantage of their strengths and mitigate their weaknesses.

The primary aim of this study is to improve the model's performance in classifying the quality of red wine. While the principles and techniques of ensemble voting classifiers apply to wine in general, our study emphasizes enhancing the classification of red wine quality. Red wines have their unique characteristics and complexities, and the study aims to tailor the ensemble classifier to better capture these intricacies and nuances, ultimately providing more accurate and refined quality assessments for red wines specifically.

The study is organized as follows: In Section 2, we provide details regarding the dataset used, data preprocessing, and the ensemble voting classifier approach. Section 3 presents our findings, including performance metrics for different machine learning algorithm combinations and visualizations. Section 4 provides insights from our study, discusses implications, and suggests potential avenues for future research in wine quality classification.

2. Materials and Methods

2.1. Dataset

The red wine dataset used in this research was obtained from the University of California, Irvine (UCI) Machine Learning Repository [16]. This dataset comprises various physicochemical characteristics of red wine originating from Portugal's Vinho Verde region. Notably, due to logistical and privacy considerations, the dataset includes the physicochemical attributes and the output variable, while details such as the wine brand and selling price have been omitted. The dataset consists of 1359 rows and is structured into 12 columns, each representing a distinct variable. Table 1 presents a detailed description of these 12 variables.

2.2. Data Preprocessing

The data preprocessing stage began with removing duplicate entries to ensure data cleanliness, as duplicates can introduce bias and noise into the model. Next, we simplified the wine quality variable by binarizing it into two distinct categories: 'low quality' and 'high quality.' Wines with quality ratings ranging from 0 to 6 were categorized as 'low quality,' while those with ratings from 7 to 10 were classified as 'high quality.' This simplification allowed us to focus on distinguishing between lower and higher quality wines in the analysis.

After binarization, an imbalance between the 'low quality' and 'high quality' classes was observed. The 'low quality' class had significantly more samples (1,382) than the 'high quality' class (217). To address this issue, we performed random undersampling on the 'low quality' class, ensuring a balanced class distribution in the dataset. Finally, the preprocessed dataset was split into training and testing subsets in an 80-20 ratio. This allowed us to train the ensemble voting classifier model on a substantial portion of the data while reserving a separate portion for model validation and testing, ensuring its ability to generalize to unseen data [17].

2.3. Ensemble Voting Classifier

This study explores the construction and application of an ensemble voting classifier for the wine quality classification task. We use hyperparameter tuning through a grid search with five-fold cross-validation to optimize the model's performance. Ensemble voting classifiers merge predictions from multiple individual models to enhance predictive accuracy and model robustness, ensuring that our approach is fine-tuned to achieve the best results.

The ensemble consists of four distinct machine learning algorithms: random forest, XGBoost, support vector machine (SVM), and AdaBoost. Each algorithm brings unique strengths to the classification task. Random forest is known for its robustness and accuracy [18, 19], XGBoost for its efficiency and predictive capabilities [20, 21], SVM for its versatility in handling both linear and non-linear datasets [22, 23], and AdaBoost for its proficiency in mitigating class imbalances by giving more weight to misclassified samples [24]. To create the ensemble, we used the hard voting method, combining predictions from the individual models [25]. This approach leverages the diverse perspectives of these models, potentially improving the overall accuracy and robustness of our wine quality classification system.

2.4. Evaluation Metrics

The trained models were evaluated on the test set to gauge their effectiveness in wine quality classification. To provide a comprehensive assessment, we utilized a range of critical evaluation metrics, each serving a specific

Table 1. Dataset descriptions.

Variable	Description
Fixed acidity	Acids associated with wine are fixed or non-volatile
Volatile acidity	The volume of volatile acids
Citric acid	It is found in minute amounts and gives wines freshness and taste
Residual sugar	The remaining sugar after fermentation has stopped.
Chlorides	The number of salts in the wine.
Free sulfur dioxide	The wine's oxidation and microbiological development are stopped by the free form of SO_2
Total sulfur dioxide	The amount of free and combined forms of SO_2
Density	Depends on the percentage of sugar content and alcohol.
pH	Describes how acidic or basic the wine is on a scale of 0-14. Acids fall under the pH range of 0-7. Wines have a pH between 3 to 4.
Sulfates	An additive to wine that contributes to SO_2 levels.
Alcohol	The alcohol percentage in the wine.
Quality	It is the output variable ranging from 0-10.

purpose in understanding the performance of our ensemble voting classifier.

Firstly, we employed accuracy, which quantifies the proportion of correctly classified instances in the test set, providing a fundamental measure of overall performance. Precision, our second metric, focused on evaluating the precision of positive predictions. It allowed us to examine the number of true positive predictions concerning all positive predictions, helping us gauge the classifier's ability to minimize false positives. Our third metric, recall, delved into the model's ability to effectively capture true positives. It measured the number of true positive predictions in relation to all actual positive instances, offering valuable insights into the classifier's capacity to minimize false negatives. Additionally, the F1-score, a harmonized metric, amalgamated precision and recall into a single value. This balanced measure provided a comprehensive assessment of the classifier's performance. The equations for these metrics are presented in Equations 1, 2, 3, and 4, respectively:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FN} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - score = 2\,x\,\frac{(Recall\ x\ Precision)}{(Recall + \ Precision)} \quad (4)$$

where TP (True Positives) represented instances correctly classified as 'low quality' wines. FP (False Positives) indicated instances wrongly classified as 'low quality' when they were high quality.' TN (True Negatives) were instances correctly classified as 'high quality' wines, and FN (False Negatives) denoted instances incorrectly classified as 'high quality' when they were 'low quality.'

3. Results and Discussion

In this study, we successfully developed a machine learning model for the classification of red wine quality, and this model underwent rigorous hyperparameter tuning to optimize its performance. The model's performance was meticulously evaluated, and the results are presented in Table 2, which showcases various combinations of machine learning algorithms and their corresponding performance metrics. The evaluation metrics for assessing the models include accuracy, precision, recall, and the F1-score. Each metric provides essential insights into the model's classification abilities.

Based on the results, several observations can be made. The individual models, such as Random Forest, XGBoost, SVM, and AdaBoost, each demonstrated respectable performance with accuracy values ranging from approximately 0.816 to 0.873. These models also showed consistent precision, recall, and F1 scores, indicating their reliability in classifying red wine quality.

When we combine two machine learning algorithms, we notice an improvement in the performance metrics, particularly in terms of accuracy. For instance, the combination of Random Forest and XGBoost achieved an accuracy of 0.885, demonstrating an enhancement in classifying red wine quality. Similar improvements in accuracy can be observed in other dual combinations, such as Random Forest with AdaBoost and XGBoost with AdaBoost.

Interestingly, combining three algorithms, Random Forest, XGBoost, and SVM, or Random Forest, XGBoost, and AdaBoost, led to comparable accuracy levels of 0.873 and 0.862, respectively. These results indicate that adding SVM to the ensemble did not significantly improve the model's overall accuracy compared to using just two algorithms. It's possible that the information captured by the Random Forest and XGBoost models largely overlapped, making the addition of SVM redundant. If

Table 2. The performance of model on the testing set.

Combination	Accuracy	Precision	Recall	F1
Random Forest	0.862	0.869	0.862	0.862
XGBoost	0.874	0.875	0.874	0.873
SVM	0.816	0.822	0.816	0.815
AdaBoost	0.839	0.842	0.839	0.839
Random Forest, XGBoost	**0.885**	**0.886**	**0.885**	**0.885**
Random Forest, SVM	0.793	0.803	0.793	0.791
Random Forest, AdaBoost	0.851	0.852	0.851	0.850
XGBoost, SVM	0.805	0.818	0.805	0.802
XGBoost, AdaBoost	0.874	0.874	0.874	0.874
SVM, AdaBoost	0.793	0.809	0.793	0.790
Random Forest, XGBoost, SVM	0.874	0.875	0.874	0.873
Random Forest, XGBoost, AdaBoost	0.862	0.869	0.862	0.862
Random Forest, SVM, AdaBoost	0.851	0.855	0.851	0.850
XGBoost, SVM, AdaBoost	0.862	0.863	0.862	0.862
Random Forest, XGBoost, SVM, AdaBoost	0.862	0.863	0.862	0.862

Bold values indicate best results

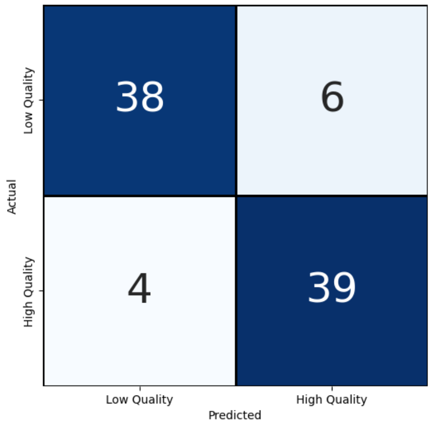

Figure 1. Confusion matrix of the testing set

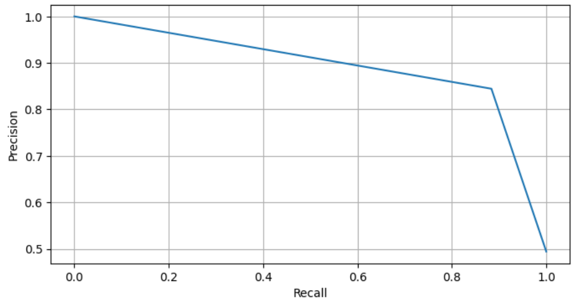

Figure 2. Precision-recall curve.

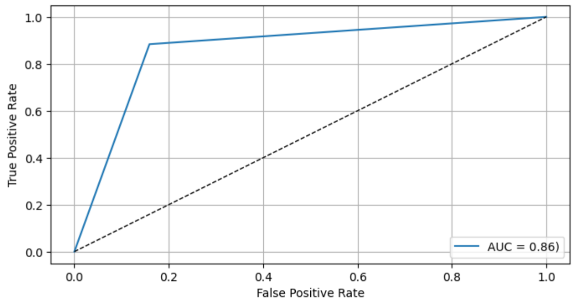

Figure 3. ROC curve.

Random Forest and XGBoost were already effective in capturing the patterns in the data, SVM might not have had much room to contribute additional discriminatory power.

Furthermore, combining all four machine learning algorithms, namely Random Forest, XGBoost, SVM, and AdaBoost, did not yield the highest accuracy. This may be due to the principle of diminishing returns when adding more models to an ensemble. Beyond a certain point, the gains in performance become increasingly minimal, and the complexities associated with managing and fine-tuning the ensemble outweigh the potential benefits. The highest accuracy of 0.885 was achieved by the combination of Random Forest and XGBoost, demonstrating that these two models worked synergistically to produce the best overall classification performance for red wine quality in this study.

In a more detailed analysis of the best-performing model, which combines Random Forest and XGBoost, we visualized the confusion matrix, as shown in Figure 1. The confusion matrix provides a breakdown of the model's classification results, distinguishing between True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN). In this case, the confusion matrix indicates that the model correctly identified 38 instances of 'high-quality' red wines (TP) while incorrectly classifying 6 'low-quality' wines as 'high-quality' (FP). Additionally, the model failed to classify 4 'high-quality' wines correctly (FN) but accurately identified 39 'low-quality' wines (TN). These results offer valuable insights into the model's performance, highlighting the trade-offs between precision and recall.

In the Figure 2, we visualized the Precision-Recall curve, which provides a detailed view of the model's precision and recall trade-off. The precision-recall curve illustrates the model's ability to minimize false positives and capture true positives across different classification thresholds. The precision-recall curve reveals how the model's precision decreases as it becomes more inclusive in classifying 'high-quality' wines, while recall represents the proportion of actual 'high-quality' wines correctly identified by the model.

Lastly, Figure 3 presents the Receiver Operating Characteristic (ROC) curve, which is a graphical representation of the model's performance across different classification thresholds. The Area Under the ROC Curve (AUC) summarizes the model's ability to distinguish between the two classes. In this case, the AUC score is 0.86, which signifies that the model has a good ability to differentiate between 'high-quality' and 'low-quality' red wines. The closer the AUC score is to 1.0, the better the model's discriminatory power. An AUC of 0.86 suggests that the Random Forest and XGBoost ensemble effectively classify red wine quality, with a high likelihood of correctly ranking 'high-quality' wines above 'low-quality' wines across various thresholds.

4. Conclusions

This study successfully developed an ensemble voting classifier for red wine quality classification, achieving an accuracy of 0.885. While the model demonstrates promise in automating wine quality assessment, it simplifies quality into binary categories and faces limitations related to class imbalance and the exclusive focus on red wines. Future research should explore more nuanced quality classifications, address class imbalances more effectively, extend the model's applicability to various wine types, and consider additional features for a more comprehensive assessment. Advancements in machine learning techniques and data collection may offer further opportunities for enhancing wine quality evaluation in the future.

Author Contributions: Conceptualization, D.J.I.S. and H.S.; methodology, D.J.I.S. and H.S.; software, D.J.I.S. and K.H.; validation, D.J.I.S., H.S., and K.H.; formal analysis, D.J.I.S. and H.S.; investigation, D.J.I.S. and H.S.; resources, H.S.; data curation, H.S. and K.H.; writing—original draft preparation, D.J.I.S. and K.H.; writing—review and editing, H.S.; visualization, D.J.I.S.; supervision, H.S.; project administration, H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study does not receive external funding.

Ethical Clearance: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The red wine dataset is available for download from the UCI Machine Learning Repository.

Acknowledgments: The author would like to express their gratitude to their respective institution.

Conflicts of Interest: All the authors declare that there are no conflicts of interest.

References

1. Buja, L. M. (2022). The history, science, and art of wine and the case for health benefits: Perspectives of an oenophilic cardiovascular pathologist, *Cardiovascular Pathology*, Vol. 60, 107446.
2. GAWEL, R. (1998). Red wine astringency: a review, *Australian Journal of Grape and Wine Research*, Vol. 4, No. 2, 74–95. doi:10.1111/j.1755-0238.1998.tb00137.x.
3. Charters, S., and Pettigrew, S. (2006). Product involvement and the evaluation of wine quality, *Qualitative Market Research: An International Journal*, Vol. 9, No. 2, 181–193.
4. Gawel, R., and Godden, P. W. (2008). Evaluation of the consistency of wine quality assessments from expert wine tasters, *Australian Journal of Grape and Wine Research*, Vol. 14, No. 1, 1–8.
5. Wulandari, L., Idroes, R., Noviandy, T. R., and Indrayanto, G. (2022). Application of chemometrics using direct spectroscopic methods as a QC tool in pharmaceutical industry and their validation, 327–379. doi:10.1016/bs.podrm.2021.10.006.
6. Idroes, G. M., Maulana, A., Suhendra, R., Lala, A., Karma, T., Kusumo, F., Hewindati, Y. T., and Noviandy, T. R. (2023). TeutongNet: A Fine-Tuned Deep Learning Model for Improved Forest Fire Detection, *Leuser Journal of Environmental Studies*, Vol. 1, No. 1, 1–8. doi:10.60084/ljes.v1i1.42.
7. Agustia, M., Noviandy, T. R., Maulana, A., Suhendra, R., Muslem, M., Sasmita, N. R., Idroes, G. M., Rahimah, S., Afidh, R. P. F., Subianto, M., Irvanizam, I., and Idroes, R. (2022). Application of Fuzzy Support Vector Regression to Predict the Kovats Retention Indices of Flavors and Fragrances, *2022 International Conference on Electrical Engineering and Informatics (ICELTICs)*, IEEE, 13–18. doi:10.1109/ICELTICs56128.2022.9932124.
8. Maulana, A., Faisal, F. R., Noviandy, T. R., Rizkia, T., Idroes, G. M., Tallei, T. E., El-Shazly, M., and Idroes, R. (2023). Machine Learning Approach for Diabetes Detection Using Fine-Tuned XGBoost Algorithm, *Infolitika Journal of Data Science*, Vol. 1, No. 1, 1–7. doi:10.60084/ijds.v1i1.72.
9. Idroes, R., Noviandy, T. R., Maulana, A., Suhendra, R., Sasmita, N. R., Muslem, M., Idroes, G. M., Kemala, P., and Irvanizam, I. (2021). Application of Genetic Algorithm-Multiple Linear Regression and Artificial Neural Network Determinations for Prediction of Kovats Retention Index, *International Review on Modelling and Simulations (IREMOS)*, Vol. 14, No. 2, 137. doi:10.15866/iremos.v14i2.20460.
10. Kumar, S., Agrawal, K., and Mandan, N. (2020). Red Wine Quality Prediction Using Machine Learning Techniques, *2020 International Conference on Computer Communication and Informatics (ICCCI)*, IEEE, 1–6. doi:10.1109/ICCCI48352.2020.9104095.
11. Aich, S., Al-Absi, A. A., Hui, K. L., Lee, J. T., and Sain, M. (2018). A classification approach with different feature sets to predict the quality of different types of wine using machine learning techniques, *2018 20th International Conference on Advanced Communication Technology (ICACT)*, IEEE, 139–143. doi:10.23919/ICACT.2018.8323674.
12. Mahima, Gupta, U., Patidar, Y., Agarwal, A., and Singh, K. P. (2020). Wine Quality Analysis Using Machine Learning Algorithms, 11–18. doi:10.1007/978-981-15-2329-8_2.
13. Noviandy, T. R., Maulana, A., Idroes, G. M., Emran, T. Bin, Tallei, T. E., Helwani, Z., and Idroes, R. (2023). Ensemble Machine Learning Approach for Quantitative Structure Activity Relationship Based Drug Discovery: A Review, *Infolitika Journal of Data Science*, Vol. 1, No. 1, 32–41. doi:10.60084/ijds.v1i1.91.

14. Sherazi, S. W. A., Bae, J.-W., and Lee, J. Y. (2021). A soft voting ensemble classifier for early prediction and diagnosis of occurrences of major adverse cardiovascular events for STEMI and NSTEMI during 2-year follow-up in patients with acute coronary syndrome, *PLOS ONE*, Vol. 16, No. 6, e0249338. doi:10.1371/journal.pone.0249338.

15. Noviandy, T. R., Maulana, A., Idroes, G. M., Irvanizam, I., Subianto, M., and Idroes, R. (2023). QSAR-Based Stacked Ensemble Classifier for Hepatitis C NS5B Inhibitor Prediction, *2023 2nd International Conference on Computer System, Information Technology, and Electrical Engineering (COSITE)*, IEEE, 220–225. doi:10.1109/COSITE60233.2023.10250039.

16. Cortez, P., Cerdeira, A., Almeida, F., Matos, T., and Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties, *Decision Support Systems*, Vol. 47, No. 4, 547–553. doi:10.1016/j.dss.2009.05.016.

17. Noviandy, T. R., Maulana, A., Idroes, G. M., Maulydia, N. B., Patwekar, M., Suhendra, R., and Idroes, R. (2023). Integrating Genetic Algorithm and LightGBM for QSAR Modeling of Acetylcholinesterase Inhibitors in Alzheimer's Disease Drug Discovery, *Malacca Pharmaceutics*, Vol. 1, No. 2, 48–54. doi:10.60084/mp.v1i2.60.

18. Le, T.-T.-H., Kim, H., Kang, H., and Kim, H. (2022). Classification and Explanation for Intrusion Detection System Based on Ensemble Trees and SHAP Method, *Sensors*, Vol. 22, No. 3, 1154. doi:10.3390/s22031154.

19. Zermane, A., Tohir, M. Z. M., Zermane, H., Baharudin, M. R., and Yusoff, H. M. (2023). Predicting fatal fall from heights accidents using random forest classification machine learning model, *Safety Science*, Vol. 159, 106023.

20. Maulana, A., Noviandy, T. R., Sasmita, N. R., Paristiowati, M., Suhendra, R., Yandri, E., Satrio, J., and Idroes, R. (2023). Optimizing University Admissions: A Machine Learning Perspective, *Journal of Educational Management and Learning*, Vol. 1, No. 1, 1–7. doi:10.60084/jeml.v1i1.46.

21. Hidayat, T., Hadinata, E., Damanik, I. S., Vikki, Z., and Irvanizam, I. (2023). Implementation of Hybrid CNN-XGBoost Method for Leukemia Detection Problem, *Infolitika Journal of Data Science*, Vol. 1, No. 1, 15–21. doi:10.60084/ijds.v1i1.87.

22. Yang, S., Li, C., Mei, Y., Liu, W., Liu, R., Chen, W., Han, D., and Xu, K. (2021). Determination of the Geographical Origin of Coffee Beans Using Terahertz Spectroscopy Combined With Machine Learning Methods, *Frontiers in Nutrition*, Vol. 8, No. June, 1–10. doi:10.3389/fnut.2021.680627.

23. Lahmiri, S., Dawson, D. A., and Shmuel, A. (2018). Performance of machine learning methods in diagnosing Parkinson's disease based on dysphonia measures, *Biomedical Engineering Letters*, Vol. 8, No. 1, 29–39. doi:10.1007/s13534-017-0051-2.

24. Noviandy, T. R., Maulana, A., Emran, T. B., Idroes, G. M., and Idroes, R. (2023). QSAR Classification of Beta-Secretase 1 Inhibitor Activity in Alzheimer's Disease Using Ensemble Machine Learning Algorithms, *Heca Journal of Applied Sciences*, Vol. 1, No. 1, 1–7. doi:10.60084/hjas.v1i1.12.

25. Habib, A.-Z. S. Bin, and Tasnim, T. (2020). An ensemble hard voting model for cardiovascular disease prediction, *2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI)*, IEEE, 1–6.