ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25nov409

Developing a Dashboard Embedded with KNN Machine Learning Algorithm for Wine Quality Prediction

Theophilus Bamise Ajala[1]; Rashid Kehinde Oloko[2]; Abiodun Richard Agboola[3]

[1,2,3] College of Computing, Caleb University, Imota, Lagos, Nigeria

Publication Date: 2025/11/21

Abstract: In the beverage sector, wine quality is crucial because of its high demand and competitive market. Classifying wine quality is a challenging task because the evaluation provided by human specialists is costly and time-consuming. The aim of this research is to develop a dashboard embedded with KNN machine learning algorithm for wine quality prediction. The user opens the GUI application, supplies the values of the wine features, the entered information serves as the dataset for the red wine quality prediction system, which utilizes it to accurately forecast outcomes based on the specified range. The output of the KNN algorithm has been estimated using various evaluation metrics. with respect to precision, recall, f1-score and accuracy are 21.277%, 52.632%, 30.303%, and 85.625% respectively. Kaggle red wine dataset serves as a benchmark for the research. The essence of this research is that the adoption of a machine learning algorithm in predicting wine quality can enhance both the efficiency and precision of wine quality evaluations before production.

Keywords: *Quality of Wine, Red Wine, KNN, Classification, Machine Learning Algorithm.*

How to Cite: Theophilus Bamise Ajala; Rashid Kehinde Oloko; Abiodun Richard Agboola (2025) Developing a Dashboard Embedded with KNN Machine Learning Algorithm for Wine Quality Prediction. *International Journal of Innovative Science and Research Technology*, 10(11), 1096-1103. https://doi.org/10.38124/ijisrt/25nov409

I. INTRODUCTION

In the beverage sector, wine quality is crucial because of its high demand and competitive market. Making sure that the quality of wine is well produced, served and enjoyed by producers and customers respectively is a thing of utmost important that must not be taken with levity hand, this leads to the acquistion of quality measures used in controlling the cycle of production Baheti et al. (2024). As more and more people begin to consume wine, the red wine sector is expanding at an incredible rate. As a result, wine companies must produce higher-quality wines to differentiate themselves in the more competitive market (Dong, 2023). Every year, Americans spend about $70 billion on wine (Dong, 2023).

According to Bhardwaj et al. (2022) the wine business relies heavily on wine quality certification. Wine production has expanded in the contemporary era as a result of excessive consumption. Due to social and normative considerations, wine consumption is on the rise in most regions of the world (Gawale, 2022). The main cause of the intense market competitiveness has been the rise in wine output. It is quite difficult for the majority of wine production enterprises to defend the quality of their wines. Companies must concentrate on wine certification in order to evaluate wine quality in the market and preserve wine quality (Cardoso et al., 2022). According to Pradnya et al. (2023) drinking low-quality wine has a negative effect on one's health, wine quality is particularly important. Wine quality is one of the main issues the wine business is dealing with. The person who defines quality determines it, regardless of their level of expertise. The wine's flavor, aroma, color, and other qualities are determined by its chemical composition. The type of grape, the environment, the microbial strains that are present during fermentation, and viticultural techniques all have an impact on the chemical composition (Bhardwaj et al., 2022). With the advent of technology, wine quality can now be predicted and the accuracy ratio increased using machine learning, deep learning, and hybrid methods (Baheti et al., 2024).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25nov409

II. STATEMENT OF PROBLEM

Wine quality is determined by the ingredients that go into its production. The drink called wine is a beverage that people around the world are familiar with; giving the notion that the older the wine, the better it tastes, but it costs more. Important metrics such free sulfur dioxide, volatile acidity, citric acid, and residual sugar are used to assess the quality of wine (Saranya et al., 2024). Wine quality depends on several stages, in recent days, businesses use product quality certifications to advertise their goods (Gupta, 2017). This is back-breaking and an expensive process that takes too much time to be evaluated by human experts. The traditional way to assess wine quality is time consuming. Also, from the literature reviewed, the researchers that made use of machine learning, implemented their code which provides less GUI user interaction. For buyers and the wine business to produce in sufficient amounts, wine quality is crucial. Wine quality classification is a difficult task because taste is the least important of the five senses Pradnya et al. (2023). To nearly accurately forecasting wine quality presents a significant challenge within the wine industry, hence, this is a research gap whereby the application of machine learning model will be investigated in this research to predict wine quality alongside graphical user interface library to enhance user interaction.

III. AIM AND OBJECTIVES OF THE RESEARCH

The aim of this project is to develop a dashboard embedded with KNN machine learning algorithm for wine quality prediction. The objectives are:

- Collection of dataset that consist of red wine features
- Design a web application for wine quality prediction
- Implement machine learning algorithm in (ii) using Python
- Evaluate the model performance in (iii) using wine classification accuracy, precision, f1_score, and recall.

IV. LITERATURE REVIEW

Many researchers published related works on wine prediction. From this background, this section provides an overview on some of the main studies conducted on different data to predict wine quality using machine learning:

In Bhardwaj et al. (2022), the authors create synthetic data and build a machine learning model using this wine data as well as available experimental data gathered from various and varied parts of New Zealand. They used 18 samples of wine consisting of Pinot noir details with 54 attributes categorized as 47 chemical and 7 physiochemical. Using the Synthetic Minority Oversampling Technique (SMOTE), they created samples consisting of 1381 from 12 original wine samples; 6 samples were reserved and used to test the model. Four different feature selection techniques were used to compare the results. Seven different machine learning algorithms were trained and tested on the wine samples. When trained and assessed without feature selection, the Adaptive Boosting (AdaBoost) classifier demonstrated 100% accuracy.

In Gupta (2018), the author explores the usage of macine learning techniques for product quality by using linear regression to ascertain how dependent the target variable is on independent variables. Important variables are chosen based on computed dependency. Additionally, the values of the dependent variable are predicted using support vector machines and neural networks. As for the trials, the author combine white and red wine datasets. This study demonstrates that when certain features (variables) are taken into account instead of all the features, a better forecast can be made.

In Saranya et al. (2024), the authors' project gives an automatic prediction of Wine quality, as good or bad, using machine learning approaches which are Neural Networks, Logistic Regression and Support Vector Machine are implemented on datasets of Portuguese “Vinho Verde” Wine. The results are compared with standard values. The assure the customers of the notable job being carried out in the wine industry, this work shows how quality testing is carried out.

Baheti et al. (2024), the authors' study examine the model used in predicting the quality of alcohol by employing regression analysis, a classification type of machine learning. The study utilzed various steps starting with preprocessing to modeling, then to evaluation to ascertain the accuracy of the predictions. Altogether, the study shows how effective linear regression is in predicting the content of alcohol, recording an accuracy of 92%, this proffer a positive implication for quality production and control in the alchol industry in alignment with other regression models most notably gradient boosting attaining an accuracy of 90% and decision tree regressor showing 83% as its accuracy.

In Dong (2023), the author utilized machine learning models to identify the factors that have the greatest influence on the overall quality of wine by analyzing 1599 wine samples, each with 11 input variables. The study's linear regression model reveals that acid and alcohol had the most effects on quality. A heat map was also used to display all of the correlations between the variables. In order to gain a more thorough understanding of the variables that have the greatest influence on quality, box plots and 3D scatter plots were employed to corroborate the results obtained from the linear regression model. These findings provide insight into the factors that have the biggest impact on wine quality.

V. PROPOSED METHODOLOGY

According to Irfan et al. (2019), another type of supervised machine learning process is classification. As for Classification, when a dataset is annotated, then a column is created that is used for prediction known as the predict class label, classification is a type of data analysis used to extract and derive a model for

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25nov409

prediction. This project utilized the K-Nearest Neighbour (KNN) algorithm, a type of supervised machine learning model. The research was structured around an experimental framework and web development that involved gathering a dataset, specifically by leveraging publicly accessible datasets related to wine. The process commenced with the importation of the wine dataset, followed by data pre-processing and feature extraction. KNN algorithm was then employed to establish the classification task after dividing the dataset into training and testing subsets. Finally, the model was assessed to evaluate the impact of the optimization on its achievement.

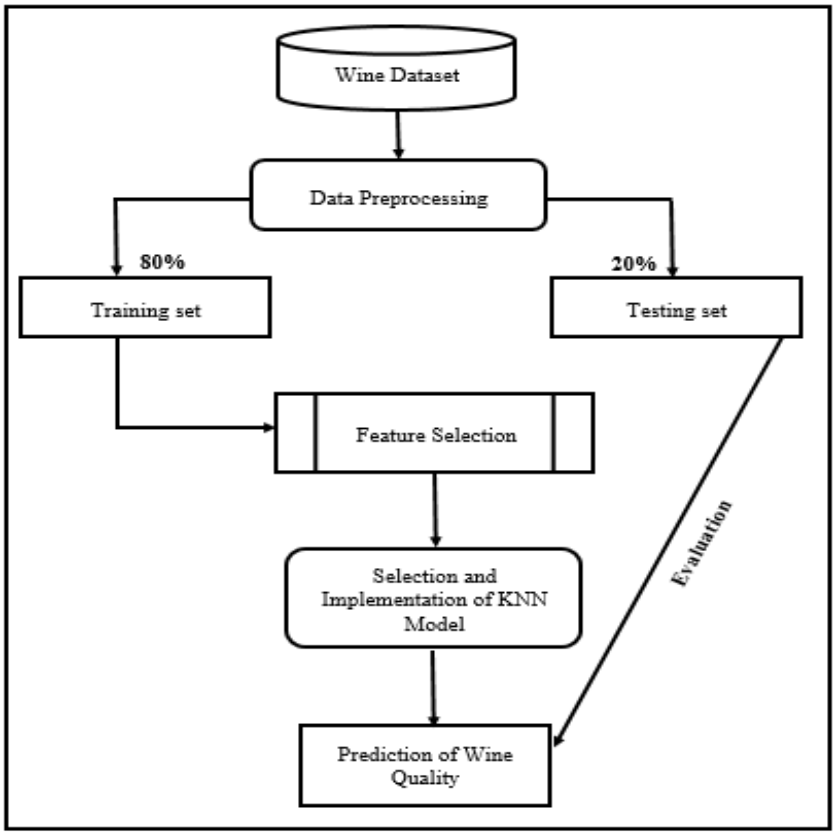
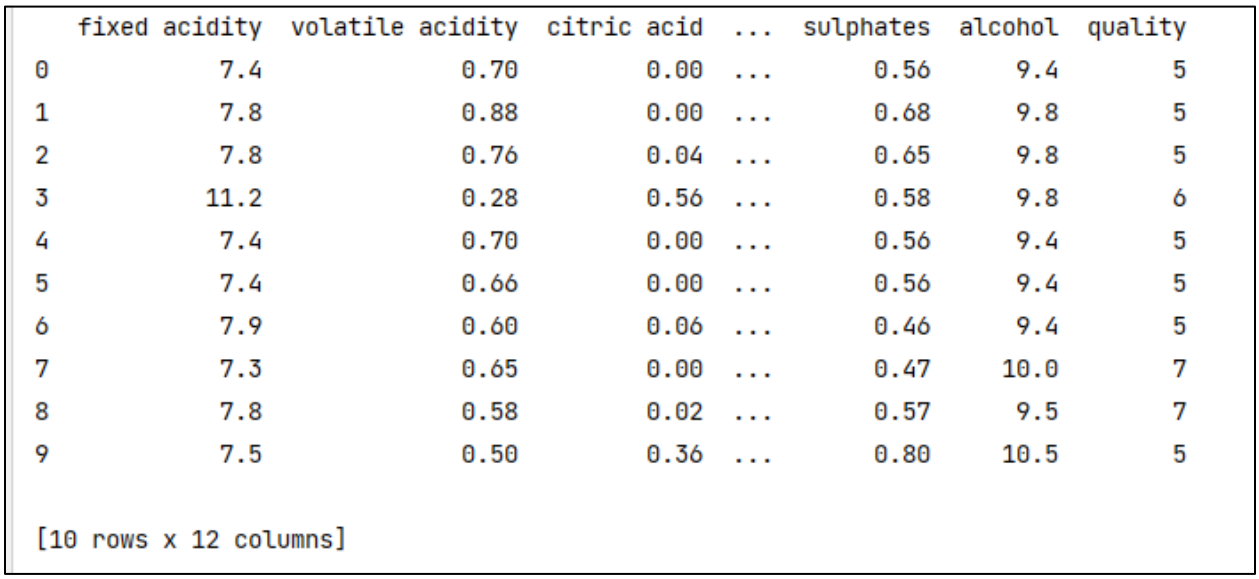


Fig 1: Workflow Diagram of the Methodology

We downloaded a dataset of 1500+ wine samples from the Kaggle Machine Learning Repository,

which has become very famous as an open-source dataset. For the dataset utilized in this research, it contains 1599 rows and 12 columns, out of the columns we have 11 different wine features which will be used for the training and testing, the 12th column handles the prediction class. This analysis focuses exclusively on red wine. The output variable represents quality, which is scored between 0 and 10.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25nov409



	fixed acidity	volatile acidity	citric acid	...	sulphates	alcohol	quality
0	7.4	0.70	0.00	...	0.56	9.4	5
1	7.8	0.88	0.00	...	0.68	9.8	5
2	7.8	0.76	0.04	...	0.65	9.8	5
3	11.2	0.28	0.56	...	0.58	9.8	6
4	7.4	0.70	0.00	...	0.56	9.4	5
5	7.4	0.66	0.00	...	0.56	9.4	5
6	7.9	0.60	0.06	...	0.46	9.4	5
7	7.3	0.65	0.00	...	0.47	10.0	7
8	7.8	0.58	0.02	...	0.57	9.5	7
9	7.5	0.50	0.36	...	0.80	10.5	5

[10 rows x 12 columns]

Fig 2: Screenshot Sample of Red Wine Dataset

Table 1: Features of the Dataset (Author, 2025)

S/N	Features (Column Name)	Datatype
1	Fixed acidity	float
2	Volatile acidity	Float
3	Citric acid	Float
4	Residual sugar	Float
5	Chlorides	Float
6	Free sulfur dioxide	Float
7	Total sulfur dioxide	Float
8	Density	Float
9	pH	Float
10	Sulphates	Float
11	Alcohol	Float
12	Prediction quality	Integer

VI. ALGORITHM

From a training set, What KNN does as a model is to classify a label class of a new instance based on the class that is more from k-neighbour. KNN checks testing set side by side with the training set which are of proximity to and have resemblance to the test data. The classified-based KNN algorithm is used to classify objects depending on the learning data that are of close proximity to the object. Learning data is estimated to be a multi-dimensional space, where each magnitude describes the features of the set of data. This space partitioned into segments is based on the categorization of learning data. A point in this space is indicated with "c", which means class, if c is the classification that is frequently seen in the k-neighbour of the close proximity to that point. KNN is a supervised machine learning model, where the results of searching newly instantiated objects are classified based on big the categories are on the KNN. Regarding the training phase, at most KNN keeps feature vectors and classification of sample data relating to the training set. For the phase that handles the classification, matching features are computed for testing set of data; which means the classification is not known ahead and it is not part of the training phase. The new vector will have its distance computed with all the sample training vectors and the close proximity k will be selected. The point with the majority classification will be included with the latest point of classification that is already predicted. (Irfan et al., 2019). In order to calculate the K-neighbour's similarity of data, the equation is as follows.

$$Similarity(T,S) = \frac{\sum_{i=1}^{n} Sim(K_i(T).K_i(S))xwi}{\sum_{i=1}^{n} w} * 100\%$$

Where:
T = testing data
S = training data
n = total of criteria
w = weight of criteria
Sim ($K_i(T)$, $K_i(S)$) = Similarity value standing as the distance between the source and case target.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25nov409

➢ *Working of KNN Algorithm*

As for the "k" in K-NN, it represents the count of neighbors consisted in the new data point. It is very paramount to decide an acceptable value for k in the KNN algorithm. In order to enhance the accuracy, it is germain that we select the accurate value of k, and this process is formally called parameter tuning. A noisy results can be obtained when the value of k is like 1 or 2, on the contrary, an extreme value can create confusion some times, with respect to the dataset used. There is no default value for k, howbeit, in practice the value that k uses is "5", that is, for the most voting process, the 5 neighbors of close proximity to the new data point are put into consideration. To prevent confusion and mistakes in the midst of two classes of data sets, overall, an odd value of k is the best or most suitable to use (Bansal et al. 2022). Another formula-based calculation for K can be done through this formula:

$$K = \sqrt{n}$$

And, **n** is the overall count of data points.

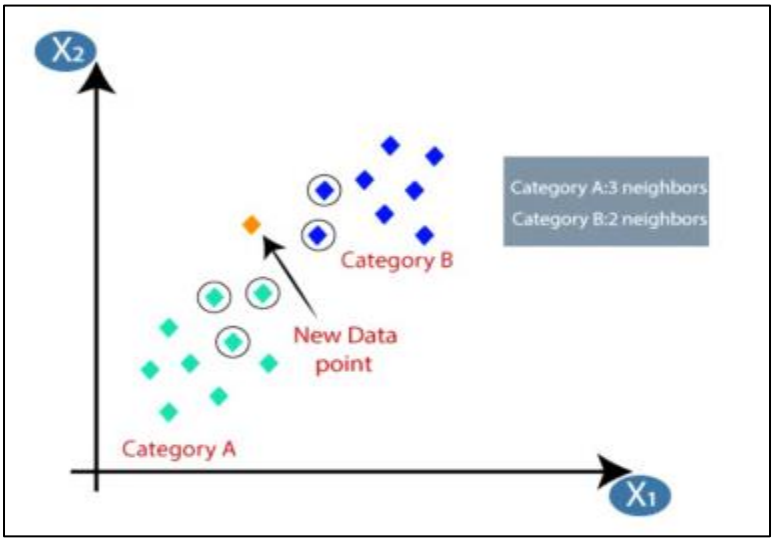


Fig 3: Schematic of KNN for Classification of New Data Point Based on Neighbors (a) (Bansal et al. 2022).

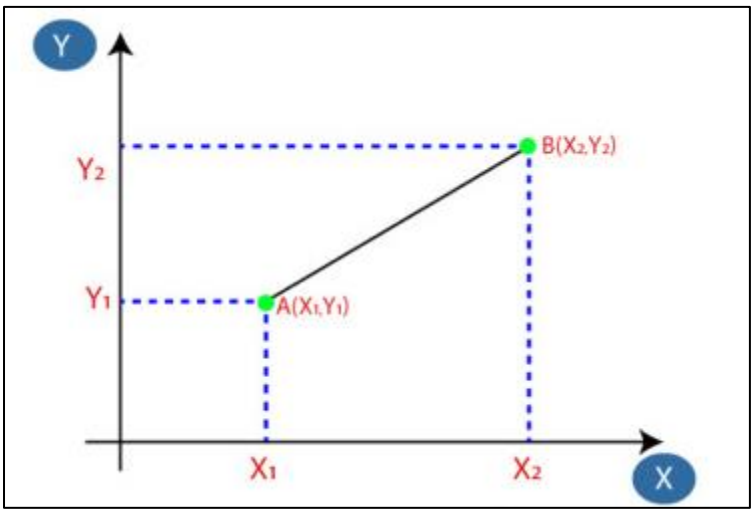


Fig 4: Schematic of KNN to Find Euclidean Distance Between Two Points (Bansal et al. 2022).

➢ *Implementation of KNN*

KNN 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 KNN classifier
Step7: Check efficiency for evaluation of model

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25nov409

VII. RESULT AND DISCUSSION

The implementation of the system involves the development of a new application based on the strategies outlined in the preceding chapter and the designated implementation environment. This chapter illustrates the execution of the plan, with the objective of delivering a system that utilizes machine learning to predict wine quality.

➢ *Model Evaluation Metrics*

Consequently, on the selected dataset, we assess and validate the performance of the selected algorithm KNN by using four evaluation metrics: accuracy, F1-score, precision and recall.

➢ *Accuracy*

Accuracy is a true prediction ratio (positive and negative) by testing the entire data into the model to obtain the target value. Accuracy is used to find the amount of data classified correctly Anggoro & Kurnia (2020). Algorithm performance in terms of accuracy can be evaluated using the following formula as given below:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\%$$

Where:
TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative

➢ *F1-score*

The F-score can be described as the weighted average involving both precision and recall depending on the weight function β, view Formula below:. The F1-score refers to the harmonic mean in the middle of recall and precision, when it is written F-score it usually means F1-score. The F-score is also called the F-measure. The F1-score can have different indices giving different weights to precision and recall (Dalianis, 2018).

$$F - score : F_\beta = (1 + \beta^2) * \frac{P * R}{\beta^2 * P * R}$$

With β= 1 the standard F-score is obtained

$$F1 - Score = \frac{2T_p}{2T_p + F_p + F_N} \; x\, 100$$

➢ *Recall*

According to Dalianis (2018), recall is calculated by dividing the number of correctly retrieved instances by the total number of correctly retrieved instances; the formula is as follows:

$$Recall = \frac{TP}{TP + FN} \; x\, 100$$

➢ *Precision*

According to Dalianis (2018), precision is calculated by dividing the number of correctly retrieved instances by the total number of retrieved instances; the formula is as follows:

$$Precision = \frac{TP}{TP + FP} \; x\, 100$$

Table 2: Model Evaluation

	K-Nearest Neighbor
Precision	21.277%
Recall	52.632%
F1 Score	30.303%
Accuracy	85.625%

Regarding the accuracy metric, it indicates that 85.625% of the total predictions, encompassing both correct positives and negatives, made by the model were accurate. In terms of the precision metric, it reveals the proportion of wines classified as 'good quality' that were indeed of good quality. The recall metric assesses the model's capability to identify all actual good quality wines, with a recall rate of 52.632% signifying that it accurately recognizes approximately half of the good quality wines. Lastly, the F1-score metric, with a score of 30.3%, demonstrates a significant imbalance between precision and recall, primarily influenced by the low precision.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25nov409

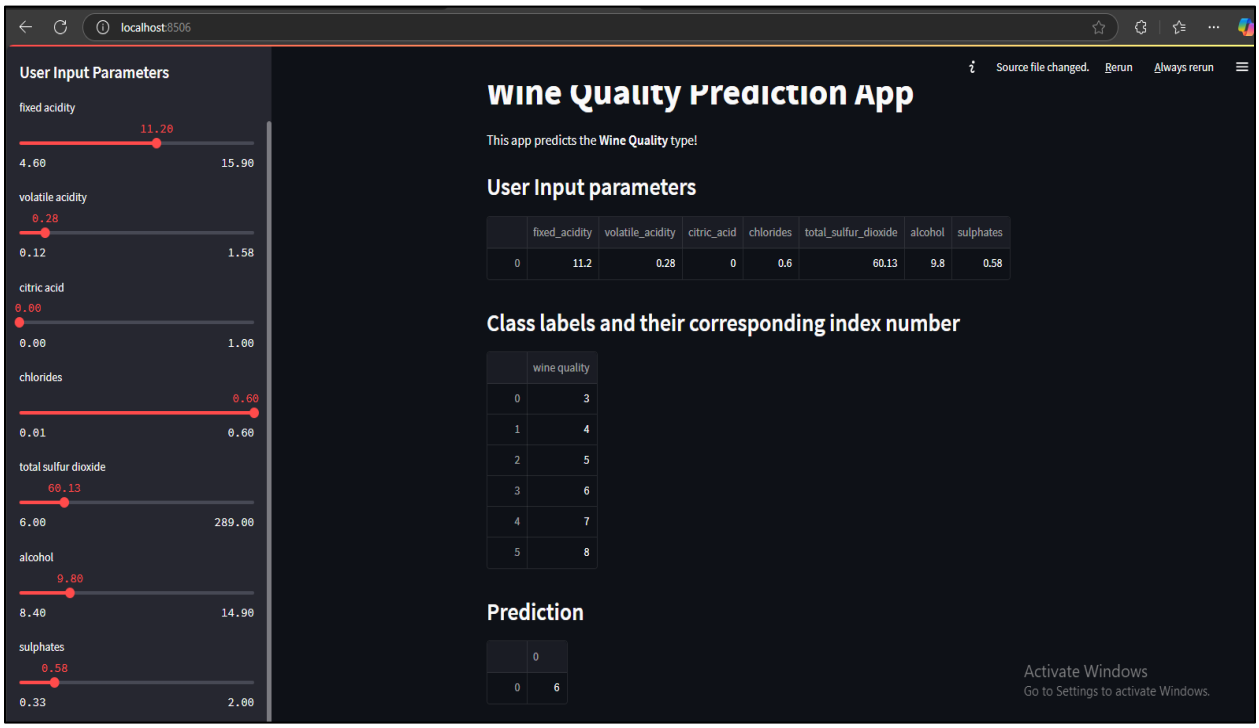


Fig 5: Screenshot of Wine_Quality_Prediction Page

The above screenshot illustrates the web page where all essential data is entered for evaluation, leading to the generation of corresponding results. The entered information through the left sliders serves as the dataset for the red wine quality prediction system, which utilizes it to accurately forecast outcomes based on the specified range.

VIII. CONCLUSION

The main aim in this project is to develop a dashboard embedded with KNN machine learning algorithm for wine quality prediction. This project utilized K-Nearest Neighbour (KNN) algorithm, a type of supervised machine learning model. The research was structured around an experimental framework that involved gathering a dataset, specifically by leveraging Kaggle publicly accessible datasets related to wine.

The result obtained in the implementation shows that when all essential data is entered by the user for evaluation, leading to the generation of corresponding results. The entered information serves as the dataset for the red wine quality prediction system, which utilizes it to accurately forecast outcomes based on the specified range. The application then analyzes these inputs to predict the quality of red wine, categorizing values from 0 to 3 as low quality, 4 to 6 as average quality, and 7 to 10 as high quality, as indicated in the prediction table.

REFERENCES

[1]. Abernathy C. (2020). “Press Release: Frequent Wine Drinking Population in the US in Decline, Led by Younger Consumers, Though Overall Participation in Wine Category Up.” Wine Intelligence, Courtney Abernathy Retrieved from: https://Www.wineintelligence.com/Wp-Content/Uploads/2018/07/logo5.Png, 2020.

[2]. Anggoro, D.A., & Kurnia, N.D. (2020). Comparison of Accuracy Level of Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) Algorithms in Predicting Heart Disease. Volume 8. No. 5, May 2020. International Journal of Emerging Trends in Engineering Research Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter32852020.pdf. doi: https://doi.org/10.30534/ijeter/2020/32852020.

[3]. Arshad, H., Naveed, H., Nasim, F., Ahmad, J., Ahmed, M., & Liaqat, M.S. (2024). Wine Quality Prediction Using Machine Learning.

[4]. Aslam, M. (2022). Wine Quality Predictin by using Machine Learning Algorithms. GSJ: Volume 10, Issue 12, December 2022, Online: ISSN 2320-9186 www.globalscientificjournal.com.

[5]. Baheti, A.S., Sawarkar, A.D., Shaikh, U.A., & Shrimankar, D.D. (2024). Alcohol Quality Analysis Using Machine Learning Regression Technique. DOI: https://doi.org/10.7759/s44389-02400227-1.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25nov409

[6]. Bansal, M., Goyal, A., & Choudhary, A. (2022). A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning. doi: https://doi.org/10.1016/j.dajour.2022.100071.
[7]. Basalekou, M., Tataridis, P., Georgakis, K., & Tsintonis, C. (2023). Measuring Wine Quality and Typicity. doi: https://doi.org/10.3390/beverages9020041.
[8]. Bhardwaj, P., Tiwari, P., Olejar, K., Parr, W., & Kulasiri, D. (2022). A machine learning application in wine quality prediction. doi: https://doi.org/10.1016/j.mlwa.2022.100261.
[9]. Butnariu, M., & Butu, A. (2020). Biotechnological Progress and Beverage Consumption.
[10]. Cardoso, C., Schwindt, V., Coletto, M. M., D´ ıaz, M. F. & Ponzoni, I. (2022). Could qsor modelling and machine learning techniques be useful to predict wine aroma?, Food and Bioprocess Technology pp. 1–19.
[11]. Dalianis, H. (2018). Evaluation Metrics and Evaluation. DOI: 10.1007/978-3-319-78503-5_6.
[12]. Dong, J. (2023). Red Wine Quality Analysis based on Machine Learning Techniques.
[13]. Gupta, Y. (2018). Selection of important features and predicting wine quality using machine learning techniques.
[14]. Irfan, M., Nurhidayat, A.R., Wahana, A., Maylawati, D.S., & Ramdhani, M.A. (2019). Comparison of K-Nearest Neighbour and support vector machine for choosing senior high school. doi: 10.1088/1742-6596/1280/2/022026.
[15]. Gawale, A.S. (2022). Wine Quality Prediction using Machine Learning and Hybrid Modeling.
[16]. Gupta, Y. (2017). Selection of important features and predicting wine quality using machine learning techniques.
[17]. Pradnya, W., Swapnil, N., Sudarshan, S., & Pravin, S. (2023). A study and analysis of wine prediction model using various machine learning techniques. DOI: https://www.doi.org/10.56726/IRJMETS38486.
[18]. Saranya, G., Chennamsetty, S., and Rachupalli, P.K. (2024). Wine quality prediction using machine learning