JPPIPA 10(10) (2024)

Jurnal Penelitian Pendidikan IPA

Journal of Research in Science Education

http://jppipa.unram.ac.id/index.php/jppipa/index

Analysis of Naïve Bayes and K-Nearest Neighbors Algorithms for Classifying Fishermen Aid Eligibility

Muhammad Nasrullah[1*], Bayu Surarso[2], Oky Dwi Nurhayati[3]

[1] Magister Information System, Universitas Diponegoro, Semarang, Indonesia
[2] Graduate School and Department of Mathematics, Universitas Diponegoro, Semarang, Indonesia
[3] Computer Engineering, Faculty of Engineering, Universitas Diponegoro, Semarang, Indonesia

Received: April 02, 2024
Revised: August 12, 2024
Accepted: October 25, 2024
Published: October 31, 2024

Corresponding Author:
Muhammad Nasrullah
muhammadnasrullahsib@gmail.com

DOI: 10.29303/jppipa.v10i10.8818

Abstract: This article analyzes the use of data mining with Naïve Bayes and K-Nearest Neighbor (KNN) algorithms to build classification models and evaluate their performance in identifying fishermen eligible for aid. The study aims to compare the effectiveness of these algorithms in handling imbalanced datasets using the Synthetic Minority Over-sampling Technique (SMOTE). The research applies SMOTE to improve the balance of the dataset before classification. Without SMOTE, Naïve Bayes achieved an accuracy of 97.01%, precision of 94.16%, recall of 96.67%, and F1-score of 95.39%. KNN, on the other hand, reached an accuracy of 94.04%, precision of 94.53%, recall of 86.00%, and F1-score of 90.06%. After applying SMOTE, both algorithms improved: Naïve Bayes attained an accuracy of 98.33%, precision of 96.86%, recall of 100.00%, and F1-score of 98.49%, while KNN reached an accuracy of 96.90%, precision of 97.72%, recall of 96.19%, and F1-score of 96.94%. The results show that Naïve Bayes, with SMOTE, outperforms KNN in managing data imbalance and accurately classifying eligible fishermen for aid.

Keywords: Classification; F1-score; Inbalance Datset; K-Nearst Neighbor; Naïve Bayes; SMOTE

Introduction

Currently, advances in information technology have developed rapidly, bringing new challenges and opportunities for various sectors, including the marine and fisheries sector. In the midst of the dynamics of the need for fast and precise decision making, it is important to have a system that is able to predict the feasibility of providing assistance to fishermen with high accuracy (Putri et al., 2021). Fishing communities are groups of people who live in coastal areas and their lives depend on natural resources in the sea, such as fish, shrimp, seaweed, shellfish, coral, and various other marine wealth (Fauziah et al., 2024; Hutajulu, 2023).

The number of fishing gears owned by a fisherman such as the number of boats, number of motorized canoes, number of canoes, number of fishing rods, and number of nets can be an indication of fishermen's productivity (Syukur et al., 2018). This is because the fishing gear can provide an idea of productivity, economic capability and access to resources. A fisherman who owns a lot of fishing gear tends to have a bigger catch (Parenrengi et al., 2020). In the context of this research, reviewing the effectiveness of Naïve Bayes and K-Nearest Neighbor algorithms in processing data to support government decision-making becomes increasingly relevant. Testing the effectiveness of these two algorithms allows us to evaluate the extent to which these two algorithms can work according to the data owned. Testing the effectiveness of the Naïve Bayes and KNN algorithms in determining the eligibility of fishermen assistance is important because the different characteristics and assumptions of each algorithm can affect the prediction results. Naïve Bayes assumes

How to Cite:
Nasrullah, M., Surarso, B., & Nurhayati, O. D. (2024). Analysis of Naïve Bayes and K-Nearest Neighbors Algorithms for Classifying Fishermen Aid Eligibility. *Jurnal Penelitian Pendidikan IPA*, *10*(10), 7652–7664. https://doi.org/10.29303/jppipa.v10i10.8818

independence between features, while KNN uses distance measurements to determine data classes (Tejawati et al., 2023). Testing the effectiveness of such algorithms helps in comparing their performance in the context of providing fishermen assistance, thus ensuring the selection of the right algorithm to support accurate and reliable decisions.

In this research, Naïve Bayes and KNN algorithms will be used to analyze the eligibility of providing relevant fishermen assistance. Classification is a method of data mining that predicts label variables based on criteria variables (Nasution et al., 2023). Naïve Bayes is a classification algorithm, which is a probabilistic machine learning technique based on Bayes Theorem, which is predicted for independent speculation (Sano et al., 2023). Naïve Bayes uses probability theory to classify data (Shang, 2024). Naïve Bayes helps develop models that provide predictive capabilities, providing a new way to understand data (Libnao et al., 2023).

The second algorithm to be used in this research is the K-Nearest Neighbor algorithm. This algorithm classifies objects by referring to the training data that is closest to the object (Martin et al., 2023). The KNN algorithm is a classification method for a set of data that is based on learning from previously classified data (Lin, 2024). KNN is a Supervised Learning algorithm, where the results of new query instances are classified based on the majority of proximity distances from the existing categories in KNN (Hasdyna & Dinata, 2020).

Several classification methods have been applied in determining aid provision such as Random Forest, Support Vector Machine, Decision Tree, and Neural Network methods have drawbacks. For example, Random Forest tends to require longer computation time due to the construction of many decision trees, while SVM can be sensitive to data scale and requires careful parameter tuning. Decision Tree is prone to overfitting, and Neural Network requires a large amount of data (Kurniadi & Larasati, 2022). In research conducted by Azhari (2021), using a dataset of 200 using classification algorithms such as Random Forest, Decision Tree, SVM and Naïve Bayes, in his research Random Forest and Decisom Tree have the least accuracy because they require longer computing time due to the construction of many decision trees and the low level of accuracy of assessment on small data.

In the context of providing aid to fishermen, most current approaches still rely on manual methods, which are time-consuming and prone to errors. Computational methods, such as classification algorithms, have been applied, but existing studies typically focus on a single algorithm or use computationally intensive algorithms, such as Random Forest and Neural Networks, which are less efficient for small to medium-sized datasets. The effectiveness of Naïve Bayes and KNN algorithms can be an alternative due to their efficient nature in handling small to medium data (Syefudin et al., 2023). They also tend to be more resistant to noise and outliers in the dataset (Yacoub et al., 2024). In the context of eligibility for fishermen assistance with a dataset of 302, the advantages of these two algorithms can work well on small and medium data scales (Kenia et al., 2023). The dataset in this study includes a dataset with an unbalanced distribution of classes in the data, to overcome the unbalanced dataset, overfiting and reduce bias towards majority data, in this case it will use the Synthetic Minority Over-sampling Technique (SMOTE) method. The main function of the SMOTE method is to increase the number of samples from the minority class by generating new synthetic samples, thus achieving balance with the majority class (Hunafa & Hermawan, 2023).

This research offers a novel approach by utilizing both Naïve Bayes and K-Nearest Neighbor (KNN) algorithms, which are more efficient for handling small to medium-sized datasets. These two algorithms are rarely used together in the context of fishermen's aid classification. Additionally, this study employs SMOTE (Synthetic Minority Over-sampling Technique) to address data imbalance issues, a method seldom applied in similar studies within the maritime sector. The combination of Naïve Bayes, KNN, and SMOTE for classifying the eligibility of fishermen for aid is an innovative approach, aimed at improving both accuracy and efficiency in decision-making processes (Danitasari et al., 2024).

This research is important because it can help ensure that the distribution of aid to fishermen is more targeted and efficient. By utilizing the Naïve Bayes and K-Nearest Neighbor (KNN) algorithms, along with the SMOTE technique to address data imbalance, this study supports faster and more accurate decision-making. The results of this research are expected to improve the welfare of fishermen and encourage the use of technology in the fisheries sector.

This research focuses on analyzing the eligibility of providing fishermen assistance using Naïve Bayes and KNN algorithms supported by using the SMOTE method. This research will build a model, evaluate and compare the effectiveness of the two algorithms in classifying the eligibility of providing fishermen assistance.

Method

Literature review

The literature review in this study will involve an extensive exploration of three main areas. First, this review requires an understanding of Naïve Bayes and KNN algorithms, including their working principles, advantages, disadvantages, and variations. Second, in the context of analyzing the eligibility for fishermen's aid, the literature review covers methodologies and frameworks for evaluating aid eligibility, as well as factors influencing its success. Lastly, this review seeks literature discussing Naïve Bayes and KNN algorithms, including case studies and applied research that highlight the strengths and limitations of these algorithms in practical contexts.

Data collection

Data collection for the eligibility assessment of aid to fishermen for this research involves gathering information from residents of Ketapang Raya Village, Keruak District, East Lombok Regency, who work as fishermen. This is achieved through direct visits to the village government for conducting surveys, observations, and interviews. The collected data will ensure accuracy, completeness, and relevance.

The data sources in this research consist of two types. First, primary data is obtained through data collection activities focused on the characteristics and profiles of individual fishermen. Second, secondary data is utilized from previous studies, particularly those that provide insights into the models used to analyze the eligibility for aid for fishermen. This secondary data includes data mining, classification, Naïve Bayes methods, K-Nearest Neighbor, and other relevant literature (Putro et al., 2020).

Preprocessing

In the preprocessing stage, data cleaning will be conducted to remove or correct inaccurate, irrelevant, or incomplete data entries (Lubis et al., 2024). This includes tasks such as eliminating duplicates, correcting typing errors, or normalizing formats. Additionally, unnecessary attributes will be removed to avoid potential impacts on classification analysis results. The data cleaning process will utilize the Replace Missing Values operator in Rapid Miner to ensure clean data output. By cleaning the data, we aim to minimize errors, biases, and anomalies that could affect the final analysis outcomes (Wang et al., 2024).

Data training and testing split

The process of dividing data into training and testing sets is a crucial stage in machine learning model development. The training data will be used to train the classification model, while the testing data will be used to evaluate the performance of the trained model. The data will be split into training and testing sets with a ratio of 80% and 20%. This data splitting process will utilize the Split Data operator available in Rapid Miner. It is important to perform this data splitting process to ensure that the developed model can generalize well. By using independent testing data, we can objectively measure the model's performance and determine whether it has good classification and prediction capabilities or not (Prasetyo et al., 2024).

Building Naïve Bayes and KNN Models

Naïve Bayes.

Naïve Bayes is a classification algorithm that has a simple algorithm structure and high computational efficiency (Chen et al., 2021). This algorithm is a simple form of Bayesian Network, in which all independent attributes are assigned variable class values (Farhana, 2021). Naïve Bayes algorithm has advantages such as simplicity, speed, and high accuracy (Imanuddin et al., 2023). In short, the Naïve Bayes algorithm is a data calcification algorithm that predicts all the probabilities of each class member (Ridwan, 2020). The following equation can be seen in equation (1):

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

Explanation:

X : Data with unknown classes
H : Hypothesis that X is a certain class
P(H|X) : The probability of hypothesis H given condition x (posteriori prob.)
P(H) : The probability of hypothesis H (prior prob.)
P(X|H) : The probability of X given that condition
P(X) : The probability of X

K-Nearest Neighbor (KNN)

The K-Nearest Neighbor algorithm is a method for object classification based on the training data closest to the object (Raysyah et al., 2021). This distance is calculated based on the proximity between the input data and the data in the group, using the values of a number of existing features (Argina, 2020). The distance or measure of dissimilarity can be calculated using the euclidean distance (Rahmahwati & Kirana, 2023). The following equation used is shown in equation 2 (Kurniawan & Barokah, 2020).

$$D(X_1, X_2) = \sqrt{\sum_{i=1}^{n}(X1 - X2)^2} \quad (2)$$

(2)

Explanation:

D : nearest distance
X1 : Sample data or training data
X2 : Test data
n : Number of attributes for each case
I : Individual attributes from 1 to n

Synthetic Minority Over-sampling Technique (SMOTE)

Synthetic Minority Over-sampling Technique (SMOTE) is a method to balance different classes by using oversampling (Bunkhumpornpat et al., 2024). SMOTE doubles the data in the minority class to balance it with the data in the majority class (Sholihah & Hermawan, 2023). Class imbalance occurs when one class has far more instances than another, which can cause machine learning models to be biased towards the majority class (Prianata, 2024). This technique produces synthetic or artificial data based on measuring the closeness of numerical data using euclidean distance, while for categorical data using mode values (Iskandar & Nataliani, 2021).

$$X_{syn} = X_i + (X_{knn} - X_i)\ x\ \delta \tag{3}$$

Model evaluation

This research uses a fold-cross evaluation validation of 10. Furthermore, Confusion Matrix can provide an overview of how well the model can distinguish between positive and negative classes sebenarnya (Normawati & Prayogi, 2021). From Confusion Matrix, we can calculate other matrices such as accuracy, presicion, and recall. The following is the equation used (Liu, et al., 2024)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} x\ 100\% \tag{4}$$

$$Presicion = \frac{TP}{TP+FP} x\ 100\% \tag{5}$$

$$Recall = \frac{TP}{TP+FN} x\ 100\% \tag{6}$$

$$F_{measure} = 2\frac{presisi \times recall}{presisi+recall} x\ 100\% \tag{7}$$

Classification model selection

Classification is the task of assessing data objects to place them into specific classes from a set of available classes (Hunafa & Hermawan, 2023). Classification model selection is a critical stage in machine learning model development aimed at choosing the most suitable and optimal model to address the problem at hand. Based on the evaluation results, the Naïve Bayes and KNN classification models will be compared to determine the most effective model for classifying the eligibility for fishermen assistance. The model that performs better in predicting correct classes and exhibits higher evaluation matrix values will be chosen.

Results and discussion

This section presents the findings and discussion of the effectiveness of Naïve Bayes and KNN algorithms for assessing the eligibility of fishermen for assistance programs, comparing which algorithm is more effective in classification. The analysis involves evaluating the results obtained from previous model evaluations. These results are then discussed in-depth to assess the performance of both classification algorithms in the context of analyzing the eligibility for fishermen assistance. The discussion compares the performance of Naïve Bayes and KNN without SMOTE and with SMOTE, based on the predefined Confusion Matrix metrics such as accuracy, precision, recall, and F1-score. Furthermore, the results and discussion include an analysis of the strengths and weaknesses of each algorithm, as well as the potential application of these algorithms in broader contexts of providing assistance to fishermen. This comprehensive analysis aims to provide insights into how Naïve Bayes and KNN algorithms can be effectively utilized in determining eligibility for fishermen assistance programs, considering both their performance metrics and practical implications.

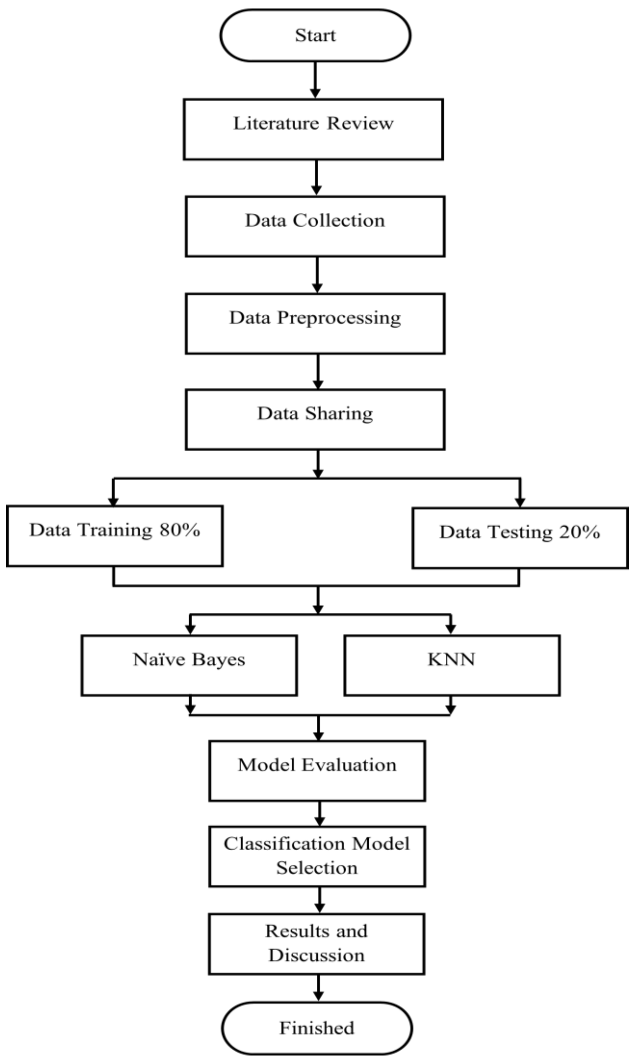

Figure 1. Research Procedures

Result and Discussion

Modeling
Dataset

The dataset utilized comprises fisherman data with attributes such as membership number, name, hamlet, number of boats, number of motorized boats, number of non-motorized boats, number of fishing rods, number of nets, and eligibility status as the label. Each of these attributes provides significant information to determine the eligibility of fishermen for receiving assistance. The dataset used in this study includes 302 entries.

Ro...	NO ANGGOTA	NAMA	DUSUN	KAPAL	SAMPAN M...	SAMPAN	PANCING	JARING	KETERANGAN
1	A1B1C52032...	A SENAN	PELEBE	0	0	1	10	0	LAYAK
2	A1B4C52032...	ABD AZIZ ALI...	LUNGKAK S...	0	0	1	0	0	LAYAK
3	A9B0C52032...	ABD RAHMAN	LUNGKAK TI...	0	1	0	18	0	LAYAK
4	A8B1C52032...	ABDUL HAFIF	KEDOME	0	0	0	12	0	LAYAK
5	A3B0C52032...	ABDUL JABAR	LUNGKAK U...	0	0	1	10	0	LAYAK
6	A2B1C52032...	ABDUL JALA ...	MERINGKIK ...	0	1	0	15	0	LAYAK
7	A4B0C52032...	ABDUL KARIM	LUNGKAK TI...	0	0	0	0	0	LAYAK
8	A1B1C52032...	ABDUL MANAF	PELEBE	0	1	1	30	60	TIDAK LAYAK
9	A1B1C52032...	ABDUL MUJIB	LUNGKAK U...	0	0	1	0	0	LAYAK
10	A4B0C52032...	ABDUL RAH...	LUNGKAK	0	2	1	34	80	TIDAK LAYAK
11	A1B1C52032...	ABDUL RASY...	PELEBE	0	1	1	50	70	TIDAK LAYAK
12	A1B0C52032...	ABDUL SAMAK	LUNGKAK	0	1	1	15	0	LAYAK
13	A1B1C52032...	ABDULLAH	PELEBE	0	0	0	5	0	LAYAK
14	A9B1C52032...	ABDULLAH	LUNGKAK TI...	0	1	1	30	70	TIDAK LAYAK
15	A8B1C52032...	ABDURRAHIM	TELAGA BAG...	0	0	0	0	0	LAYAK

ExampleSet (302 examples,0 special attributes,9 regular attributes)

Figure 2. Dataset to Be Used

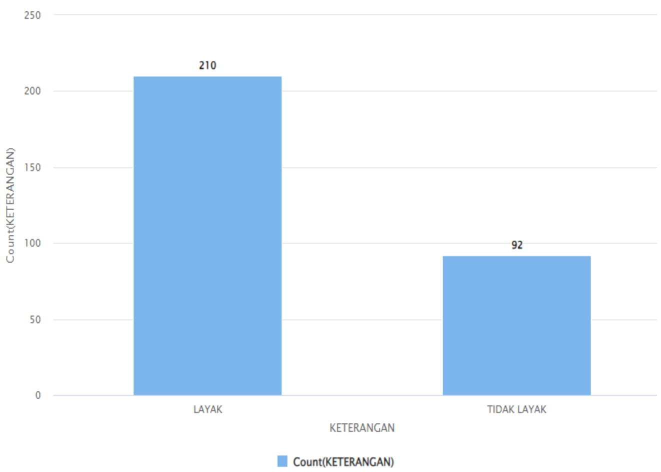

Figure 3. Data distribution based on labels

Based on the data distribution in Figure 3, there are 210 individuals eligible for assistance and 92 individuals ineligible.

Preprocessing

The data preprocessing process, which includes data cleaning, is crucial to ensure the data is in optimal condition before being used for analysis or machine learning model training. This process uses the Replace Missing Value operator in Rapid Miner. During data cleaning, it is essential to ensure the dataset is free from missing values, duplicates, unwanted outliers, and unnecessary attributes such as membership number, name, and hamlet. Removing these attributes is intended to avoid influencing the attributes used for classification. This stage also involves determining or selecting the label, where the chosen label attribute is the eligibility status. By following these steps, we can significantly improve the quality and accuracy of the machine learning model. The data cleaning process is illustrated in Figures 3 and Figure 4.

	DUSUN polynominal	KAPAL polynominal	SAMPAN M... polynominal	SAMPAN polynominal	PANCING polynominal	JARING polynominal
1	DUSUN	Kapal	Sampan Motor	Sampan	Pancing	Jaring
2	PELEBE	0	0	1	10	0
3	LUNGKAK SELA...	0	0	1	0	0
4	LUNGKAK TIMUR	0	1	0	18	0
5	KEDOME	0	0	0	12	0
6	LUNGKAK UTARA	0	0	1	10	0
7	MERINGKIK BAR...	0	1	0	15	0
8	LUNGKAK TIMUR	0	0	0	0	0
9	PELEBE	0	1	1	30	60
10	LUNGKAK UTARA	0	0	1	0	0
11	LUNGKAK	0	2	1	34	80
12	PELEBE	0	1	1	50	70

PREVI...

no problems.

Figure 4. Attribute Removal and Label Selection Process

Figure 4 illustrates the process of removing unnecessary attributes such as membership number, name, and hamlet, aiming to enhance model performance. This process also includes selecting the eligibility status attribute as the label (Peretz et al., 2024).

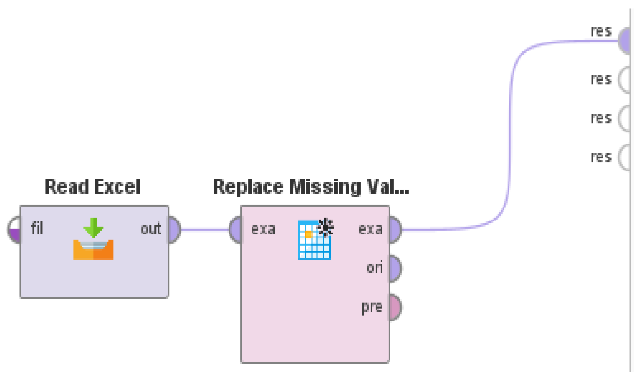

Figure 5. Process Of Inputting The Replace Missing Values Operator

Figure 5 shows the process of inputting the Replace Missing Values operator. This operator functions to detect and identify missing values in the dataset, and it also helps in simplifying the dataset (Tan et al., 2024).

Data split

Splitting the data into training and testing datasets is an essential step in developing a machine learning model to ensure objective evaluation and avoid overfitting. At this stage, the Split Data operator in Rapid Miner is used to divide the data into a training dataset (80%) and a testing dataset (20%).

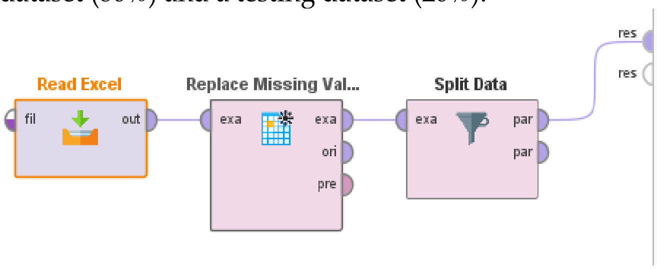

Figure 6. Data Split Process

Figure 6 depicts the process of data splitting using the Split Data operator. The data is divided into 80% for training and 20% for testing to ensure that the model built is not only accurate on the training data but also capable of generalizing and performing well on new data (Ariyanti & Iswardani, 2020).

Implementation of the Naïve Bayes Algorithm without SMOTE

As many as 302 fishermen data are utilized in classification using the Naïve Bayes algorithm without SMOTE and will be tested using 10-fold cross validation in Rapid Miner. Additionally, evaluation will be conducted to obtain accuracy, precision, and recall values for optimal results.

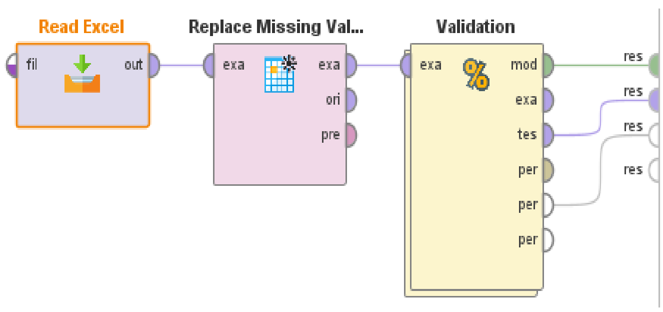

Figure 7. The process of inputting the Cross Validation operator

Figure 7 depicts the input process of the Cross Validation operator, aimed at dividing the dataset into several subsets (folds) (Ariyanti & Iswardani, 2020).

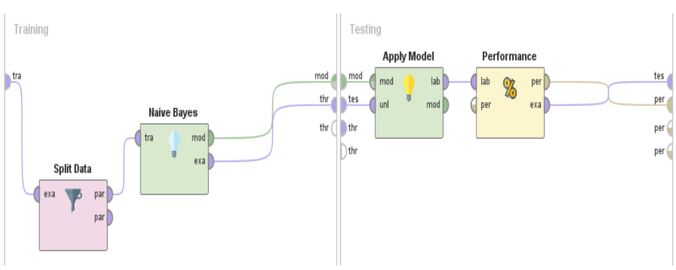

Figure 8. Data Processing Using Naïve Bayes Without SMOTE

In Figure 8, the process illustrates data processing using the Naïve Bayes algorithm. This process involves integrating operators within the Naïve Bayes classification model. Subsequently, the Apply Model operator functions to apply the trained model generated by the Naïve Bayes operator. This operator produces predictions for test data based on the trained model. Finally, the Performance operator is employed to evaluate the model's performance based on the generated predictions (Surampudi & Kumar, 2024). This operator computes the Confusion Matrix metrics such as accuracy, precision, and recall to assess how well the model performs.

Table 1. Data processing using Naïve Bayes without SMOTE

	Folds	Result of naïve bayes algorithm Processing without SMOTE		
		Accuracy	Presicion	Recall
Scenario 1	10	97.01%	94.16%	96.67%

Table 1 shows the results of testing using the K-Fold Cross Validation method with 10 folds on the Naïve Bayes algorithm without SMOTE, indicating an accuracy of 97.01%. This demonstrates that the Naïve Bayes algorithm without SMOTE can classify data effectively.

Implementation of the Naïve Bayes Algorithm with SMOTE

Figures 8 and 9 below explain the process of training data to generate a classification model for determining the eligibility of assistance for fishermen using the Naïve Bayes algorithm along with SMOTE in the Rapid Miner tool (Chachoui et al., 2024).

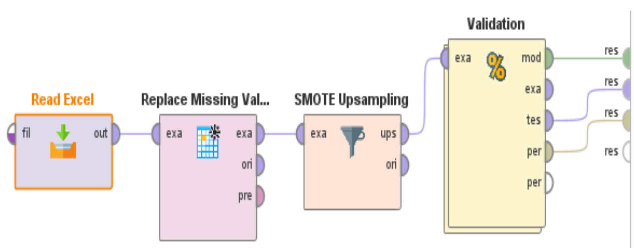

Figure 9. Processing Data using Naïve Bayes with SMOTE

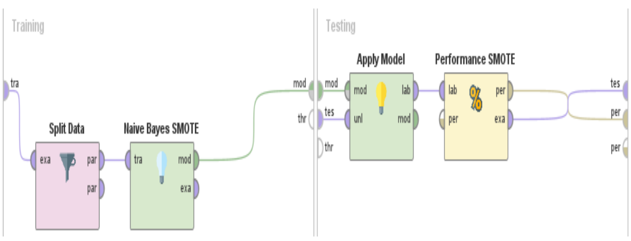

Figure 10. Continuing Data Processing using Naïve Bayes with SMOTE

Table 2. Results of Data Processing Using Naïve Bayes Algorithm with SMOTE

	Upsampling	Neighbor(k)	Folds	Results of naïve bayes algorithm processing without SMOTE		
				Accuracy	Presicion	Recall
Scenario 2	118	1	10	97.86%	96.81%	99.05%
Scenario 3	118	2	10	98.33%	96.86%	100%
Scenario 4	118	3	10	98.33%	96.86%	100%
Scenario 5	118	4	10	97.86%	96.02%	100%
Scenario 6	118	5	10	98.10%	97.27%	99.05%
Scenario 7	118	6	10	98.10%	96.84%	99.52%
Scenario 8	118	7	10	97.62%	96.00%	99.52%
Scenario 9	118	8	10	97.86%	96.02%	100%
Scenario 10	118	9	10	98.10%	96.82%	99.52%
Scenario 11	118	10	10	98.10%	96.84%	99.52%

Table 2 displays the results of data processing using the Naïve Bayes algorithm with SMOTE. The number of synthetic data generated (SMOTE upsampled data) is 118, and the training data is divided into 10 folds. In this process, the value of k varies from k = 1 to k = 10. The best classification models were achieved in scenarios 3 and 4 with k = 2 and k = 3, both achieving an accuracy of 98.33%.

Implementation of the KNN Algorithm Without SMOTE

Next, we will implement the KNN algorithm without SMOTE and test it using 10-fold cross-validation, utilizing Rapid Miner tools. This process aims to determine the accuracy, precision, and recall values of the KNN algorithm. The data will be tested using 10 folds of K-Fold Cross Validation (Hutapea & Silalahi, 2023). Figures 10 and 11 below illustrate the data processing steps with the KNN algorithm.

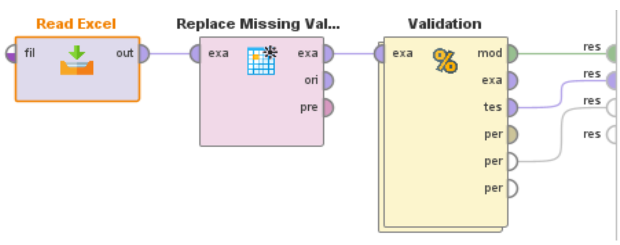

Figure 11. The process of inputting the Cross Validation operator

The purpose of inputting the Cross Validation operator is to divide the dataset into several subsets (folds) (Ariyanti & Iswardani, 2020).

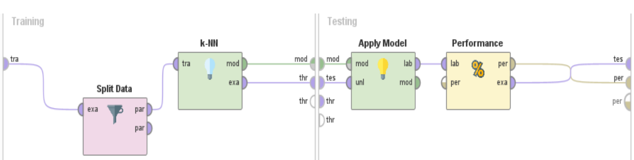

Figure 12. Continuing Data Processing using KNN without SMOTE

In Figure 12, the data processing process with the KNN algorithm without SMOTE is depicted. The process begins with the insertion of the KNN classification model operator, where the model classifies based on the majority class of the k nearest neighbors. Next, the Apply Model operator is introduced to apply the trained model using the KNN operator. This operator generates predictions for test data based on the trained model. Finally, the Performance operator is included to evaluate the model's performance based on the generated predictions.

The results of the testing using the K-Fold Cross Validation method with 10 folds on the KNN algorithm without SMOTE can be seen in Table 3.

Table 3. Results of data processing using KNN without SMOTE

	Folds	Results of processing the KNN algorithm without SMOTE		
		Accuracy	Presicion	Recall
Scenario 12	10	94.04%	94.53%	86.00%

From the test results using K-Fold Cross Validation with 10 folds, shown in Table 3, an accuracy of 94.04% was obtained. This indicates that the KNN algorithm without SMOTE is capable of classifying the data effectively.

Implementation of the KNN Algorithm with SMOTE

Figures 12 and 13 below illustrate the data training process to generate a classification model for determining fishermen's eligibility for aid using the KNN algorithm with SMOTE in Rapid Miner.

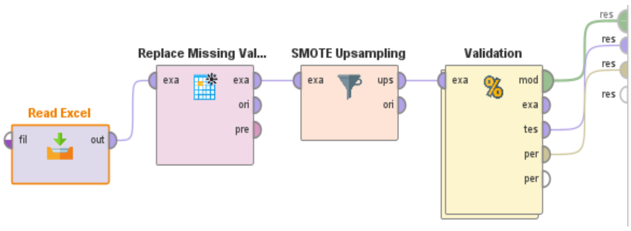

Figure 1. Data processing using KNN with SMOTE

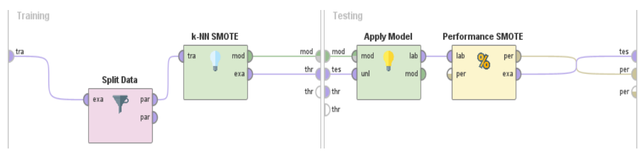

Figure 2. Continuing data processing using KNN with SMOTE

Table 4. Data Processing Results of the KNN Algorithm with SMOTE

	Upsampling	Neighbor (k)	(Folds)	Data processing results of the KNN algorithm with SMOTE		
				Accuracy	Presicion	Recall
Scenario 13	118	1	10	96.67%	97.75%	95.71%
Scenario 14	118	2	10	95.71%	97.72%	93.81%
Scenario 15	118	3	10	96.90%	97.72%	96.19%
Scenario 16	118	4	10	95.95%	97.72%	94.29%
Scenario 16	118	5	10	96.43%	97.72%	95.24%
Scenario 17	118	6	10	95.95%	97.72%	94.29%
Scenario 18	118	7	10	96.67%	97.75%	95.71%
Scenario 19	118	8	10	95.71%	97.70%	93.81%
Scenario 20	118	9	10	96.19%	97.29%	95.24%
Scenario 21	118	10	10	95.71%	97.70%	93.81%

Table 4 above presents the results of data processing using the KNN algorithm with SMOTE. A total of 118 synthetic data points (SMOTE upsampled data) were generated, and the training data was divided into 10 folds. In this process, the value of k used varied from $k = 1$ to $k = 10$. The best classification model was obtained in the scenario 15 with a k value of 2, achieving an accuracy of 96.90%, as shown in Table 4.

Evaluation

The Naïve Bayes Algorithm

One common method for evaluating a classification model, such as Naïve Bayes, is by using a Confusion Matrix and F1-score. By utilizing the Confusion Matrix and F1-score, a more comprehensive picture of the Naïve Bayes algorithm's performance can be obtained. The Confusion Matrix and F1-score for the Naïve Bayes algorithm without SMOTE achieved the best result of 97.07%, as seen in Table 1, while the Naïve Bayes algorithm with SMOTE achieved the best result of 98.33%, as shown in Table 2. Below is the Confusion Matrix for the best accuracy of the Naïve Bayes algorithm without SMOTE.

Table 5. Confusion Matrix of Naïve Byaes Algorithm without SMOTE

Accuracy: 97.01%, Presicion: 94.16%, Recall: 96.67%			
Class	Eligible	Ineligible	Class of presicion
Eligible	204	3	98.55%
Ineligible	6	89	93.68%
Class of recall	97.14%	96.74%	

$$F_{measure} = 2\frac{94.16 \times 96.67}{94.16+96.67} x\ 100\% = 95.39\%$$

From the calculation results using equation (7), the F1-score of the Naïve Bayes algorithm without SMOTE is 95.39%.

Table 6. The Confusion Matrix with the best accuracy from the Naïve Bayes algorithm with SMOTE

Accuracy: 98.33%, Presicion: 96.86%, Recall: 100%			
Class	Eligible	Ineligible	Class of presicion
Eligible	203	0	100.00%
Ineligible	7	210	96.77%
Class of recall	96.67%	100.00%	

$$F_{measure} = 2\frac{96.86 \times 100}{96.86 + 100} x\ 100\% = 98.49\%$$

From the calculation using equation (7), the F1-score of the Naïve Bayes algorithm with SMOTE is 98.49%.

KNN Algorithm

Next, the KNN algorithm will also be evaluated using Confusion Matrix and F1-score. By employing Confusion Matrix and F1-score, a more comprehensive overview of the KNN algorithm's performance can be obtained. Below are the Confusion Matrix and F1-score results for KNN without SMOTE, achieving the best result of 94.04% as seen in Table 3, and KNN with SMOTE achieving the best result of 96.90% as seen in Table 4. Additionally, the Confusion Matrix for the best accuracy of KNN without SMOTE can be found in Table 7 below.

Table 7. Confusion Matrix of KNN Algorithm Without SMOTE

Accuracy: 94.04%, Presicion: 94.53%, Recall: 86.00%			
Class	Eligible	Ineligible	Class of presicion
Eligible	205	13	94.04%
Ineligible	5	79	94.05%
Class of recall	97.62%	85.87%	

$$F_{measure} = 2\frac{94.53 \times 86.00}{94.53 + 86.00} x\ 100\% = 90.06\%$$

From the calculation using equation (7), the F1-score obtained for the KNN algorithm without SMOTE is 90.06%.

Table 8. Confusion Matrix of KNN Algorithm with SMOTE

Accuracy: 96.90%, Presicion: 97.72%, Recall: 96.19%			
Class	Eligible	Ineligible	Class of presicion
Eligible	205	8	96.24%
Ineligible	5	202	97.58%
Class of recall	97.62%	96.19%	

$$F_{measure} = 2\frac{97.72 \times 96.19}{97.72 + 96.19} x\ 100\% = 96.94\%$$

The table above shows the Confusion Matrix with the best accuracy from the KNN algorithm with SMOTE. From the calculation using equation (7), the F1-score obtained for the KNN algorithm with SMOTE is 96.94%.

Comparison Analysis of Naïve Bayes and K-Nearest Neighbor Algorithms with and without SMOT

In classification tasks, selecting the right algorithm is crucial to achieving optimal results. The Naïve Bayes and KNN algorithms have been evaluated using Confusion Matrix and F1-Score. Evaluation is based on the Confusion Matrix metrics such as accuracy, precision, recall, and F1-score, reflecting the highest results of both algorithms. Accuracy measures the proportion of correct predictions, precision indicates the proportion of truly eligible instances among all predicted eligible instances, recall measures the proportion of truly eligible instances successfully identified by the model, and F1-score balances precision and recall. By comparing the results of both algorithms without using SMOTE and with SMOTE, we can understand how applying SMOTE affects model performance, as well as the strengths and weaknesses of each approach in identifying fishermen eligible for assistance. Below is a comparison table of the classification of Naïve Bayes and KNN algorithms without SMOTE and Naïve Bayes and KNN algorithms using SMOTE.

Table 9. Comparison of Naïve Bayes and KNN Algorithms Without SMOTE

Algorithm	Accuracy	Presicion	Recall	F1-score
Naïve bayes	97.01%	94.16%	96.67%	95.39%
K-nearst neighbor	94.04%	94.53%	86.00%	90.06%

Table 9 shows the results of processing data for classifying eligibility for assistance among fishermen, demonstrating different performance between the two algorithms used: Naïve Bayes and KNN without SMOTE. Naïve Bayes achieved an accuracy of 97.01%, with precision of 94.16%, recall of 96.67%, and an F1-score of 95.39%. These results indicate that Naïve Bayes performs well in correctly classifying data, accurately identifying fishermen eligible for assistance, and maintaining a balance between precision and recall. On the other hand, KNN achieved an accuracy of 94.04%, with precision of 94.53%, recall of 86.00%, and an F1-score of 90.06%. Although KNN has slightly higher precision than Naïve Bayes, its lower recall suggests that KNN is less effective in identifying all eligible fishermen for assistance. Overall, Naïve Bayes without SMOTE demonstrates superior performance in classifying eligibility for assistance among fishermen, particularly in terms of recall and F1-score, indicating better ability to detect fishermen truly eligible for assistance.

Meanwhile, the comparison between Naïve Bayes and KNN using SMOTE is shown in the following table 10. The results of this analysis provide a clearer picture of the performance of both algorithms in handling data imbalance.

Table 10. Comparison of Naïve Bayes and KNN Algorithms Without SMOTE

Algorithm	Accuracy	Presicion	Recall	F1-score
Naïve bayes + SMOTE	98.33%	96.86%	100.00%	98.49%
KNN + SMOTE	96.90%	97.72%	96.19%	96.94%

The results of processing classification data for determining eligibility for assistance among fishermen show improved performance when using the Naïve Bayes and KNN algorithms with SMOTE. In the Naïve Bayes algorithm with SMOTE, an accuracy of 98.33% was achieved, with precision of 96.86%, recall of 100.00%, and an F1-score of 98.49%. These results indicate that the Naïve Bayes algorithm with SMOTE performs very well in correctly classifying data, identifying all eligible fishermen without error, and maintaining a very high balance between precision and recall. Meanwhile, the KNN algorithm with SMOTE achieved an accuracy of 96.90%, with precision of 97.72%, recall of 96.19%, and an F1-score of 96.94%. Although the accuracy of KNN with SMOTE is slightly lower compared to Naïve Bayes with SMOTE, its high precision and F1-score suggest that KNN with SMOTE is also very effective in identifying eligible fishermen, albeit slightly less perfect in detecting all eligible fishermen compared to Naïve Bayes. Overall, both models demonstrate excellent performance with the application of SMOTE, but Naïve Bayes with SMOTE excels in terms of accuracy, recall, and F1-score, indicating superior capability.

Tables 9 and 10 present the analysis and processing results of classifying eligibility for assistance among fishermen, highlighting that applying the Naïve Bayes algorithm with SMOTE yields better performance. From this comparison, it is evident that Naïve Bayes combined with SMOTE produces superior results. This underscores that Naïve Bayes with SMOTE is more effective in handling data imbalance and ensuring that no eligible fishermen are overlooked in the classification process. Based on these findings, Naïve Bayes with SMOTE emerges as the preferable choice for classifying eligibility for assistance among fishermen under conditions of unbalanced data.

Conclusion

Based Based on the analysis of the data, there is a significant performance difference between the Naïve Bayes and KNN algorithms in determining the eligibility for assistance among fishermen, especially when using SMOTE. Without SMOTE, Naïve Bayes outperformed with an accuracy of 97.01%, recall of 96.67%, and an F1-score of 95.39%, while KNN achieved an accuracy of 94.04%, recall of 86.00%, and an F1-score of 90.06%. After applying SMOTE, both algorithms showed improved performance, but Naïve Bayes remained superior with an accuracy of 98.33%, recall of 100.00%, and an F1-score of 98.49%. In contrast, KNN with SMOTE reached an accuracy of 96.90%, precision of 97.72%, recall of 96.19%, and an F1-score of 96.94%. This indicates that Naïve Bayes is better at handling data imbalance and accurately identifying fishermen eligible for assistance.

Acknowledgments

Thank you to everyone who has helped and supported the implementation of this research. Without the assistance and guidance from various parties, this research would not have been successfully completed. I hope the results of this research can be beneficial and make a positive contribution.

Author Contribution

Conceptualization, methodology, initial drafting, formal analys, investigation, and visualization, M. N. Wraiting review and editing, validation, supervision, and resources, B. S. and O. D. N.

Funding

This research was funded by the researcher himself.

Conflicts Interest

The authors declare no conflict of interest.

References

Putri, H. A. I. Purnamasari, A. R., Dikananda, O. Nurdiawan, S., Anwar. (2021). Penerima Manfaat Bantuan Non Tunai Kartu Keluarga Sejahtera Menggunakan Metode NAÏVE BAYES dan KNN. *Building of Informatics, Technology and Science (BITS),* 3,(3), 331–337. https://doi.org/10.47065/bits.v3i3.1093.

Fauziah, E., Araafi, A., Mauliyand, S., & Hasibuan, A. (2024). Analisis Potensi Bahaya Lingkungan Kerja Pada Nelayan Tradisional di Wilayah Pesisir. Alahyan Jurnal Pengabdian Masyarakat *Multidisiplin(ECOS-PRENEURS)*, 2(1), 45-51. https://doi.org/10.61492/ecos-preneurs.v2i1.96.

Parenrengi, S., Yunas, S., & Hilmiyah, N. (2020). Sosial Ekonomi Dan Kesejahteraan Nelayan Di Wilayah Teluk Jakarta: Literature Review. *Jurnal Riset Manajemen dan Bisnis (JRMB) Fakultas Ekonomi UNIAT,* 5(1), 93–104. http://jrmb.ejournal-feuniat.net/index.php/JRMB/article/view/274.

Hutajulu, H. (2023). Analysis of Vulnerability and Resilience of Fisherman Households in Facing the Covid-19 Pandemic in Jayapura-Papua City. *Journal of Research in Science Education*, 9(9), 7146–7153. https://doi.org/10.29303/jppipa.v9i9.4618

Syukur, A., Mahrus, M., & AR, S. (2018). Relevansi Budidaya Ramah Lingkungan Terhadap Perlindungan Lamun Skala Lokal Di Pesisir Lombok Timur. *Journal of Research in Science Education*, 5(1). https://doi.org/10.29303/jppipa.v5i1.150

Sano, A. V. D., Stefanus, A. A., Madyatmadja, E. D., Nindito, H., Purnomo, A., & Sianipar, C. P. M. (2023). Proposing a visualized comparative review analysis model on tourism domain using Naïve Bayes classifier. *Procedia Computer Science*, 227, 482–489. https://doi.org/10.1016/j.procs.2023.10.549.

Libnao, M., Misula, M., Andres, C., Mariñas, J., & Fabregas, A. (2023). Traffic incident prediction and classification system using naïve bayes algorithm.

Procedia Computer Science, 227, 316–325. https://doi.org/10.1016/j.procs.2023.10.530.

Martín-Martín, M., Bullejos, M., Cabezas, D., & Alcalá, F. J. (2023). Using python libraries and k-Nearest neighbors algorithms to delineate syn-sedimentary faults in sedimentary porous media. *Marine and Petroleum Geology,* 153. https://doi.org/10.1016/j.marpetgeo.2023.106283.

Hasdyna, N., & Kesuma D. R. (2020). Analisis Matthew Correlation Coefficient pada K-Nearest Neighbor dalam Klasifikasi Ikan Hias. *Informatics Journal*, 5(2), 57-64. https://doi.org/10.19184/isj.v5i2.18907

Azhari, M., Situmorang, Z., & Rosnelly, R. (2021). Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA,* 5(2), 640–651. https://doi.org/10.30865/mib.v5i2.2937.

Azizah, H., Rintyarna, B. S., Cahyanto, T. A. (2022) "Sentimen Analisis Untuk Mengukur Kepercayaan Masyarakat Terhadap Pengadaan Vaksin Covid-19 Berbasis Bernoulli Naive Bayes. BIOS : *Jurnal Teknologi Informasi dan Rekayasa Komputer,* 3(1), 23-29.. https://doi.org/10.37148/bios.v3i1.36.

Merdeka, P. H. (2022). Manajemen Peningkatan Kesejahteraan Masyarakat Pesisir Melalui Pemberdayaan Usaha Lokal Masyarakat. *A Review,* 1(1), 1-9. https://doi.org/10.56855/analysis.v1i1.180.

Wahab, A. (2019). Faktor-Faktor Yang Berhubungan dengan Keluhan Nyeri Punggung Bawah (Low Back Pain) Pada Nelayan di Desa Batu Karas Kecamatan Cijulang Pangandaran. *Biomedika*. 11(1), 25-35. https://doi.org/10.23917/biomedika.v11i1.7599.

Chen, H., Hu, S., Hua, R., & Zhao, X. (2021). Improved naive Bayes classification algorithm for traffic risk management. *Eurasip Journal on Advances in Signal Processing*, no. 1. https://doi.org/10.1186/s13634-021-00742-6.

Farhana, S. (2021). Classification of Academic Performance for University Research Evaluation by Implementing Modified Naive Bayes Algorithm. *Procedia Computer Science,* vol. 194, 224–228. https://doi.org/10.1016/j.procs.2021.10.077.

Imanuddin, S. H., Adi, K., & Gernowo, R. (2023). Sentiment Analysis Naive Bayes Method on SatuSehat Application. Jurnal Penelitian Pendidikan IPA, 9(7), 5524–5531. https://doi.org/10.29303/jppipa.v9i7.4054.

Putro, H. F., Vulandari, R. T. W., & Saptomo, L. Y. (2020). Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan. *Jurnal Teknologi Informasi dan Komunikasi (TIKomSiN),* 8(2), 78-85. https://doi.org/10.30646/tikomsin.v8i2.500.

Ridwan, A. (2022). Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus. *Jurnal Sistem Komputer dan Kecerdasan Buatan,* 7(1), 15-21. https://doi.org/10.47970/siskom-kb.v4i1.169.

Raysyah, S., Arinal, V., & Mulyana, D. I. (2021). Klasifikasi Tingkat Kematangan Buah Kopi Berdasarkan Deteksi Warna Menggunakan Metode KNN dan PCA. *Sistem Informasi*, 8(2), 88–95. https://doi.org/10.30656/jsii.v8i2.3638.

Sahu, P., Singh, B. K., & Nirala, N. (2024). Optimized k-nearest neighbors for classification of prosthetic hand movements using electromyography signal *Engineering Applications of Artificial Intelligence,* no. 133. https://doi.org/10.1016/j.engappai.2024.108390.

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

Rahmahwati, R., & Kirana, E. T. (2023). Implementation of C4.5 and K-Nearest Neighbor to Predict Palm Oil Fruit Production on Local Plantations. *Journal of Research in Science Education*, 9(9), 7454–7461. https://doi.org/10.29303/jppipa.v9i9.4498.

Kurniawan, Y. I., & Barokah, T. I. (2020). Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan K-Nearest Neighbor. *Jurnal Ilmiah MATRIK,* 22(1), 73-82. https://doi.org/10.33557/jurnalmatrik.v22i1.843.

Kurniadi, F. I., & Larasati, P. D. (2022). Light Gradient Boosting Machine untuk Deteksi Penyakit Stroke. *Jurnal Sistem Komputer dan Kecerdasan Buatan,* 6 (1)67-72. https://doi.org/10.47970/siskom-kb.v6i1.328.

Yacoub, M. H., Ismail, S. M., Said, L. A., Madian, A. H., & Ridwan, A. G. (2024). Reconfigurable hardware implementation of Knearest neighbor algorithm on FPGA. *AEU - International Journal of Electronics and Communications, no. 173.* https://doi.org/10.1016/j.aeue.2023.154999.

Kenia, S., Loka, P., & Marsal, A. (2023). Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier Untuk Klasifikasi Status Gizi Pada Balita. *Indonesian Journal of Machine Learning and Computer Science*, 1(3), 8-14. https://doi.org/10.57152/malcom.v3i1.474.

Prianata, W. (2024). Dampak Pengambilan Sampel Data untuk Optimalisasi Data Tidak Seimbang pada Klasifikasi Penipuan Transaksi E-Commerce Wowon Priatna. *Indonesian Journal of Computer Science Attribution*, 13(2), 3070. https://doi.org/10.33022/ijcs.v13i2.3698.

Hunafa, M. R., & Hermawan. A. (2023). Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbor PadaImbalace Class Dataset Penyakit Diabetes. *Media Online*, 4(3), 1551-1561. https://doi.org/10.30865/klik.v4i3.1486.

Ariyanti, D. I. K. (2020). Teks Mining Untuk Klasifikasi Keluhan Masyarakat Menggunakan Algoritma Naive Bayes. *Jurnal IKRA-ITH Informatika,* 4(3), 125-132. http://repository.upm.ac.id/id/eprint/4613.

Shang, Y. (2024). Prevention and detection of DDOS attack in virtual cloud computing environment using Naive Bayes algorithm of machine learning. *Measurement: Sensors*, 31. https://doi.org/10.1016/j.measen.2023.100991.

Nasution, B., Ritonga, W., Siagian, R. C., Pandara, P. D., Alfaris, L., Muhammad, A. C., & Nurahman, A. (2023). Relationship Between BE4DBE2 and Variables n and z: A Comprehensive Analysis Using Linear Regression, Nonparametric Regression, Naive Bayes Classification, Decision Tree Analysis, SVM Analysis, K-Means Clustering, and Bayesian Regression. *Journal of Research in Science Education,*, 9(11), 9532–9546. https://doi.org/10.29303/jppipa.v9i11.4483.

Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI),* 5(2), 697-711. http://dx.doi.org/10.30645/j-sakti.v5i2.369.

Hutapea, M. I., & Silalahi, A. P. (2023). Moderna's Vaccine Using the K-Nearest Neighbor (KNN) Method: An Analysis of Community Sentiment on Twitter. *Journal of Research in Science Education*, 9(5), 3808–3814. https://doi.org/10.29303/jppipa.v9i5.3203.

Lubis, R. D., Iqbal1, M., & Wahyuni, S. (2024). The Influence of Gadget Use and Teacher Creativity Through Motivation on Students Learning Outcomes of Science. *Journal of Research in Science Education,* 10(9), 6287-6297. https://doi.org/10.29303/jppipa.v10i9.8594.

Tan, Y., Sherwood, B., & Shenoy, P. P. (2024). A naïve Bayes regularized logistic regression estimator for low-dimensional classification. *International Journal of Approximate Reasoning*, 172. https://doi.org/10.1016/j.ijar.2024.109239.

Surampudi, S., & Kumar, V. (2024). Hybrid Naïve Bayes Gaussian mixture models and SAR polarimetry based automatic flooded vegetation studies using PALSAR-2 data. *Remote Sensing Applications: Society and Environment,* 36. https://doi.org/10.1016/j.rsase.2024.101361.

Iskandar, J. W., & Nataliani, Y. (2021). Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi),* 5(6), 1120–1126. https://doi.org/10.29207/resti.v5i6.3588.

Bunkhumpornpat, C., Boonchieng, E., Chouvatut, V., & Lipsky, D. (2024). FLEX-SMOTE: Synthetic over-sampling technique that flexibly adjusts to different minority class distributions. *Patterns,* 101073. https://doi.org/10.1016/j.patter.2024.101073

Sholihah, N. N., & Hermawan, A. (2023). Implementation Of Random Forest And Smote Methods For Economic Status Classification In Cirebon City. *Jurnal Teknik Informatika (Jutif),* 4(6), 1387–1397. https://doi.org/10.52436/1.jutif.2023.4.6.1135.

Biyantoro, A. & Prasetyo, B. (2024). Application of Decision Tree for Health Status Classification, Compared to KNN and Naive Bayes. *Indonesian Journal of Informatic Research and Software Engineering*, 4(1). 47–55. https://journal.irpi.or.id/index.php/ijirse.

Nugroho, A., & Religia, Y. (2021). Analisis Optimasi Algoritma Klasifikasi Naive Bayes menggunakan Genetic Algorithm dan Bagging. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi),* 5(3)504–510. https://doi.org/10.29207/resti.v5i3.3067.

Liu, Y., Fan, J., Qi, X., Shen, B., Zhang, R., & Yao, K. (2024). Adaptive ferroelectric states in KNN-based piezoceramics: Unveiling the mechanism of enhancing piezoelectric properties through multiple phase boundary engineering. *Nano Energy,* 128. https://doi.org/10.1016/j.nanoen.2024.109972.

Danitasari, F., Ryan, M., Handoko, D., & Pramuwardani, I. (2024). Improving Accuracy of Daily Weather Forecast Model at Soekarno-Hatta Airport Using BILSTM with SMOTE and ADASYN. *Journal of Research in Science Education,* 10(1), 179–193. https://doi.org/10.29303/jppipa.v10i1.5906.

Chachoui, Y., Azizi, N., Hotte, R., & Bensebaa, T. (2024). Enhancing algorithmic assessment in education: Equi-fused-data-based SMOTE for balanced learning. Computers and Education: *Artificial Intelligence,* 6. https://doi.org/10.1016/j.caeai.2024.100222.

Prasetiyo, W. T., Farikhin, & Sugiharto, A. (2024). Comparative Analysis of User Satisfaction of End User Computing Satisfaction, DeLone & McLean and Webqual 4.0 Methods. *Journal of Research in Science Education,* 10(9), 6826–6834. https://doi.org/10.29303/jppipa.v10i9.8484.

Wang, H., Hu, Z., Guo, W., Zhu, H., Xing, Z., Wang, H., & Cai, Z. (2024). Effect of A and B-site ion doping on the structure and properties of KNN-based ceramic coatings. *Ceramics International,* 50(20), 37809–37819. https://doi.org/10.1016/j.ceramint.2024.07.145.

Peretz, O., Koren, M., & Koren, O. (2024). Naive Bayes classifier – An ensemble procedure for recall and precision enrichment. *Engineering Applications of Artificial Intelligence,* 136. https://doi.org/10.1016/j.engappai.2024.108972.

Syefudin, S., Hendry, H., & Iriani, A. (2023). Analysis of Student Satisfaction with the Quality of Education Services and Lecturer Performance Using the Survey and Naive Bayes Methods. Journal of Research in Science Education, 9(11), 9423–9430. https://doi.org/10.29303/jppipa.v9i11.5367.