Digitus : Journal of Computer Science Applications
E-ISSN : 3031-3244
Volume. 2, Issue 4, October 2024
Page No: 222-244

Analysis of Broiler Chicken Production Success Classification Using K-Nearest Neighbors and Naive Bayes Methods at PT. Jandela Jaga Kaloka (Jajaka)

Tukiyat[1], Sajarwo Anggai[2], Agnia Bilqisti[3]
[123]Universitas Pamulang, Indonesia
Correspondent : bilqisti25@gmail.com[3]

Received : September 25, 2024
Accepted : October 15, 2024
Published : October 28, 2024

Citation: Tukir., Agnia, S., & Bilqisti, A. (2024). Analysis of Broiler Chicken Production Success Classification Using K-Nearest Neighbors and Naive Bayes Methods at PT. Jandela Jaga Kaloka (Jajaka). Digitus : Journal of Computer Science Applications, 2(4), 222-244.

https://doi.org/10.61978/digitus.v2i4.396

ABSTRACT: The livestock subsector, particularly broiler chickens, provides animal protein sources in Indonesia. However, low production efficiency, managerial challenges, and productivity fluctuations remain the primary obstacles to achieving sustainability in this sector. This study aims to analyze the success rate of broiler chicken production at PT. Jandela Jaga Kaloka (JAJAKA) using a data mining classification approach with the K-Nearest Neighbors (K-NN) and Naive Bayes algorithms. The research population comprises broiler production data from various branches of PT. JAJAKA, with a sample of 200 datasets selected based on representative criteria. The study employs the hold-out method with data splits of 60:40 and 70:30 for training and testing the models. The success rate of production is classified into three categories: good, less good, and excellent. The findings reveal that the K-NN algorithm outperforms with an accuracy of 92.59%, compared to Naive Bayes, which achieves 76.67%. Regarding recall, K-NN records a value of 96.67%, higher than Naive Bayes at 71.67%. However, Naive Bayes shows slightly better precision (94.29%) than K-NN (93.55%). These results affirm that the K-NN algorithm is more effective for classifying the success rate of broiler chicken production, supporting PT. JAJAKA in making more precise and strategic managerial decisions. Furthermore, this study contributes significantly to developing data mining methods in the poultry farming sector to improve efficiency and productivity sustainably. It provides valuable insights for PT. Jandela Jaga Kaloka will evaluate the success rate of broiler chicken production, facilitating more accurate managerial decision-making.

Keywords: K-Nearest Neighbors, Naive Bayes, Data Mining, Classification, Broiler Chickens.

INTRODUCTION

The livestock subsector is a food commodity sector that contributes significantly to the nation's foreign exchange earnings. Therefore, developing the livestock sector as a source of animal protein is crucial. The target for animal protein consumption is 6 grams per capita per day; however, the current realization falls far short.

Indonesia faces several key challenges in developing its livestock sector. These include the lack of equitable distribution and the failure to meet national nutritional standards. Additionally, the country struggles with limited export opportunities, inadequate feed resources, a low number of high-quality domestic products, substandard product quality, minimal productivity, underutilized human resources, insufficient collaboration among livestock stakeholders, and a lack of

commitment.

These challenges underscore the urgent need for significant efforts to advance the livestock sector, which is needed to meet the nation's nutritional goals and enhance its contribution to economic development (Sopiyanti et al., 2021).

Broiler chickens are among the most popular commodities in Indonesia's livestock agribusiness sector. To date, broiler farming remains one of the most rapidly growing and remarkable ventures in the livestock industry. Since their more intensive development, broiler chickens have outpaced other livestock commodities in meeting the demand for animal protein.

Broiler farming is highly promising, as consumer preference for the taste of broiler chicken is exceptionally high across all social strata. Additionally, the potential profit from broiler farming can be pretty substantial if managed efficiently. (Tamalluddin, 2014).

A survey conducted by the Department of Agriculture reveals that the consumption of broiler chicken meat remains consistently high and tends to increase annually. According to data from Statistics Indonesia (BPS), broiler chicken meat consumption in 2019 reached 12.79 kg per capita per year, an increase from 11.5 kg per capita per year in 2018. This data highlights the significant growth potential and opportunities in broiler chicken farming.

The main challenges small-scale broiler chicken farming businesses face are low efficiency and inadequate implementation of biosecurity measures. In the broiler chicken industry, efficiency is critical in ensuring product competitiveness in terms of price and quality.

Vertical integration, involving small-scale businesses within a partnership model, is expected to help sustain these small-scale operations. Such integration can enhance their capacity to compete and align their practices with industry standards, ensuring long-term viability. (Ilham, 2020)

The production of broiler chickens involves a wide range of complex variables, including feed types, environmental conditions, chicken health, and other factors that can significantly impact production outcomes. Within this context, data mining offers a promising solution to uncover hidden patterns and relationships, providing deeper insights into the dynamics of chicken production(Zai, 2022).

To achieve optimal performance in broiler chickens, the primary influencing factors are breed quality, feed, and management practices. The success of broiler production is reflected in measurable performance indicators, such as mortality rate, feed consumption, final body weight, feed conversion ratio (FCR), and performance index (PI). The performance index serves as a critical benchmark for evaluating broiler management outcomes. A higher performance index (PI) indicates better productivity, leading to more efficient feed utilization and improved overall performance (Baihaqi et al., 2019).

PT. Jandela Jaga Kaloka, a broiler chicken farming company with several branches in West Java, recognizes the need for continuous guidance for farmers at each branch. This necessity arises from differences in farmers' experience levels, with some being highly experienced and others having

minimal experience. The company faces significant challenges in modeling the success rate of broiler production across its branches, primarily due to manual data processing. Production data is classified into three categories based on performance index: very good, sound, and less good. This classification aims to facilitate more targeted and effective decision-making processes.

Another issue faced by PT. Jandela Jaga Kaloka is a decline in production due to overpopulation. Out of the total 40,000 chickens introduced in each production cycle, the company experiences losses with mortality rates ranging from 1,500 to 2,000 chickens. The causes of this mortality include diseases, poor weather conditions, and weaknesses in management practices.

Other issues affecting production include uneven feed distribution, inadequate poultry house sanitation, and suboptimal health monitoring of the chickens. To address these challenges, a practical classification analysis is needed to assess the success rate of production. Through such classification, PT. Jandela Jaga Kaloka can identify key factors influencing success and take the necessary steps to improve production quality while reducing potential losses.

Based on the issues outlined above, the author proposes using Classification Analysis of Production Success through data mining techniques, such as K-Nearest Neighbors (KNN) and Naive Bayes, as an effective tool to classify the success rate of broiler chicken production(Kartarina et al., 2021). With a better understanding of the factors influencing chicken production performance, management can take proactive steps to improve the success rate of production and address potential risks (Akmal et al., 2023).

K-Nearest Neighbor (K-NN) is part of the instance-based learning group. This algorithm is also one of the lazy learning techniques. K-NN works by finding the k nearest (most similar) objects in the training data to a new data point or testing data. A classification system is needed to search for information(Sahar, 2020).

The Naïve Bayes algorithm is one method for classifying data. Bayesian classification is a statistical classification that can be used to predict a class's membership probability. The Naïve Bayes algorithm functions to find knowledge or patterns of similarity in characteristics within a particular group or class (Damuri, 2021).

Based on the literature studies observed by the researchers, data mining has been utilized in several existing studies using various methods(Safitri et al., 2023). For instance, research by Basuki and colleagues successfully applied data mining techniques using the Naïve Bayes algorithm to classify the success rate of broiler chicken production in Riau(Basuki, 2023). The data used in the study were directly obtained from a core company in Riau, comprising 952 datasets, which were divided into three comparison scenarios: 90:10, 80:20, and 70:30. Testing using Google Colab Python yielded the best results with a 90:10 comparison, achieving an accuracy of 89.58%, a precision of 89.89%, and a recall of 90.16%. These findings indicate that the Naïve Bayes classification model applied performs well in predicting the success rate of broiler chicken production. This result positively impacts understanding and managing broiler chicken production in Riau and can serve as a basis for better decision-making in production contexts.

Another study by Basuki et al. (2023) focused on classifying the success rate of broiler chicken production in Riau using the k-nearest Neighbor (k-NN) algorithm. The experimental results showed variations in classification accuracy based on the k values used. The highest accuracy was achieved with k=3k = 3k=3, yielding an accuracy of 86.49%, a precision of 75.00%, and a recall of 70.21%. The study demonstrated that the k value significantly influences the final results. Future studies should use other data normalization methods and develop data splitting using 10-fold cross-validation(Rafi Nahjan et al., 2023; Ramadhan et al., 2020).

Thus, classifying broiler chicken production success rates using data mining techniques is crucial for improving management efficiency in determining production success(Bratsas et al., 2021; Rozi, 2023). This research is expected to provide a clear understanding of the key factors influencing the success of broiler chicken production at PT. Jandala Jaga Kaloka's chicken farm. Applying the K- NN and Naïve Bayes methods is anticipated to significantly support decision-making, leading to more optimal and focused poultry farm management(Priambodo & Falani, 2020).

METHOD

1. Dataset

The dataset used in this study was obtained from PT. Jandela Jaga Kaloka (JAJAKA) manages broiler chicken production data from December 2023 to March 2024. The dataset comprises 200 samples, including information on production results and broiler chicken performance across various PT. JAJAKA branches. It includes 11 key attributes: age, gender, initial weight, final weight, feed intake, mortality rate, temperature, humidity, health status, feed quality, and harvest output(Nugroho & Astuti, 2021).

This dataset aims to provide a comprehensive overview of production patterns and the factors influencing harvest results. Through in-depth analysis, strategies to improve efficiency and productivity in poultry farming can be identified(Kurnia et al., 2019).

Next, classification methods using Altair AI Studio (Rapid Miner) with the K-Nearest Neighbors (KNN) and Naive Bayes algorithms are applied to evaluate the data(Fauzi, 2023). Daily production data collection and environmental information from poultry house sensors play an integral role in supporting this research(Audita et al., 2022; Ilham, 2020). The evaluation of this data is expected to assist PT. JAJAKA in formulating strategic steps to optimize broiler chicken production management (Muharrom, 2023).

2. Data Collection Methods

a. Literature Review

This study was conducted by searching various references from books, journals, and online sources using the keyword "chicken production." These sources include relevant books or journals supporting this thesis's development.

b. Observation

Observation was carried out to gain a direct understanding of the broiler chicken production process at PT Jandela Jaga Kaloka (JAJAKA), covering technical, operational, and managerial aspects. This activity involved direct observation at the farm, such as maintenance processes, feed distribution, and harvesting procedures, to obtain a comprehensive view of the production

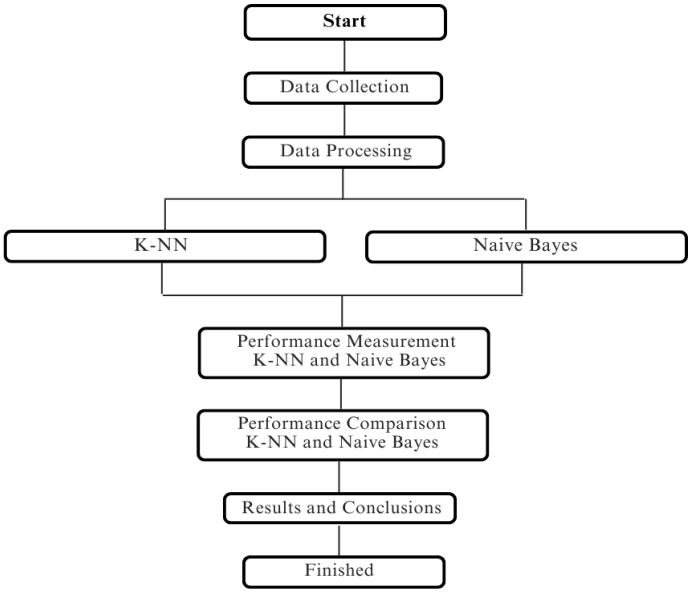
practices implemented.

c. Interviews

Interviews were conducted with various parties involved in the broiler chicken production process, including production managers, operational staff, and maintenance technicians. These interviews aimed to gather in-depth information about the challenges faced, strategies implemented to improve efficiency, and perspectives on the factors affecting production success.

3. Research Design

In this study, the classification of broiler chicken production success at PT. Jandela Jaga Kaloka (JAJAKA) uses the K-Nearest Neighbor (KNN) and Naïve Bayes algorithms. This process consists of several stages, which can be explained in the following flow:



Gambar 1 Alur Penelitian

4. Analysis Technique

In this study, several processes are carried out to analyze the data. The process begins with data collection, including important chicken production variables, such as depletion, average harvest weight, FCR, harvest age, and production index. The collected data then goes through cleaning, organizing, and transformation stages to be ready for analysis. Next, the K-Nearest Neighbors

(KNN) and Naive Bayes methods are applied as classification techniques to analyze and predict production outcomes based on the processed attributes(Noviriandini et al., 2019). The performance of both methods is evaluated using metrics such as accuracy, precision, and recall to assess their effectiveness(Ikhromr, 2023). After that, a performance comparison between KNN and Naive Bayes is conducted to determine which method is superior in classification. Based on the analysis and comparison, conclusions summarize the research findings, provide insights into the strengths and weaknesses of each method, and offer recommendations for future research or applications (Ayu et al., 2024). The research process concludes with preparing the final report, including all the steps and results.

5. Analysis Technique: K-Nearest Neighbors (KNN) Algorithm

K-Nearest Neighbor (KNN) is one of the data mining algorithms that uses a selection of suitable values for *k* (the number of records closest to an object), where the classification in this algorithm depends on the value of *k*. When determining the value of *k*, the most straightforward technique is to repeatedly apply the algorithm until a different *k* value is obtained, and the *k* value with the best performance is chosen (Aini, 2022). The following is Equation 2, which is used in the K-Nearest Neighbor algorithm:

Where the result of d(xi,xj)d(x_i, x_j)d(xi,xj) represents the difference in each attribute squared and summed to the smallest value with the test data(Cahyanti et al., 2021). This algorithm determines the KNN based on the shortest distance from the test sample to the training sample. After gathering the KNN, most of the KNN is selected to make predictions from the test sample. The proximity or distance of the neighbors is usually calculated based on Euclidean distance. The steps to calculate the K-Nearest Neighbor method include:

- Determine the Parameter K
- Calculate the Distance Between Training Data and Test Data

The Euclidean distance is the most commonly used distance calculation in the KNN algorithm. The formula is as follows:

$$euc = \sqrt{((a\_1 - b\_1)^2 + \cdots + (a\_n - b\_n)^2} \quad \text{.....(1)}$$

Where :

pi = training sample / training data

qi = test sample / test data

i = data variable

n = data dimension

- Sorting the formed distances
- Determining the closest distance up to the K-th order
- Pairing the corresponding class
- Finding the number of classes from the nearest neighbors and assigning that class as the class of the data to be evaluated

Accuracy

Accuracy is defined as the degree of closeness between the predicted value and the actual value. The accuracy formula is presented in equation 3.

Precision

Precision is the ratio of relevant items selected to all selected items. Precision can be interpreted as

the match between an information request and the response to that request. The precision formula is shown in equation 4.

Recall

Recall is the ratio of relevant items selected to the total number of relevant items available. Equation 5 explains the recall formula.

F-Measure

The F-measure is the harmonic mean between the precision and recall values. The F-measure is sometimes also called the F1-Score. The F-measure formula is detailed in equation 6.

Accuracy = (TP + TN) / (TP + TN + FP + FN)..... (3)

Precision = TP / (TP + FP).....(4)

Recall = TP / (TP + FN).....(5)

F-Measure = 2 × (Precision × Recall) / (Precision + Recall).....(6)

Variable Description:

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

6. Naive Bayes Algorithm Analysis Technique

Naive Bayes is a simple probability classification that can calculate all possibilities by combining several combinations and frequencies of a value from the obtained database. This algorithm utilizes Bayes' theorem and estimates all independent and mutually exclusive attributes that a value can assign to the class variable. Naive Bayes is a classification method based on probability and calculations discovered by an English scientist, Thomas Bayes, which produces predictions of upcoming probabilities based on previous experiences (Rachman & Handayani, 2021). Bayes' theorem has the following general form:

P(H\X) = P (H\X) P (H) P(H)..... (7)

Where:
X = The class of data that is unknown H = The hypothesis that data X belongs to a specific class P(H|X) = The probability of the hypothesis H given the condition X (posterior probability) P(H) = The probability of the hypothesis H (prior probability) P(X|H) = The probability of X given the hypothesis H P(X) = The probability of X

The Naive Bayesian Classifier assumes that the presence of an attribute (variable) is independent of the presence of other attributes (variables), as the attributes are assumed to be conditionally independent, expressed by the formula:

$$P(X|Ci) = \sum_{k=1}^{n} P(Xk|Ci) \quad \text{.....(8)}$$

After obtaining the results from all the data in each class, the final result can be calculated using the formula:

$$P(X|Ci) = \arg\max P(X|Ci) \times P(Ci). \quad \text{.....(9)}$$

7. Model Testing

A Confusion Matrix test is performed to determine the accuracy of the built model. The Confusion Matrix is a matrix representation that illustrates a classification model's performance by comparing the model's predicted results with the actual data. This matrix consists of four main elements: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). These elements help calculate evaluation metrics such as accuracy, precision, recall, and F1-score, providing a more comprehensive overview of the model's ability to classify data correctly. Visually, the confusion matrix can be constructed as shown in Figure 2 below:

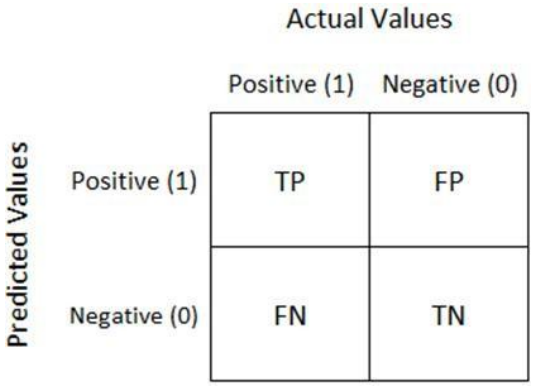


Figure 2: Confusion Matrix

To understand the figure above, this study uses the case of "chicken production success." The study uses a classification model with a dataset that has labels or classes of "Good," "Very Good," and "Not Good." Below is an explanation of the four sections in the confusion matrix shown in the table above:

- **True Positive (TP):** This represents the number of correct predictions, where the model in this study predicts that the production result falls into a specific category (Good, Very Good, or Not Good), which is indeed correct. For example, the model predicts that the production result is "Good," and in reality, it is indeed "Good."
- **True Negative (TN):** Represents the number of correct predictions, where the model in this study predicts that the production result does not fall into a specific category and is indeed correct. For example, the model predicts that the production result is not in the "Not Good" category, and in reality, the production result is indeed not "Not Good."

- **False Positive (FP):** Represents the number of predictions where the model in this study predicts that the production result falls into a specific category, but it is incorrect. For example, the model predicts that the production result is "Very Good," but in reality, the production result is either "Good" or "Not Good."
- **False Negative (FN):** Represents the number of predictions where the model in this study predicts that the production result does not fall into a specific category but is actually incorrect. For example, the model predicts that the production result is not "Good," but in fact, the production result is "Good."

RESULT AND DISCUSSION

At the data collection stage, 200 datasets were obtained from PT. Jandela Jaga Kaloka for classification testing. The collected data includes essential variables such as population size, tonnage, depletion, average harvest weight, and Feed Conversion Ratio (FCR). This data calculates the Performance Index (IP) value to measure production success. The input analysis aims to evaluate each factor affecting the production success rate and to understand the patterns in the data, thus supporting a more accurate classification of broiler chicken production success.

The IP data has two categories and three numerical variables stored in (.xls) file format, as shown in Table 1.

Table 1. Overall Production Dataset

NO	ID	ɔpulation (Head)	LETION (%)	Chickens Harvested (Head)	Average Harvest Weight (KG)	TOTAL PAKAN	Tonnage	FCR	Iarvest Time	IP
1	1223-1	44760	0,9	44359	0,2	7800	8406	0,9	7	288
2	1223-2	44589	1,2	44045	0,2	9700	8192	1,2	7	221
3	1223-3	42028	3,2	40704	0,2	5900	6879	0,8	7	275
4	1223-4	43215	5,2	40971	0,2	6200	6596	0,9	7	232
5	1223-5	44128	0,1	44099	0,5	25600	21410	1,2	14	290
6	1223-6	43794	0,1	43747	0,5	25750	20889	1,2	14	277
7	1223-7	44099	0,1	44060	0,6	29350	24762	1,2	15	316
8	1223-8	43747	0,1	43706	0,5	28150	23645	1,2	15	303
9	1223-9	44060	0,1	44002	0,6	32650	27699	1,2	16	334
10	1223-10	43706	0,1	43662	0,6	31650	27332	1,2	16	338
11	1223-11	40398	0,6	40373	0,5	21050	18592	1,1	14	290
12	1223-12	40644	0,1	40594	0,5	21050	18531	1,1	14	286
13	1223-13	44002	0,1	43950	0,6	36300	28480	1,3	17	299
14	1223-14	43662	0,1	43611	0,7	35300	28500	1,2	17	310

15	1223-15	40373	0,1	40328	0,5	24050	20345	1,2	15	284
16	1223-16	40594	0,1	40552	0,5	24500	20823	1,2	15	291
…..	…..	…..	…..	…..	…..	…..	…..	…..	…..	…..
17	0325-33	7692	0,5	7656	2,8	155250	21322	1,6	39	443

The level of production success, measured using the Performance Index (IP) value, is calculated based on data from various aspects of broiler chicken production. These indicators include population data, tonnage, depletion, average harvest weight, and FCR, which are processed to provide an overview of overall performance and production effectiveness. The IP calculation results are then categorized into Very Good, Good, and Not Good.

Based on interactions with PT. Jandela Jaga Kaloka, the level of broiler chicken production success is classified into three categories based on the Performance Index (IP) value as follows:

Table 2. Production Success Levels

NO	Index Performa (IP)	Description
1	>400	Very Good
2	300 - 399	Good
3	<300	Not Good

From Table 2, the index of production success levels from Haris Wawandaa and Pengusha shows that the IP is classified into three categories: Not Good, Good, and Very Good. The performance evaluation of production based on IP is divided into three categories. The Very Good category applies if the IP value is greater than 400, indicating that the production is in optimal condition with excellent performance. The Good category is assigned if the IP value is between 300 and 399, suggesting that production has reached an adequate target, although there is still room for improvement. Meanwhile, the Not Good category applies if the IP value is less than 300, indicating a significant need for improvement in the production process to meet the expected success standards. This classification helps PT. Jandela Jaga Kaloka evaluate production effectiveness and guide the necessary improvement steps. The data results showing how the broiler production success levels are classified into three categories based on the Performance Index (IP) value can be seen in Table 3.

Table 3. Production Data Selection

NO	ID	LETION (%)	Average Harvest Weight	FCR	Iarvest Time	IP
1	1223-1	0,9	0,2	0,9	7	Not Good
2	1223-2	1,2	0,2	1,2	7	Not Good
3	1223-3	3,2	0,2	0,8	7	Not Good
4	1223-4	5,2	0,2	0,9	7	Not Good
5	1223-5	0,1	0,5	1,2	14	Not Good

6	1223-6	0,1	0,5	1,2	14	Not Good
7	1223-7	0,1	0,6	1,2	15	Good
8	1223-8	0,1	0,5	1,2	15	Good
9	1223-9	0,1	0,6	1,2	16	Good
10	1223-10	0,1	0,6	1,2	16	Good
11	1223-11	0,6	0,5	1,1	14	Not Good
12	1223-12	0,1	0,5	1,1	14	Not Good
13	1223-13	0,1	0,6	1,3	17	Not Good
14	1223-14	0,1	0,7	1,2	17	Good
15	1223-15	0,1	0,5	1,2	15	Not Good
16	1223-16	0,1	0,5	1,2	15	Not Good
…..	…..	…..	…..	…..	…..	…..
17	0325-33	0,5	2,8	1,6	39	Very Good

To create consistency in the scale and range of data values, normalization is performed using Standard Score Normalization (Z-Score). This normalization transforms the data into a distribution resembling a standard normal distribution (Gaussian Distribution), with a mean of zero and a standard deviation of one. This process helps ensure that the data is on a comparable scale, thus preventing certain variables from dominating the analysis results simply because they have a larger scale. The results of the data normalization are displayed in Table 4.

Table 4. Production Data (Normalization)

NO	ID	DEPLETION (%)	Average Harvest Weight	FCR	Harvest Time	IP
1	1223-1	130.595	-175.711	-260.367	-225.380	Not Good
2	1223-2	193.500	-176.271	-120.164	-225.380	Not Good
3	1223-3	572.892	-178.993	-304.728	-225.380	Not Good
4	1223-4	973.907	-180.274	-255.438	-225.380	Not Good
5	1223-5	-.32563	-128.320	-114.687	-135.664	Not Good
6	1223-6	-.24700	-129.601	-.94971	-135.664	Not Good
7	1223-7	-.28631	-116.072	-120.711	-122.847	Good
8	1223-8	-.28631	-119.434	-117.973	-122.847	Good
9	1223-9	-.20768	-105.265	-125.093	-110.031	Good
10	1223-10	-.26666	-105.826	-135.498	-110.031	Good
11	1223-11	.71623	-132.323	-149.190	-135.664	Not Good
12	1223-12	-.22734	-132.963	-147.000	-135.664	Not Good
13	1223-13	-.22734	-102.303	-.71421	-.97214	Not Good
14	1223-14	-.22734	-101.423	-.91137	-.97214	Good

15	1223-15	-.24700	-125.278	-122.354	-122.847	Not Good
16	1223-16	-.26666	-123.837	-125.640	-122.847	Not Good
…..	…...	…...	…...	…...	…...	…...
17	0325-33	.46068	239.837	109.309	184.751	Very Good

1. Implementation of the K-Nearest Neighbor Method

In this stage, three classification tests will be conducted using the K-Nearest Neighbor (KNN) algorithm with the Altair AI Studio (RapidMiner) tool, applying the hold-out data splitting method of 60:40, 70:30, and 80:20 between training and testing data. The results of this testing stage can be seen in Table 4.5.

Table 5. K-Nearest Neighbor Testing Results

accuracy: 92.59%

	true Baik	true Kurang Baik	true Sangat Baik	class precision
pred. Baik	58	2	2	93.55%
pred. Kurang Baik	1	9	0	90.00%
pred. Sangat Baik	1	0	8	88.89%
class recall	96.67%	81.82%	80.00%	

From Table 5, it can be explained that the broiler chicken production data using the K-Nearest Neighbor method with a 60:40 training-to-testing data ratio resulted in the following predictions: In the Good category, there were 58 data points, in the Not Good category, there were 9 data points, and in the Very Good category, there were 8 data points. Based on the accuracy value of the K-NN mining algorithm calculation, the explanation is as follows:

Precision = TP/(TP+FP)

= 58/(58+2+2)

= 58/62 X 100%

= 93,55%

Recall = TP/(TP+FN)

= 58/(58+1+1)

= 58/60 X 100%

= 96,67%

Akurasi= (TP+TN)/(TP+TN+FP+FN)

= (58+9+8)/(58+2+2+1+9+0+1+0+8)

= 75/81 X 100%

= 92,59%

Based on the results of the K-Nearest Neighbor (K-NN) method test with a 60:40 training-to-testing data split, an accuracy of 92.59% was achieved from the total data. Additionally, the recall value reached 96.96%, and the precision was 93.55%. A subsequent test was performed with a 70:30 hold-out data split, as shown in Table 6.

Table 6. K-Nearest Neighbor Testing Results 70:30

accuracy: 86.67%

	true Baik	true Kurang Baik	true Sangat Baik	class precision
pred. Baik	41	2	2	91.11%
pred. Kurang Baik	3	6	0	66.67%
pred. Sangat Baik	1	0	5	83.33%
class recall	91.11%	75.00%	71.43%	

From Table 4.6, it can be explained that the broiler chicken production data using the K-Nearest Neighbor (K-NN) method with a 70:30 training-to-testing data ratio resulted in the following predictions: In the Good category, there were 41 data points, in the Not Good category, there were 6 data points, and in the Very Good category, there were 5 data points. Based on the accuracy value of the K-NN mining algorithm calculation, the explanation is as follows:

Precision = 41/(41+2+2)

= 41/45 X 100%

= 91,11%

Recall = 41/(41+3+1)

= 41/45 X 100%

= 91,11%

Akurasi = (41+6+5)/(41+2+2+3+6+0+1+0+5)

= 52/60 X 100%

= 86,67%

Based on the results of the K-Nearest Neighbor (K-NN) method test with a 70:30 training-to-testing data split, an accuracy of 86.67% was achieved from the total data. Additionally, the recall value reached 91.11%, and the precision was 91.11%. A subsequent test was performed with an 80:20 hold-out data split, as shown in Table 7.

Table 7. K-Nearest Neighbor Testing Results 80:20

accuracy: 87.50%

	true Baik	true Kurang Baik	true Sangat Baik	class precision
pred. Baik	28	1	2	90.32%
pred. Kurang Baik	1	4	0	80.00%
pred. Sangat Baik	1	0	3	75.00%
class recall	93.33%	80.00%	60.00%	

From Table 4.7, it can be explained that the broiler chicken production data using the K-Nearest Neighbor (K-NN) method with an 80:20 training-to-testing data ratio resulted in the following predictions: In the Good category, there were 28 data points, in the Not Good category, there were 4 data points, and in the Very Good category, there were 3 data points.

Based on the accuracy value of the K-NN mining algorithm calculation, the explanation is as follows:

Precision = 28/(28+1+2)

= 28/31 X 100%

= 90,32%

Recall = 28/(28+1+1)

= 28/30 X 100%

= 93,33%

Akurasi = (28+4+3)/(28+1+2+1+4+0+1+0+3)

= 35/40 X 100%

= 87,50%

Based on the results of the K-Nearest Neighbor (K-NN) method test with a 70:30 training-to-testing data split, an accuracy of 87.50% was achieved from the total data. Additionally, the recall

value reached 93.33%, and the precision was 90.32%.

Through the classification process conducted to measure the success level of broiler chicken production using the K-Nearest Neighbor (K-NN) Classifier algorithm, the performance results of each test will be obtained. Below is the table and graph comparing the results with the three data split scenarios, as shown in Table 8.

Table 8. Comparison of K-NN Algorithm Performance Results

PERCENTAGE	MODEL		
	K-Nearest Neighbor (K-NN)		
	Accuracy	*Recall*	*Precision*
60% : 40%	92,59%	96,67%	93,55%
70% : 30%	86,67%	91,11%	91,11%
80% : 20%	87,50%	93,33%	90,32%

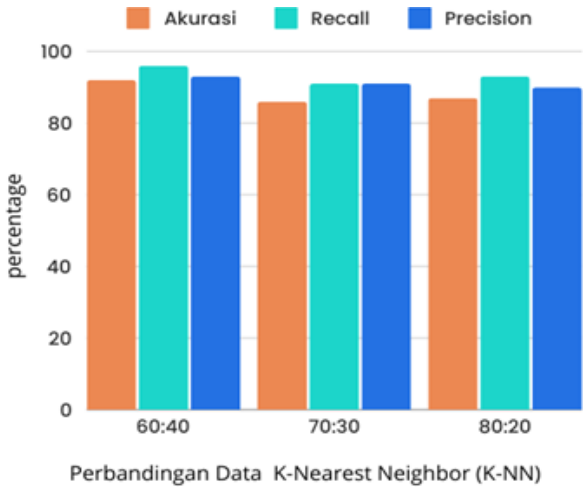


Figure 3. Graph of K-NN Algorithm Performance Comparison

As shown in Table 4.8 and Figure 4.14, based on the comparison results of the three classification tests conducted using the K-Nearest Neighbor (K-NN) algorithm on the Altair AI Studio (RapidMiner) tool with the hold-out data splitting method of 60:40, 70:30, and 80:20, the highest accuracy of 92.59% was achieved with the 60:40 data split, with a recall value of 96.67% and precision of 93.55%.

2. Implementation of the Naive Bayes Method

this stage, three classification tests will be conducted using the Naïve Bayes algorithm with the Altair AI Studio (RapidMiner) tool, applying the hold-out data splitting method of 60:40, 70:30, and 80:20 between training and testing data. The results of this testing stage can be seen in Table 9.

Table 9. Naive Bayes Testing Results 60:40

accuracy: 74.07%

	true Baik	true Kurang Baik	true Sangat Baik	class precision
pred. Baik	43	1	3	91.49%
pred. Kurang Baik	10	10	0	50.00%
pred. Sangat Baik	7	0	7	50.00%
class recall	71.67%	90.91%	70.00%	

From Table 9, it can be explained that the broiler chicken production data using the Naive Bayes method with a 60:40 training-to-testing data ratio resulted in the following predictions: In the Good category, there were 43 data points, in the Not Good category, there were 10 data points, and in the Very Good category, there were 7 data points.

Based on the accuracy value of the Naive Bayes mining algorithm calculation, the explanation is as follows:

Precision = 43/(43+1+3)

= 43/47 X 100%

= 91,49%

Recall = 43/(43+10+7)

= 43/60 X 100%

= 71,67%

Akurasi = (43+10+7)/(43+1+3+10+10+0+7+0+7)

= 60/81 X 100%

= 74,07%

Based on the results of the Naive Bayes method test with a 60:40 training-to-testing data split, an accuracy of 74.07% was achieved from the total data. Additionally, the recall value reached 71.67%, and the precision was 91.49%. A subsequent test was performed with a 70:30 hold-out data split, as shown in Table 10.

Table 10. Naive Bayes Testing Results 70:30

accuracy: 76.67%

	true Baik	true Kurang Baik	true Sangat Baik	class precision
pred. Baik	33	1	1	94.29%
pred. Kurang Baik	7	7	0	50.00%
pred. Sangat Baik	5	0	6	54.55%
class recall	73.33%	87.50%	85.71%	

From Table 10, it can be explained that the broiler chicken production data using the Naive Bayes method with a 70:30 training-to-testing data ratio resulted in the following predictions: In the Good category, there were 33 data points, in the Not Good category, there were 7 data points, and in the Very Good category, there were 6 data points.

Based on the accuracy value of the Naive Bayes mining algorithm calculation, the explanation is as follows:

Precision = 33/(33+1+1)

= 33/35 X 100%

= 94,29%

Recall = 33/(33+7+5)

= 33/45 X 100%

= 73,33%

Akurasi = (33+7+6)/(33+1+1+7+7+0+5+0+6)

= 46/60 X 100%

= 76,67%

Based on the results of the Naive Bayes method test with a 70:30 training-to-testing data split, an accuracy of 76.67% was achieved from the total data. Additionally, the recall value reached 73.33%, and the precision was 94.29%. A subsequent test was performed with an 80:20 hold-out data split, as shown in Table 11.

Table 11. Naive Bayes Testing Results 80:20

accuracy: 70.00%

	true Baik	true Kurang Baik	true Sangat Baik	class precision
pred. Baik	20	1	1	90.91%
pred. Kurang Baik	6	4	0	40.00%
pred. Sangat Baik	4	0	4	50.00%
class recall	66.67%	80.00%	80.00%	

From Table 11, it can be explained that the broiler chicken production data using the Naive Bayes method with an 80:20 training-to-testing data ratio resulted in the following predictions: In the Good category, there were 20 data points, in the Not Good category, there were 4 data points, and in the Very Good category, there were 4 data points.

Based on the accuracy value of the Naive Bayes mining algorithm calculation, the explanation is as follows:

Precision = 20/(20+1+1)

= 20/22 X 100%

= 90,91%

Recall = 20/(20+6+4)

= 20/30 X 100%

= 66,67%

Akurasi = (20+4+4)/(20+1+1+6+4+0+4+0+4)

= 28/40 X 100%

= 70,00%

Based on the results of the Naive Bayes method test with an 80:20 training-to-testing data split, an accuracy of 70.00% was achieved from the total data. Additionally, the recall value reached 66.67%, and the precision was 90.91%. Through the classification process performed to measure the success level of broiler chicken production using the Naive Bayes Classifier algorithm, the performance results of each test will be obtained. Below is the table and graph of the comparison results with the three data splitting scenarios, as shown in Table 12.

Table 12. Performance Comparison Results of Naive Bayes Algorithm

PERCENTAGE	MODEL		
	Naive Bayes Classifier		
	Accuracy	*Recall*	*Precision*
60% : 40%	74,07%	71,67%	91,49%
70% : 30%	76,67%	73,33%	94,29%
80% : 20%	70,00%	66,67%	90,91%

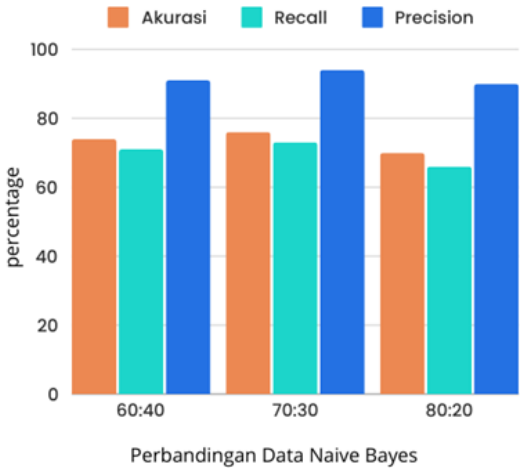


Figure 4. Performance Comparison Graph of Naive Bayes Algorithm

As shown in Table 12 and Figure 4, based on the comparison results from the three classification tests conducted using the Naive Bayes algorithm on Altair AI Studio (RapidMiner) with the 60:40, 70:30, and 80:20 hold-out data splitting methods, the highest accuracy of 76.67% was obtained with the 60:40 data split, with a recall value of 73.33% and precision of 93.55%.

3. Compilation of the Model

By performing a classification process to measure the success rate of broiler chicken production using the K-Nearest Neighbor (K-NN) and Naive Bayes Classifier algorithms, the performance results of each algorithm were obtained. The following table and graph show the comparison results under three data partitioning scenarios, which can be seen in Table 13 and Figure 5.

Table 13: Algorithm Performance Comparison Results

PERCENTAGE	MODEL					
	K-Nearest Neighbor (K-NN)			Naive Bayes Classifier		
	Accuracy	*Recall*	*Precision*	Accuracy	*Recall*	*Precision*

60% : 40%	92,59%	96,67%	93,55%	74,07%	71,67%	91,49%
70% : 30%	86,67%	91,11%	91,11%	76,67%	73,33%	94,29%
80% : 20%	87,50%	93,33%	90,32%	70,00%	66,67%	90,91%

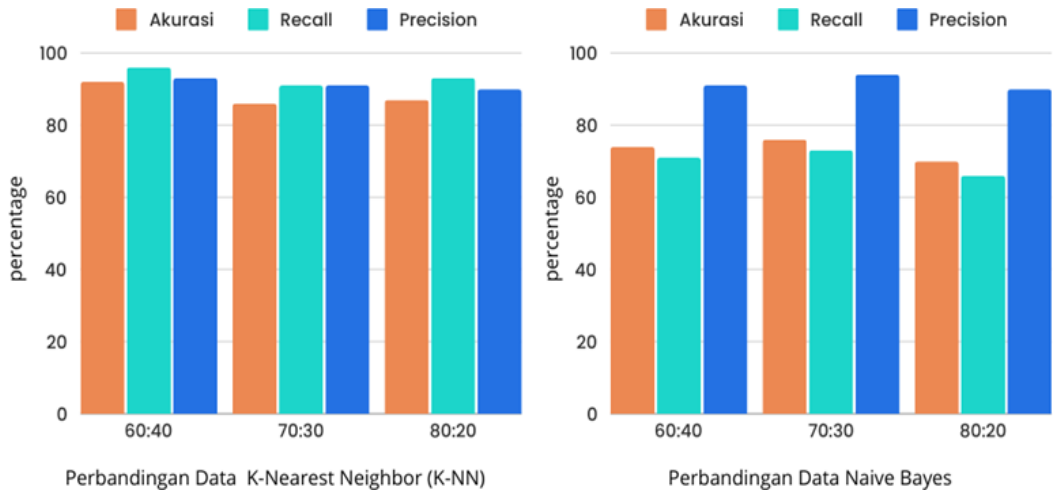


Figure 5: Algorithm Performance Comparison Chart

The best algorithm based on the results obtained between K-Nearest Neighbor (K-NN) and Naive Bayes is K-Nearest Neighbor (K-NN). This algorithm demonstrated a higher accuracy rate, reaching 92.59% in the 60:40 hold-out data partitioning, compared to Naive Bayes, which achieved only 76.67% accuracy in the 70:30 hold-out partitioning.

In terms of recall, K-NN also excelled with the highest value of 96.67% in the 60:40 data partitioning, while Naive Bayes only reached 71.67% under the same scenario. However, the best precision score was achieved by Naive Bayes at 94.29% in the 70:30 hold-out partitioning, which was slightly higher than K-NN's precision of 93.55% in the 60:40 partitioning.

CONCLUSION

This study concludes that the success rate of broiler chicken production at PT. Jandela Jaga Kaloka (JAJAKA) is dominated by the "Good" IP category, indicating that the production process has met established standards(Argina, 2020).

Based on the analysis of algorithm performance, K-Nearest Neighbor (K-NN) proved to be superior to Naive Bayes in classifying the success rate of broiler chicken production. K-NN achieved the highest accuracy of 92.59% in the 60:40 hold-out data partition scenario, outperforming Naive Bayes, which attained 76.67% accuracy in the 70:30 partition. Additionally, K-NN exhibited a recall value of 96.67%, significantly higher than Naive Bayes' recall of 71.67%. However, in terms of precision, Naive Bayes was slightly better, achieving 94.29% in the 70:30 data partition compared to K-NN's 93.55% in the 60:40 scenario(Ardiansyah & Walim, 2018).

Overall, the K-Nearest Neighbor (K-NN) algorithm was selected as the best method for classifying broiler chicken production success at PT. JAJAKA, with Naive Bayes serving as a viable alternative(Amam, 2019). This research is important to provide a crucial foundation for improving production management efficiency and optimizing decision-making through the appropriate utilization of data mining algorithms(Ayuningtyas et al., 2022).

Limitations encountered during the research include:

1. **Limited dataset size:** The study used 200 datasets, which, while representative, might not be large enough to capture the variability of broiler chicken production data at PT. JAJAKA.
2. **Non-optimized algorithm parameters:** The study did not delve into parameter optimization for K-NN (e.g., selection of the value of kkk) or Naive Bayes, which could influence algorithm performance.
3. **Focus on only two algorithms:** The research only compared K-NN and Naive Bayes, whereas other methods like Random Forest or Neural Networks might provide more accurate or relevant results.
4. **Limited analysis factors:** Other factors influencing production success, such as environmental conditions, feed availability, or chicken stress levels, were not analyzed in- depth.

Recommendations for future research include:

1. **Increasing dataset size:** Future studies are advised to use a larger dataset covering a more extended period to improve data representation and result accuracy.
2. **Exploring other algorithms:** Comparing the performance of additional algorithms, such as Random Forest, Gradient Boosting, or Deep Learning, is necessary to identify superior classification methods.
3. **Optimizing algorithm parameters:** Future research should focus on tuning parameters for the algorithms used, such as the kkk value in K-NN, to enhance classification performance.
4. **Enriching analysis factors:** Incorporating additional variables, such as coop temperature, water quality, or health monitoring frequency, is recommended to provide a more comprehensive analysis.
5. **Integrating data mining-based systems:** Developing a data mining-based system that management at PT. JAJAKA can directly use for real-time monitoring and evaluating production success.

By addressing these limitations, future studies are expected to provide more comprehensive and practical solutions to support the sustainability of the broiler chicken farming sector.

REFERENCE

Aini, D. N. (2022). Seleksi Fitur untuk Prediksi Hasil Produksi Agrikultur pada Algoritma K-Nearest Neighbor (KNN)'. *Jurnal Sistem Komputer dan Informatika (JSON*, *4*(1), 140. https://doi.org/10.30865/json.v4i1.4813.

Akmal, K., Faqih, A., & Dikananda, F. (2023). Perbandingan Metode Algoritma Naïve Bayes Dan K-

Nearest Neighbors Untuk Klasifikasi Penyakit Stroke'. *JATI (Jurnal Mahasiswa Teknik Informatika*, *7*(1), 470–477. https://doi.org/10.36040/jati.v7i1.6367.

Amam, A. (2019). Usaha Ternak Ayam Pedaging Sistem Kemitraan Pola Dagang Umum: Pemetaan Sumber Daya dan Model Pengembangan'. *Sains Peternakan*, *17*(2), 5. https://doi.org/10.20961/sainspet.v17i2.26892.

Ardiansyah, D., & Walim, W. (2018). Algoritma C4.5 untuk Klasifikasi Calon Peserta Lomba Cerdas Cermat Siswa SMP dengan Menggunakan Aplikasi RapidMiner'. *Jurnal Inkofar*, *1*(2), 5–12. https://politeknikmeta.ac.id/meta/ojs/index.php/inkofar/article/view/29/45.

Argina, A. M. (2020). Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes'. *Indonesian Journal of Data and Science*, *1*(2), 29–33. https://doi.org/10.33096/ijodas.v1i2.11.

Audita, I., Damanik, I. S., & IRAWAN, E. K. A. (2022). Pemetaan Hasil Produksi Buah-Buahan Dengan Teknik Data Mining K-Medoids'. *Jurnal Teknik Mesin, Industri, Elektro dan Informatika*, *1*(3), 39–53. https://doi.org/10.55606/jtmei.v1i3.535.

Ayu, P., Purnama, W., & Putra, T. A. (2024). Klasifikasi Penjualan Produk Menggunakan Algoritma Naive Bayes pada Konter HP Bayu Cell'. *Remik: Riset dan E-Jurnal Manajemen Informatika Komputer*, *8*(1), 286–292. https://doi.org/10.33395/remik.v8i1.13207.

Ayuningtyas, N., R, N., Basysyar, M., & F. (2022). Penerapan Data Mining pada Penjualan Produk MS Glow Menggunakan Metode Naive Bayes untuk Strategi Pemasaran'. *Jurnal Accounting Information System (AIMS*, *5*(2), 157–166. https://doi.org/10.32627/aims.v5i2.503.

Baihaqi, D. I., Handayani, A. N., & Pujianto, U. (2019). *Perbandingan Metode Naïve Bayes dan C4.5 untuk Memprediksi Mortalitas pada Peternakan Ayam Broiler'* (Vol. 10, Issue 1, pp. 383–390). Jurnal Teknik Mesin, Elektro dan Ilmu Komputer. https://doi.org/10.24176/simet.v10i1.2846.

Basuki, B. (2023). Klasifikasi Tingkat Keberhasilan Produksi Ayam Broiler di Riau Menggunakan Algoritma Naïve Bayes'. *Jurnal Sistem Komputer dan Informatika (JSON*, *4*(3), 510. https://doi.org/10.30865/json.v4i3.5800.

Bratsas, C., Chondrokostas, E., Koupidis, K., & Antoniou, I. (2021). The use of national strategic reference framework data in knowledge graphs and data mining to identify red flags. *Data*, *6*(1), 1–20. https://doi.org/10.3390/data6010002

Cahyanti, F. L. D., Gata, W., & Sarasati, F. (2021). Implementasi Algoritma Naïve Bayes dan K-Nearest Neighbor Dalam Menentukan Tingkat Keberhasilan Immunotherapy Untuk Pengobatan Penyakit Kanker Kulit'. *Jurnal Ilmiah Universitas Batanghari Jambi*, *21*(1), 259. https://doi.org/10.33087/jiubj.v21i1.1189.

Damuri, A. (2021). Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako'. *JURIKOM (Jurnal Riset Komputer*, *8*(6), 219. https://doi.org/10.30865/jurikom.v8i6.3655.

Fauzi, A. (2023). *Sistem Pakar Diagnosa Penyakit Hewan Ternak Sapi Menggunakan Metode Naïve Bayes'* (Vol. 1, Issue 1, pp. 281–284).

Ikhromr, F. N. (2023). Implementasi Data Mining Untuk Memprediksi Penyakit Diabetes Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor'. *INTECOMS: Journal of Information Technology and Computer Science*, *6*(1), 416–428. https://doi.org/10.31539/intecoms.v6i1.5916.

Ilham, N. (2020). Implementasi Hubungan Antara Pelaku Usaha pada Usaha Kemitraan Ayam Pedaging Skala Kecil di Indonesia'. In *Relationships between Stakeholders in Small Scale Broiler Business Partnerships in Indonesia.* http://download.garuda.kemdikbud.go.id/article.php?article=2203772&val=7169&title=Implementation

Kartarina, K., Sriwinarti, N. K., & Juniarti, N. L. P. (2021). Analisis Metode K-Nearest Neighbors (K-NN) Dan Naive Bayes Dalam Memprediksi Kelulusan Mahasiswa'. *JTIM: Jurnal Teknologi Informasi dan Multimedia*, *3*(2), 107–113. https://doi.org/10.35746/jtim.v3i2.159.

Kurnia, E., Daulay, R., & Nugraha, F. (2019). Dampak Faktor Motivasi dan Fasilitas Kerja Terhadap Produktivitas Kerja Karyawan Pada Badan Usaha Milik Negara di Kota Medan'. *Proseding Seminar Nasional Kewirausahaan*, *1*(1), 365–372.

Muharrom, M. (2023). Analisis Komparasi Algoritma Data Mining Naive Bayes, K-Nearest Neighbors dan Regresi Linier Dalam Prediksi Harga Emas'. *Bulletin of Information Technology (BIT*, *4*(4), 430–438. https://doi.org/10.47065/bit.v4i4.986.

Noviriandini, A., Handayani, P., & Syahriani. (2019). *Prediksi Penyakit Liver Dengan Menggunakan Metode Naïve Bayes Dan K-Nearest Neighbour (KNN)'.* Prosiding TAU SNAR-TEK Seminar Nasional Rekayasa dan Teknologi.

Nugroho, M., & Astuti, F. Y. (2021). Analisis Kelayakan Usaha Peternakan Ayam Pedaging'. *Jurnal Manajemen Dayasaing*, *23*(1), 59–72. https://doi.org/10.23917/dayasaing.v23i1.14065.

Priambodo, S. A., & Falani, A. Z. (2020). Pemanfaatan Data Mining Untuk Klasterisasi Potensi Produksi Beras di Kabupaten Blitar dengan Menggunakan Metode Fuzzy C-Means'. *Jurnal SPIRIT*, *12*(2), 30–36. https://www.blitarkab.go.id/.

Rachman, R., & Handayani, R. N. (2021). Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM'. *Jurnal Informatika*, *8*(2), 111–122. https://doi.org/10.31294/ji.v8i2.10494.

Rafi Nahjan, M., Heryana, N., & Voutama, A. (2023). Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell'. *JATI (Jurnal Mahasiswa Teknik Informatika*, *7*(1), 101–104. https://doi.org/10.36040/jati.v7i1.6094.

Ramadhan, D., Hanuranto, A. T., & Mayasari, R. (2020). *Implementasi Kandang Ayam Pintar Berbasis Internet of Things'.* e-Proceeding of Engineering.

Rozi, M. F. (2023). Penerapan Data Mining Menggunakan Metode Naive Bayes Untuk Klasifikasi Data Penentuan Hasil Penjualan Dalam Strategi Pemasaran'. *Jurnal Komputer Teknologi Informasi dan Sistem Informasi (JUKTISI*, *2*(2), 444–454. https://doi.org/10.62712/juktisi.v2i2.137.

Safitri, D., Hilabi, S. S., & Nurapriani, F. (2023). Analisis Penggunaan Algoritma Klasifikasi Dalam

Prediksi Kelulusan Menggunakan Orange Data Mining'. *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, *8*(1), 75–81. https://doi.org/10.36341/rabit.v8i1.3009.

Sahar, S. (2020). Analisis Perbandingan Metode K-Nearest Neighbor dan Naïve Bayes Classifier Pada Dataset Penyakit Jantung'. *Indonesian Journal of Data and Science*, *1*(3), 79–86. https://doi.org/10.33096/ijodas.v1i3.20.

Sopiyanti, A., Hendro Pudjiantoro, T., & Santikarama, I. (2021). Klasifikasi Penjualan Jus Dengan Menggunakan Algoritma K-Nearest Neighbor (K-NN) Untuk Penerapan Konsep Up-Selling'. *Seminar Nasional Informatika dan Aplikasinya (SNIA*.

Zai, C. (2022). Implementasi Data Mining Sebagai Pengolahan Data'. *Portaldata.org*, *2*(3), 12. http://portaldata.org/index.php/portaldata/article/view/107.