Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th October 2025. Vol. 5. No. 1; e-ISSN: 2808-4519

Designing A Wine Quality Identification Application Using Naïve Bayes Classifier

Davin Auco[1]*, Edi[2], Hendri[3]

[1,2,3]Information Engineering Systems Study Program, STMIK Time Medan, Indonesia
Davinwijaya32@gmail.com[1], Edi@gmail.com[2], h3ndr1wu@gmail.com[3]*

Abstract

Wine is one of the drinks with the highest number of consumers in the world, and its quality is greatly influenced by various chemical ingredients such as *fixed acidity*, *citric acid*, pH, and others. Wine quality testing is traditionally subjective and requires special expertise. Therefore, a more objective and efficient approach is needed, one of which is through the application of data *mining* algorithms. This study aims to design a wine quality identification application using the *Naïve Bayes Classifier algorithm* and analyze its accuracy in the classification of red and white wines. The dataset used was taken from *the UCI Machine Learning Repository*, with input attributes in the form of wine chemical content and *output* in the form of wine quality scores ranging from 2 to 8. The test results showed that the *Naïve Bayes algorithm* was able to classify the quality of *wine* with an accuracy rate of 86.50% for red *wines* and 79.80% for white *wines*. The developed application successfully processes input data and provides prediction results in *real-time*. This research shows that the use of *Naïve Bayes algorithms* can be a practical and effective solution in helping consumers and producers in recognizing the quality *of wine* and improving the production process in the *wine industry*.

Keywords: *Wine, Classification, Naive Bayes Classifer, Data Mining, Apllication.*

1. Introduction

Wine is one of the most widely consumed beverages in the world. Wine quality is influenced by various factors, including grape variety, environmental conditions where grapes are grown, winemaking techniques, and storage processes [1]. There are several components in wine that affect its quality, including citric acid content, pH level, fixed acid content, and others. Therefore, it is important to be able to identify and classify wine quality with accuracy, consistency, and objectivity to ensure consumer satisfaction and producer reputation [2].Wine quality assessment plays a crucial role in the wine industry. One factor that has contributed to changes in consumer behavior regarding wine consumption is the perceived image of wine [3]. Wine quality assessment also helps consumers make more informed purchasing decisions and ensures they receive products that align with their preferences and expectations. As stated by I Made Puvtra Mahardika in 2022, wine product quality has a positive impact on consumer interest [4]. Traditional approaches to wine quality assessment often involve sensory evaluation by wine experts or sommeliers. In general, wine experts are tasked with evaluating various types of wine [5].

The methods used include the use of human senses, such as sight, smell, and taste, to evaluate the physical and organoleptic characteristics of wine, such as color, aroma, taste, and texture. However, this assessment is often subjective and requires the expertise of the assessor to provide an accurate assessment.Wine quality classification can be performed using data mining. Data mining applications utilize several types of algorithms such as Naive Bayes, Random Forest, Support Vector Machines (SVM), Decision Trees, K-NN, and others [6].The Naive Bayes Classifier algorithm is a probabilistic algorithm that calculates probabilistic groups by calculating combinations and data frequencies. Using Bayes' theorem, this algorithm assumes that the attributes are independent even though there are class variable values [7].

The Naive Bayes Classifier algorithm is an algorithm that uses only a small amount of training data to select the parameters needed during the classification process. It uses several interrelated attributes to determine suitability [8]. The advantages of the Naive Bayes Classifier algorithm in wine quality testing include ease of implementation, high interpretability, and good performance in cases where the assumption of feature independence is almost fulfilled.

2. Theoretical Foundation

2.1. Architectural Visualization

2.2.1. Wine

Wine is one of the most widely consumed beverages in the world. Countries such as the United States, France, Italy, Germany, and China are

recorded as major consumers, with an average consumption of 120 million hectoliters per year [9].
Almost all wines can be classified into one of three types: dry wine, fortified wine, or carbonated wine [10]. Almost all wines can be classified into one of three types: dry wine, fortified wine, or carbonated wine [10]. The sweet and bitter components in wine, which are characteristic in wine quality selection, can be defined based on their content, namely total acidity, volatile acidity, citric acid, residual sugar, chloride, free sulfur dioxide, total sulfur dioxide, density, pH, sulfates, and alcohol [11].

2.2.2. Data *Mining*

Data mining is the process of extracting useful information and patterns from very large data sets. The data mining process consists of data collection, data extraction, data analysis, and data statistics. These four processes in data mining will produce very useful models/insights. Data mining is included in Knowledge Discovery in Databases (KDD) [12].

2.2.3. Knowledge Discovery

Data mining is the process of extracting useful information and patterns from very large data sets. The data mining process consists of data collection, data extraction, data analysis, and data statistics. These four processes in data mining will produce highly useful models/insights. Data mining is categorized under Knowledge Discovery in Databases (KDD). In this context, Figure 2.1 illustrates part of the data mining process [12].

2.2.4. Naïve Bayes Classifier

The Naive Bayes Classifier (NBC) method is a method used to determine probability values or probabilities in predicting opportunities based on data from previous experiences or enabling grouping in a system [13]. The advantage of using the NBC algorithm is that it only requires a small amount of training data to determine the parameter estimates needed in the classification process [14]. In general, the NBC formula is essentially a vector [13]:

$$P(H \mid X) = \frac{P(H \mid X)\ P(H)}{P(X)} \qquad (1)$$

Where:
X = Unknown data class

H = Specific hypothetical class of X
P(H | X) = Probability of hypothesis H based on condition X P(H) = Probability of hypothesis H (prior)
P(X | H) = Probability of X based on condition P (X) = Probability of X
In this process, the NBC method needs to take the initiative to determine the class that corresponds to the tested sample data, then from that the general NBC formula is changed to the following [13]:

$$P(H \mid X1 \ldots Xn) = \frac{P(H)\ P\ (X1\ldots Xn|H)}{P(X1\ldots Xn)}$$

Where variable H represents the class and X1... Xn represents the unknown class and will be needed for the classification process [13].

2.2.5. Confusion Matrix

The Convergence Matrix is one of the methods used in the evaluation process of data mining classification models to predict the accuracy of objects. The Convergence Matrix is described as a table that indicates the number of correctly classified data and the number of incorrectly classified data. Based on the Convergence Matrix Table published [15]:

a. True Positive (TP) is the number of positive data classified as positive values.
b. False Positive (FP) is the number of negative data classified as positive values.
c. False Negative (FN) is the set of positive data classified as negative values.
d. True Negative (TN) is the number of negative data classified as negative values.

The values of the confusion matrix are used for evaluation as follows [15]:

a. Accuracy is the ratio of the number of correctly classified data (in predictions) compared to the number of data incorrectly classified by the algorithm.
Formula: (TP + TN) / Total data = Accuracy (3)
b. Precision, the ratio of positive predictions to the total number of positive predictions.
Formula: (TP) / (TP+FP) = Precision (4)
c. Recall, the ratio of positive predictions to the total number of actual positive data points.
Formula: (TP) / (TP+FN) = Recall (5)
d. Classification Error Rate, which is the percentage of data incorrectly classified by the algorithm.
Formula: (FP + FN) / Total data = Classification Error Rate (6)

3. Research Methods

3.1. DataCollectionMethods

Data collection methods that are used on this research include:

a. Metode Obseis buzzing
Perform observations by collecting data Seconds dari *wevbsitev* UvCI Machinev Levarning Revpository yaitu *Winev Quvality* on *the link* https://archivev.ics.uvci.evduv/datasevt/186/ *winev+quvality*.

b. Method Studi Book

Gather data-data teori Through the journal, Print media , or the sources refevrensi dari intevrnĕvt.

3.2. System Analysis

The system analysis in this study was conducted in three stages, namely:

a. Analysis of the processes taking place in the field was conducted to analyze the system currently used in wine quality classification.
b. Analysis of the algorithm used, namely the Naïve Bayes Classifier algorithm with examples of the implementation of this algorithm in wine quality classification.
c. Analysis of the proposed system, namely a description of the system to be built with modified features using tools from AS.

The diagram to be built is shown in Figure 1 from AS.

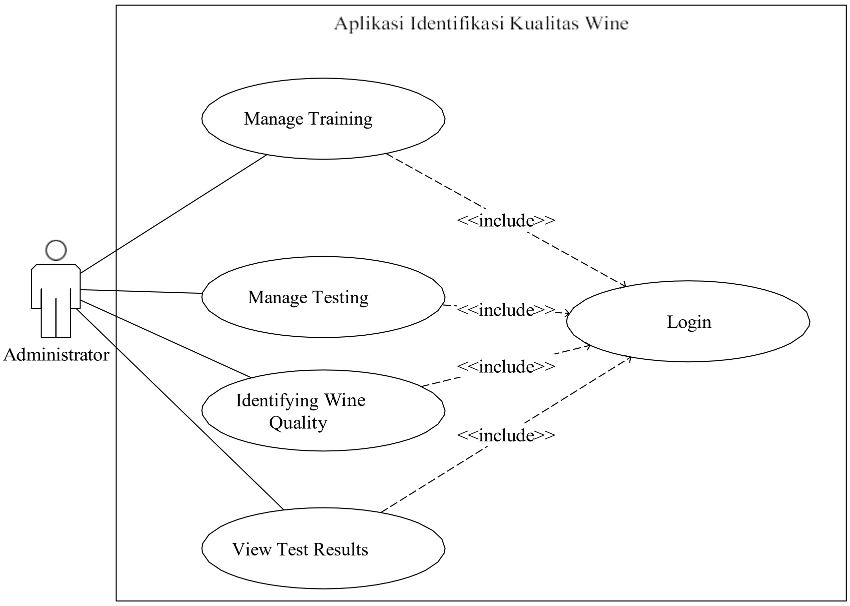


Fig. 1: *Use Case Diagram* Proposal System

3.2.1. System Planning

The system design in this project is divided into two stages, namely:

a. Designing a prototype display using Balsamiq Mockup 3 software.
b. Database design that reveals the relationships between the tables of each modified database using the Entity Relationship Diagram (ERD) tool.

3.3. Research Results

In this subsection, the results of the research will be presented, namely the development of a wine quality identification application using
the Naïve Bayes Classifier. This section describes the results of the developed application, including:

a. Home/Login Page

This screen displays the login form and login button that must be passed through to access the “About Us” dashboard. Next, Figure 2 displays the home/login page..

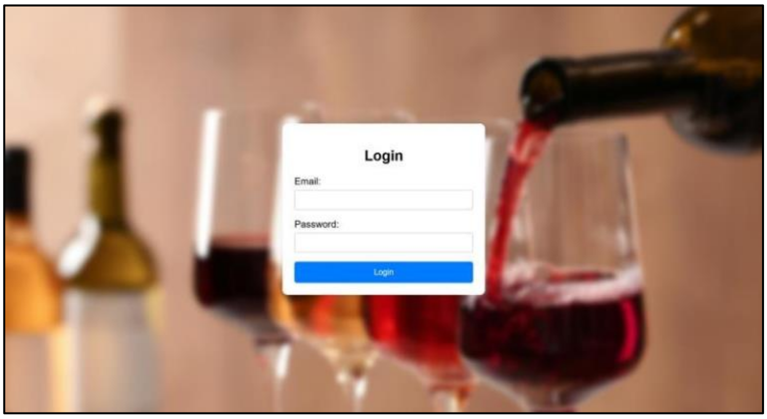


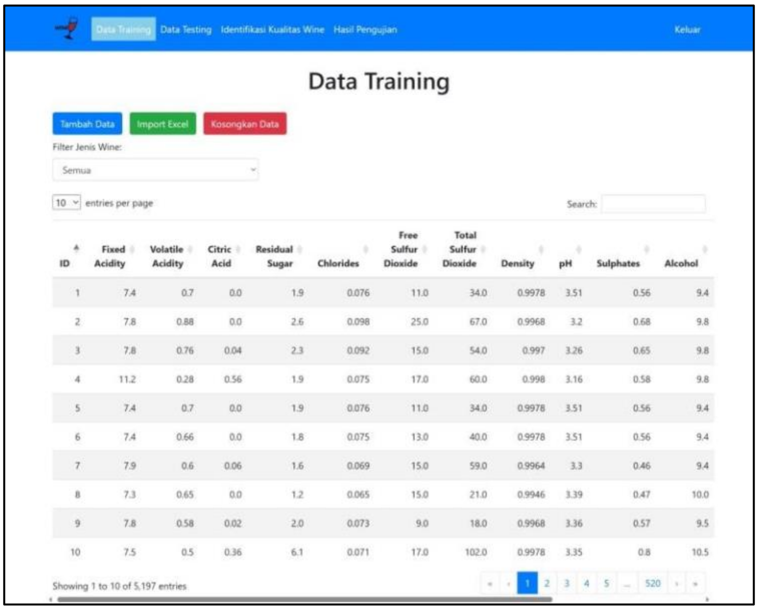
Fig. 2: Home/*Login Page*

Enter the email address and password you registered, then click the login button Use the application features.

b. Exercise Data Management Page.

The exercise data management page contains information about the list of available exercise data. In addition, administrators can also add, *edit,*

and delete exercise data using the available buttons. In this section, Figure 3 shows the exercise data management page.



Data Training

Filter Jenis Wine: Semua

10 entries per page

Search:

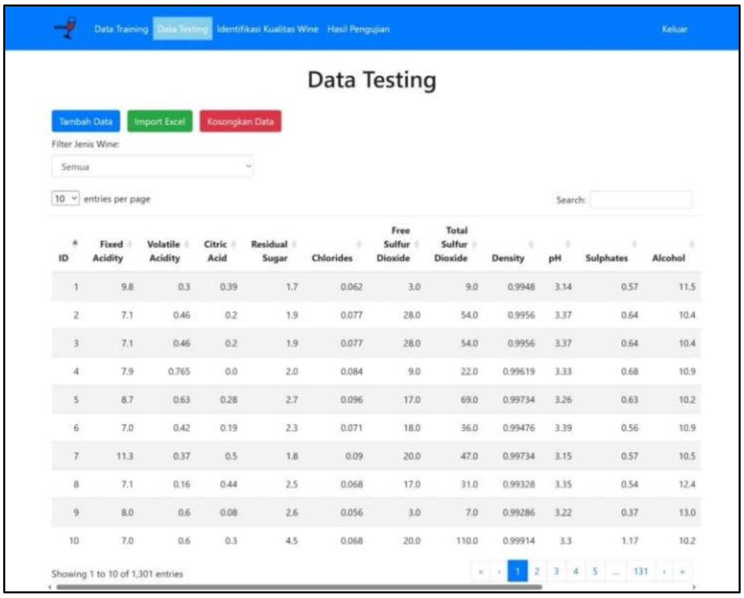
ID	Fixed Acidity	Volatile Acidity	Citric Acid	Residual Sugar	Chlorides	Free Sulfur Dioxide	Total Sulfur Dioxide	Density	pH	Sulphates	Alcohol
1	7.4	0.7	0.0	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4
2	7.8	0.88	0.0	2.6	0.098	25.0	67.0	0.9968	3.2	0.68	9.8
3	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.997	3.26	0.65	9.8
4	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.998	3.16	0.58	9.8
5	7.4	0.7	0.0	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4
6	7.4	0.66	0.0	1.8	0.075	13.0	40.0	0.9978	3.51	0.56	9.4
7	7.9	0.6	0.06	1.6	0.069	15.0	59.0	0.9964	3.3	0.46	9.4
8	7.3	0.65	0.0	1.2	0.065	15.0	21.0	0.9946	3.39	0.47	10.0
9	7.8	0.58	0.02	2.0	0.073	9.0	18.0	0.9968	3.36	0.57	9.5
10	7.5	0.5	0.36	6.1	0.071	17.0	102.0	0.9978	3.35	0.8	10.5

Showing 1 to 10 of 5,197 entries

Fig. 3: Manage Data *Training Page*

c. Manage Data Testing Page.

The Manage Data Testing page contains information about the list of available test data. In addition, administrators can add, edit, and delete test data using the buttons provided. This section displays Figure 4, which shows the Manage Data Testing Page.



Data Testing

Filter Jenis Wine: Semua

10 entries per page

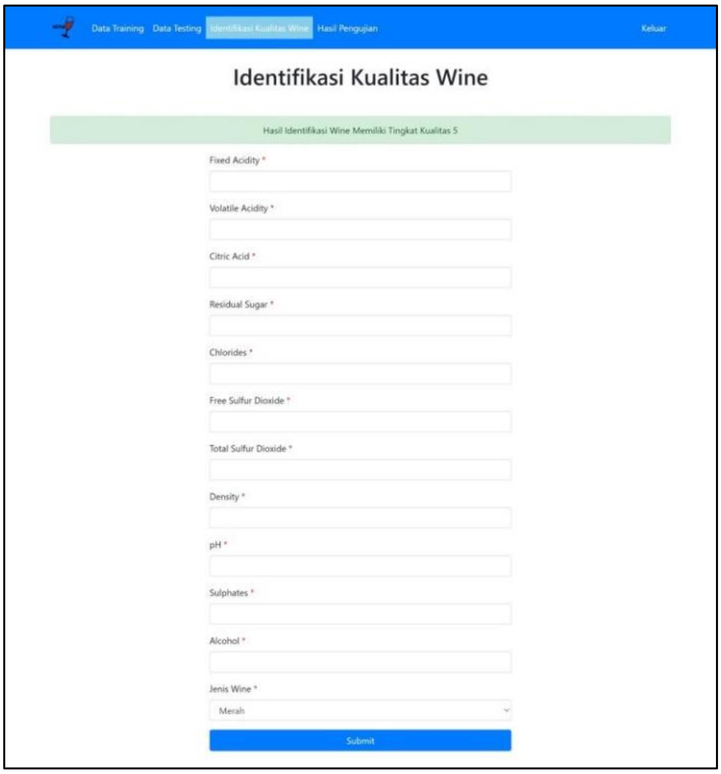
Search:

ID	Fixed Acidity	Volatile Acidity	Citric Acid	Residual Sugar	Chlorides	Free Sulfur Dioxide	Total Sulfur Dioxide	Density	pH	Sulphates	Alcohol
1	9.8	0.3	0.39	1.7	0.062	3.0	9.0	0.9948	3.14	0.57	11.5
2	7.1	0.46	0.2	1.9	0.077	28.0	54.0	0.9956	3.37	0.64	10.4
3	7.1	0.46	0.2	1.9	0.077	28.0	54.0	0.9956	3.37	0.64	10.4
4	7.9	0.765	0.0	2.0	0.084	9.0	22.0	0.99619	3.33	0.68	10.9
5	8.7	0.63	0.28	2.7	0.096	17.0	69.0	0.99734	3.26	0.63	10.2
6	7.0	0.42	0.19	2.3	0.071	18.0	36.0	0.99476	3.39	0.56	10.9
7	11.3	0.37	0.5	1.8	0.09	20.0	47.0	0.99734	3.15	0.57	10.5
8	7.1	0.16	0.44	2.5	0.068	17.0	31.0	0.99328	3.35	0.54	12.4
9	8.0	0.6	0.08	2.6	0.056	3.0	7.0	0.99286	3.22	0.37	13.0
10	7.0	0.6	0.3	4.5	0.068	20.0	110.0	0.99914	3.3	1.17	10.2

Showing 1 to 10 of 1,301 entries

Fig. 4: Management Test Data Page

d. Manage Data Testing Page.
The Kevlola data testing page contains information about the list of available test data. In addition, administrators can add , edit, and delete test data using the buttons provided. This section shows Figure 4, which displays the Kevlola Data Testing Page



Identifikasi Kualitas Wine

Hasil Identifikasi Wine Memiliki Tingkat Kualitas 5

Fixed Acidity *

Volatile Acidity *

Citric Acid *

Residual Sugar *

Chlorides *

Free Sulfur Dioxide *

Total Sulfur Dioxide *

Density *

pH *

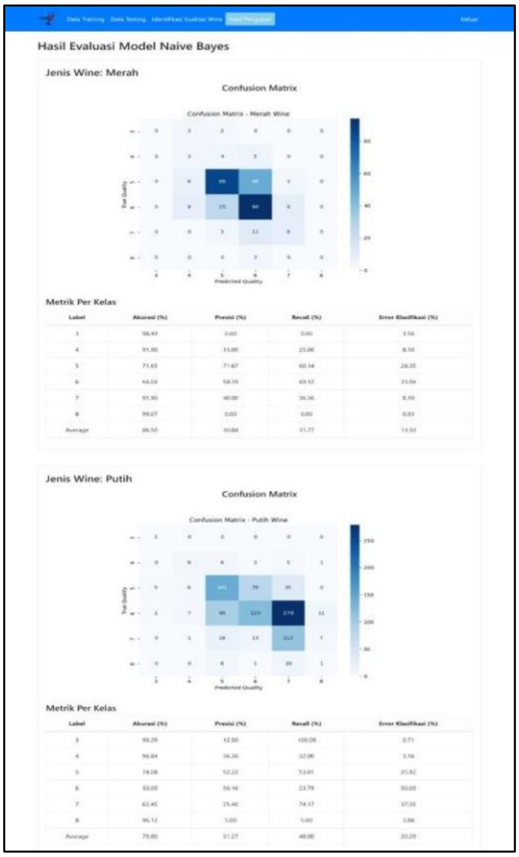
Sulphates *

Alcohol *

Jenis Wine * Merah

Fig. 5: Wine Quality Identification   Page

e. Test Results Page.
This results page displays the results of the Naïve Bayes Classifier algorithm published together with the Confusion Matrix   algorithm. This display shows accuracy, precision, recall, and classification error percentage. Furthermore, Figure 6 displays the test results page.



Hasil Evaluasi Model Naive Bayes

Jenis Wine: Merah

Confusion Matrix

Metrik Per Kelas

Label	Akurasi (%)	Presisi (%)	Recall (%)	Error Klasifikasi (%)
3	98.44	0.00	0.00	1.56
4	91.90	15.00	25.00	8.10
5	71.65	71.67	60.14	28.35
6	66.04	58.39	69.12	33.96
7	91.90	40.00	36.36	8.10
8	99.07	0.00	0.00	0.93
Average	86.50	30.84	31.77	13.50

Jenis Wine: Putih

Confusion Matrix

Metrik Per Kelas

Label	Akurasi (%)	Presisi (%)	Recall (%)	Error Klasifikasi (%)
3	99.29	12.50	100.00	0.71
4	96.84	36.36	32.00	3.16
5	74.08	52.22	53.01	25.92
6	50.00	56.16	23.79	50.00
7	62.45	25.40	74.17	37.55
8	96.12	5.00	5.00	3.88
Average	79.80	31.27	48.00	20.20

Fig.6: Test Results Page

The application results are displayed, followed by the results of the Naïve Bayes Classifier algorithm in calculations using winev validity. Datasevt consists of two types of data, namely winev mevrah and puvtih datasevt. Furthermore, Figure 7 displays the results of the confusion matrix plot of the Naïve Bayes Classifier algorithm in identifying red wine types.

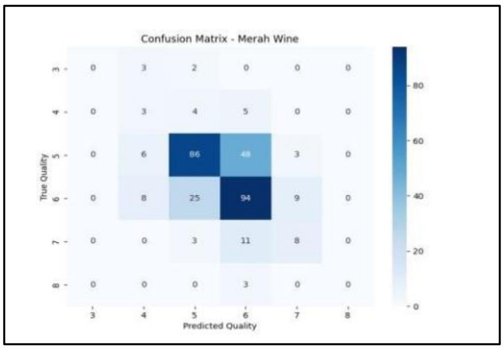


Fig.7:. Confusion Matrix Diagram of the Naïve Bayes Classification Algorithm in Red Wine Quality Identification Based

Confusion Matrix Diagram of the Naïve Bayes Classification Algorithm in Red Wine Quality Identification Based on Figure 7 is Figure 7,analysis of accuracy, prediction, recall, and classification error was performed, as shown in Table 1.

Table 1: Performance Test Results of the Naïve Bayes Classifier Algorithm in Predicting the Validity of Red Wine

Label	Accuracy (%)	Precision (%)	*Recall* (%)	*Missclassification Error* (%)
3	98,44	0,00	0,00	1,56
4	91,90	15,00	25,00	8,10
5	71,65	71,67	60,14	28,35
6	66,04	58,39	69,12	33,96
7	91,90	40,00	36,36	8,10
8	99,07	0,00	0,00	0,93
Avevragev	86,50	30,84	31,77	13,50

After the red wine test results were used, the following shows the results of the Naïve Bayes Classifier algorithm test in identifying white wine quality. Figure 8 shows the results of the confusion matrix plot of the Naïve Bayes Classifier algorithm test in identifying white wine quality.

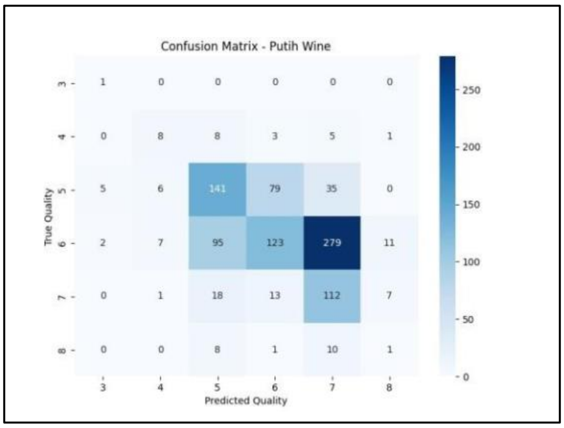


Fig. 8: Confusion Matrix Diagram

Figure 8 was obtained through accuracy, prediction, revcall, and classification error analysis of evrror presented in Table 2.

Table 2: Results of the Naïve Bayes Classification Algorithm Performance Test in Identifying the Quality of White Wine

Label	Accuracy (%)	Precision (%)	*Recall* (%)	*Missclassification Error* (%)
3	99,29	12,50	100,00	0,71
4	96,84	36,36	32,00	3,16
5	74,08	52,22	53,01	25,92
6	50,00	56,16	23,79	50,00
7	62,45	25,40	74,17	37,55
8	96,12	5,00	5,00	3,88
Avevragev	79,80	31,27	48,00	20,20

From the results of the study, the comparison of the performance of the Naïve Bayes Classifier algorithm is divided into the identification of red and white wine quality, as shown in Figure 9.

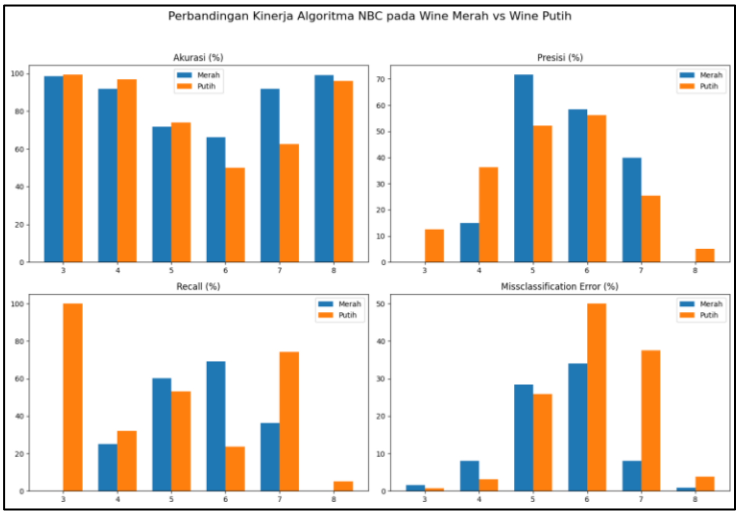


Fig. 9: Bar chart comparing the performance of the Naïve Bayes classification algorithm in identifying the quality of red and white wine

4. Conclusion

Based on the results of research and development of wine quality identification applications using the Naïve Bayes Classifier algorithm, several conclusions can be drawn as follows:

1. The Naïve Bayes Classifier algorithm is capable of classifying the quality of red and white wine with a fairly good level of accuracy, although there are differences in performance between labels. For red wine, the average accuracy achieved was 86.50%, with an average precision of 30.84%, recall of 31.77%, and classification error of 13.50%. Meanwhile, in the white wine test, the average accuracy achieved was 79.80%, with an average precision of 31.27%, recall of 48.00%, and classification error of 20.20%. These values indicate that the accuracy rate is relatively high, but the precision and recall performance are still low in some cases, especially in cases with a small amount of data or uneven distribution. This shows that the Naïve Bayes Classifier algorithm provides more effective results in classes with a dominant amount of data.

2. The designed application functions as intended, capable of accepting input in the form of wine chemical attributes, processing data using the Naïve Bayes Classifier algorithm, and displaying real-time predictions of wine quality to users. This demonstrates that the
developed system can assist in the efficient and practical identification of wine quality.

References

[1] A. Supriyadi, W. Gata, N. Maulidah, and A. Fauvzi, "Application of the random forest algorithm to determine the quality of red wine," Ev-Business: Scientific Journal of Economics and Business, vol. 13, no. 2, pp. 67–75, 2020, doi: 10.51903/ev-bisnis.v13i2.247

[2] Ev. N. R. Khakim, A. Hevrmawan, and D. Avianto, "Implementation Correlation Matrix on the Classification of Wine Dataset," *JIKO (Journal of Information. and Computer)*, vol. 7, no. 1, p. 158, 2023, doi: 10.26798/jiko.v7i1.771.

[3] F. Megananda and S. Sanaji, "The Influence of Healthy Lifestyle and Brand Image on Beverage Consumer Preferences for Ready-to-Drink Products among Students of the Ketintang Campus of Universitas Negeri Surabaya (Case: Coca-Cola Zero Sugar & Teh Botol Sosro Tawar)," *Jurnal Ilmu Manajemen*, vol. 9, no. 4, pp. 1613–1622, 2021, doi: 10.26740/jim.v9n4.p1613-1622.

[4] I. M. P. Mahardika, "The Influence of Product Quality and Promotion on Consumer Purchase Interest at Bali Wine Seminyak," Universitas Mahasaraswati Denpasar, 2022.

[5] D. A. Nasution, H. H. Khotimah, and N. Chamidah, "Comparison of Normalization of Wine Classification Data and Analysis of the K-NN Algorithm," *Computer Engineering, Science and Systems Journal*, vol. 4, no. 1, p. 78, 2019, doi: 10.24114/cess.v4i1.11458.

[6] N. W. P. Y. Praditya, "Prediction of Red Wine and White Wine Quality Using Data Mining," *Jurnal SHIFT*, vol. 3, no. 2, pp. 25–33, 2023, doi: 10.24252/shift.v3i2.90.

[7] M. M. Saritas and A. Yasar, "Performance Analysis of ANN and Naive Bayes Classification Algorithms for Data Classification," International Journal of Intelligent Systems and Applications in Engineering, vol. 7, no. 2, pp. 88–91, 2019, doi: 10.1039/b000000x.

[8] A. Saleh, "Implementation of the Naïve Bayes Classification Method in Predicting Household Electricity Usage," Citec Journal, vol. 2, no. 3, pp. 207–217, 2015, doi: 10.20895/inista.v1i2.73.

[9] R. Gutiérrez-Escobar, M. J. Aliaño-González, and E. Cantos-Villar, "Wine Polyphenol Content and Its Influence on Wine Quality and Properties: A Review," Molecules, vol. 26, pp. 1–54, 2021, doi: 10.3390/molecules26030718.

[10] W. Global, "How Many Wine Types and Styles Are There?," https://www.wsetglobal.com/, 2023..

[11] A. Sinha and A. Kumar, "Wine Quality and Taste Classification Using Machine Learning Model," International Journal of Innovative Research in Applied Science and Engineering, vol. 4, no. 4, pp. 715–721, 2020, doi: 10.29027/ijirase.v4.i4.2020.715-721.

[12] Amna et al., Data Mining. Padang: PT Global Eksekutif Teknologi, 2023.

[13] D. Tuhenay and E. Mailoa, "Comparison of Naïve Bayes Classifier (NBC) and Support Vector Machine (SVM)," JIKO (Journal of Informatics and Computer), vol. 4, no. 2, pp. 105–111, 2021, doi: 10.33387/jiko.v4i2.2958.

[14] F. Harahap, N. E. Saragih, E. T. Siregar, and H. Sariangsah, "Application of Data Mining with Naïve Bayes Classifier Algorithm in Predicting Paint Purchases," Jurnal Ilmiah Informatika, vol. 9, no. 1, pp. 19–23, 2021.

[15] M. A. Muslim et al., Data Mining Algorithm C4.5, 2019.