BAREKENG: Journal of Mathematics and Its Applications
December 2024 Volume 18 Issue 4 Page 2153–2164
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss4pp2153-2164

COMPARISON OF NAÏVE BAYES AND K-NEAREST NEIGHBOR MODELS FOR IDENTIFYING THE HIGHEST PREVALENCE OF STUNTING CASES IN EAST JAVA

Teguh Herlambang[1]*, Vaizal Asy'ari[2], Ragil Puji Rahayu[3], Aji Akbar Firdaus[4], Nyoman Juniarta[5]

[1,3]*Information System Department, FEBTD, Universitas Nahdlatul Ulama Surabaya*
Jl. Raya Jemursari 51-57, Surabaya, 60237, Indonesia

[2]*Postgraduate of Department of Information Technology,*
Universitas Pembangunan Nasional "Veteran" Jawa Timur
Jl. Rungkut Madya No. 1, Gunung Anyar, Surabaya, 60294, Indonesia

[4] *Department of Engineering, Faculty of Vocational, Universitas Airlangga*
Jl. Dharmawangsa Dalam Selatan 28-30, Surabaya 60286, Indonesia

[5]*Université de Lorraine, Nancy, France*
2 rue Jean Lamour, 54519 Vandoeuvre les

*Corresponding author's e-mail: ** ***teguh@unusa.ac.id***

ABSTRACT

Article History:

Received: 28th December 2023
Revised: 9th May 2024
Accepted: 5th July 2024
Published: 14th October 2024

Keywords:

K-nearest Neighbor;
Naive Bayes;
Stunting

Indonesia will experience a demographic bonus in 2030, where the productive age group will dominate the population and become a buffer for the economy. However, this potential is in vain if human resources experience stunting. According to WHO (2015), stunting is a disorder of child growth and development due to chronic malnutrition and repeated infections, characterized by below-standard length or height. Based on the background of the problem, the author wants to compare the prediction of the prevalence of the highest stunting cases in East Java using the Naive Bayes method and the K-Nearest Neighbor method. The stages carried out in this study are data collection, initial data processing, advanced data processing using the Naïve Bayes Method and K-Nearest Neighbor, and comparative analysis. The results of the implementation of the Naïve Bayes and K-Nearest Neighbor methods are in the form of stunting prevalence prediction charts with variables that affect LBW and TTD. The results of simulations conducted in 6 regions, the Naive Bayes method gets the highest accuracy value of 83.33% in simulation one and 66.67%. The smallest RMSE value is 0.382 simulation 1 and 0.469 simulation 2. This shows that the Naive Bayes method can predict well.

How to cite this article:

T. Herlambang, V. Asy'ari, R. P. Rahayu, A. A. Firdaus and N. Juniarta., " COMPARISON OF NAÏVE BAYES AND K-NEAREST NEIGHBOR MODELS FOR IDENTIFYING THE HIGHEST PREVALENCE OF STUNTING CASES IN EAST JAVA," *BAREKENG: J. Math. & App.*, vol. 18, iss. 4, pp. 2153-2164, December, 2024.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article • Open Access

1. INTRODUCTION

Indonesia will experience a demographic bonus in 2030, where the productive age force will dominate the population and become an economic buffer. The demographic bonus that Indonesia will be consisted of are the productive age force (15-64 years) which is predicted to reach 68 percent of the total population, and the old force (65 years and over) around 9 percent. However, that potential will becomes useless if human resources are stunted [1]. According to WHO (2015), stunting is a disorder of growth and development of children due to chronic malnutrition and repeated infections, which is characterized by their length or height being below the standard. Meanwhile, in 2020, WHO states that stunting is divided into short or very short based on length or height for age (less than -2 standard deviations on the WHO growth curve) which is irreversible due to inadequate nutritional intake and/or repeated or chronic infections that occur in the first 1000 days of life. [2]. Based on WHO data, a country is said to experience stunting problems if the number of cases is above 20% [3]. According to the 2022 Indonesian Nutritional Status Survey (SSGI) Pocket Book, the prevalence of stunting in Indonesia decreased to 21.6% from 24.4% in 2021. In East Java province, stunting prevalence has also decreased from 23.5% in 2021 to 19.2% in 2022. However, the Deputy Governor of East Java, Emil Elestianto Dardak, said that he targeted a reduction in stunting prevalence in East Java to 13.5% by 2024. This target is lower than the central government's target of 14% in accordance with the 2020-2024 National Medium-Term Development Plan (RPJMN) [4].

In previous studies, several forecasting methods were used for estimation of closed hotels and restaurants in Jakarta as impact of corona virus disease spread using adaptive neuro fuzzy inference system [5], neural network algorithm for breast cancer diagnosis [6], estimation of closed hotels and restaurants as impact of Covid-19 spread using backpropagation neural network [7], forecasting of occupied rooms in the hotel using linear support vector machine [8], forecasting the number of dengue hemorrhagic fever patients using the fuzzy logic [9], analysis of demand and supply of blood using panel data regression [10], profitability estimation using H-Infinity and Ensemble Kalman Filter (EnKF) [11], stock price estimation using Ensemble Kalman Filter Square Root (EnKF-SR) [12], prediction of student graduation on time using decision tree [13], electronic nose for classifying civet coffee using Support Vector Machine (SVM), k-nearest neighbors (k-NN), and decision tree [14]. The Naïve Bayes method is a simple probabilistic classification that calculates a set of probabilities by summing the frequencies and combinations of values from a given dataset. While the K-Nearest Neighbor method is an algorithm for classifying data into several classes that have been grouped based on the closest distance or have similarities in the data with training data [15]. Previously, there were several studies that made comparisons between the two methods. Comparative analysis of Naïve Bayes, K-Nearest Neighbor and C.45 method in weather forecast [16] with KNN algorithm results has a higher level of accuracy compared to Naïve Bayes and C4.5 in the case of weather prediction with $k = 7$ and fold $= 5$. Classification of nutritional status of toddlers using Naïve Bayes and K-Nearest Neighbor [17] by generating a website-based application for Classification of Nutritional Status of Toddlers can speed up determining decisions for nutritional status of toddlers. The comparison of accuracy, recall, and precision KNN obtained an accuracy value of 91.79%, recall 91.79%, precision 91.17% while for Naïve Bayes obtained an accuracy value of 80.60%, recall 80.60%, precision 79.66%. Comparative analysis of decision tree, KNN, and Naïve Bayes algorithms for prediction of start-up success [18] by producing a comparison between the three algorithms for classifying 923 start-up data, shows that the Decision Tree algorithm is the most suitable algorithm to be used among KNN and Naïve Bayes algorithms. The accuracy of Decision Tree is 79.29%, while the KNN algorithm with 66.69%, and Naïve Bayes with 64.21%. Furthermore, for the precision value, Decision Tree is still superior with a value of 78.99%, followed by the KNN algorithm with 55.13%, and Naïve Bayes with 51.32%. From the recall performance results, it turns out that the Naïve Bayes algorithm showed the best results with 79.16%, while the Decision Tree 56.27% and KNN with 40.14%. Forecasting average room rate (ARR) using k-nearest neighbor (k-NN) [19] obtained the best RMSE of 6,335 can be used Hotel S in determining the price of various type of rooms that are appropriate and expected to provide benefits for management.

Based on the background of the problem and previous research that has been described, this study compares prevalence predictions regarding stunting cases using 2 methods, namely the Naïve Bayes method and the K-Nearest Neighbor method. This research is very important to be carried out, to produce predictions of the development of the prevalence of stunting cases in the future, which can be used as reference material for the East Java Provincial Health Office to realize the government's mission in reducing the prevalence of stunting cases and find out which methods can be used to make the best predictions with the highest level of accuracy.

2. RESEARCH METHODS

2.1 Dataset

The dataset used in the study was obtained from the East Java Health Office. The dataset is stunting data consisting of several variables, namely, the prevalence of babies with Low Birth Weight (BBLR), the prevalence of mothers who receive Blood Supplement Tablets (TTD) during pregnancy, stunting rates, and stunting status compared to the previous month. The dataset used from January 2019 to December 2022 is shown in **Table 1**.

Table 1. Stunting Dataset in East Java (period 2019 to 2022)

	2019				2020			
Month	**TTD**	**BBLR**	**Stunting**	**Status**	**TTD**	**BBLR**	**Stunting**	**Status**
January	84,5	4,6	14,2	Decreased	90,3	3,5	11,1	Decreased
February	87,2	4,3	13	Decreased	92,6	3,5	10,7	Decreased
March	90,6	4,2	12,6	Decreased	87	3,7	12	Increased
April	93,5	3,7	11,8	Decreased	93,8	3,3	10,3	Decreased
May	96,3	3,5	11,5	Decreased	85,6	4,1	13,6	Increased
June	101,1	3	10,3	Decreased	89,7	3,6	11,6	Decreased
July	94,5	3,6	11,7	Increased	91,5	3,5	10,8	Decreased
August	91,2	4	12,4	Increased	90,8	3,5	10,9	Increased
September	97	3,4	11,4	Decreased	84,9	4,2	15,8	Increased
October	85	4,4	13,5	Increased	85,7	4	13,3	Decreased
November	89,6	4,3	12,8	Decreased	86,5	3,8	12,3	Decreased
December	93,5	3,8	12,4	Decreased	88,4	3,7	11,6	Decreased
	2021				2022			
Month	**TTD**	**BBLR**	**Month**	**TTD**	**BBLR**	**Month**	**TTD**	**BBLR**
January	92,7	3,4	8,5	Decreased	87,7	4,3	7,9	Decreased
February	89	3,8	9,2	Increased	87,5	4,4	8,5	Increased
March	93,3	3,3	8,1	Decreased	87,9	4,3	7,4	Decreased
April	90,7	3,5	9	Increased	86,8	4,7	8,5	Increased
May	90,4	3,5	9,1	Increased	87,6	4,3	8	Decreased
June	91,5	3,5	8,6	Decreased	88,9	4	7,4	Decreased
July	85,7	4,1	10	Increased	89,3	4	7,4	Decreased
August	85,5	4,4	10,4	Increased	90,8	3,8	6,8	Decreased
September	86,3	4	9,8	Decreased	91,4	3,6	6,8	Decreased
October	90	3,7	9,1	Decreased	89,7	3,9	7,3	Increased
November	85,2	4,6	11,6	Increased	89,9	4	7,1	Decreased
December	86,5	3,8	9,4	Decreased	90,3	3,9	6,9	Decreased

2.2 Naïve Bayes Method

Naive Bayes is a simple probabilistic classification that calculates a set of probabilities by summing the frequency and combination of values from a given dataset. The Naive Bayes algorithm, introduced by the British scientist Thomas Bayes, predicts future probabilities based on past experiences and is therefore associated with Bayes' Theorem. The distinctive feature of the Naive Bayes Classifier is its strong (naive) assumption of independence between each condition or event. A conditional probability is an occurrence probability calculation, h, when another event, x, has occurred, which is noted as $P(h|x)$, which incorporates both h and b probabilities **[20]**. The algorithm uses Bayes' theorem and assumes all independent or non-interdependent attributes given by values on class variables **[21]**. Bayes' theorem is expressed mathematically in **Equation (1)**.

$$P(h|x) = \frac{P(x|h).\ P(h)}{P(h)} \tag{1}$$

Where is the conditional probability: $P(x|h)$, Represents the probability of occurrence x if known h and:

$$P(x) = \sum_{i=1}^{n} P(x\ |h_i).P(h_i) \tag{2}$$

where
$P(h|x)$: the probability h occurs with proof that x has occurred (superior probability)
$P(x|h)$: the probability of x occurring with evidence that h has occurred
$P(h)$: probability of h occurrence
$P(x)$: probability of occurrence x

This model is very simple and sophisticated classification method, and it performed well even in complicated scenarios **[22]**. The advantages of using the Naïve Bayes algorithm **[23]** are that it is simple and easy to implement, does not require a lot of training data, handles continuous and discrete data, is highly scalable with the number of predictors and data points, is fast and can be used to make realtime predictions, and is not sensitive to irrelevant features.

2.3 K-nearest Neighbor Method

K-Nearest Neighbor is an algorithm for classifying data into several classes that have been grouped based on the closest distance or have similarities between the data and training data. KNN performs classification based on the nearest, far or nearest distance calculated from Euclidean **[24]**. Euclidean Distance is a calculation used to find the distance between 2 points in Euclidean space. The calculation of Euclidean distances is shown in **Equation (3)**.

$$Euc = \sqrt{(x_2 - x_1)^2 + \left(y_2 - y_1\right)^2} \tag{3}$$

$d\left(x_1, x_j\right)$ = euclidean distance
x_i = record data to i
x_j = record data to j
a_r = r-th virgin with i, j

The advantages of using the k-nearest neighbor algorithm **[25]** are that it is easy to implement, adaptable, has little hyperparameter. The k-NN algorithm operates by identifying the closest distance between the data point being evaluated and the nearest k neighbors in the training data. The k-NN process involves the following steps **[26]**, **[27]**, **[28]**:

1. Define the parameter k (number of nearest neighbors).
2. Compute the Euclidean distance between each object and the sample data.
3. Sort the objects into groups based on the smallest Euclidean distance.
4. Collect the classification category of the nearest neighbors.
5. Predict the value of the query instance using the most common category among the nearest neighbors.

2.4 Comparative Analysis

At this stage, a comparative analysis of the results obtained from the two methods used is carried out. The results that appear are in the form of accuracy percentage, where if the accuracy percentage in one of the methods gets the highest level of accuracy, then it can be known that the method is better at predicting case prevalence. Accuracy is obtained from the amount of data that has been grouped according to the group (TP) plus the number of true negative data (TN) divided by the sum of all data. To calculate the accuracy, **Equation (4)** is used **[24]**.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$TP =$ the number of positive data that has been accurately grouped.
$FP =$ the number of negative data that has been grouped as positive data.
$TN =$ the number of negative data that has been grouped is accurate.
$FN =$ the number of positive data that has been grouped as negative data.

In addition, it is also to find out which method is better from the Root Mean Square Error (RMSE) value. RMSE is used to measure the error of a method. The closer the RMSE value is to 0, the more accurate the predicted results are. The RMSE calculation can be seen in **Equation (5)**.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - F_i)^2}{n}} \tag{5}$$

n = Lots of data from predictions
X_i = Forecast data is calculated from the model used at the time/year i
F_i = Actual data

2.5 Stages of implementation

The study was conducted with several stages in conducting comparative research on prevalence predictions in stunting cases as shown in **Figure 1**.

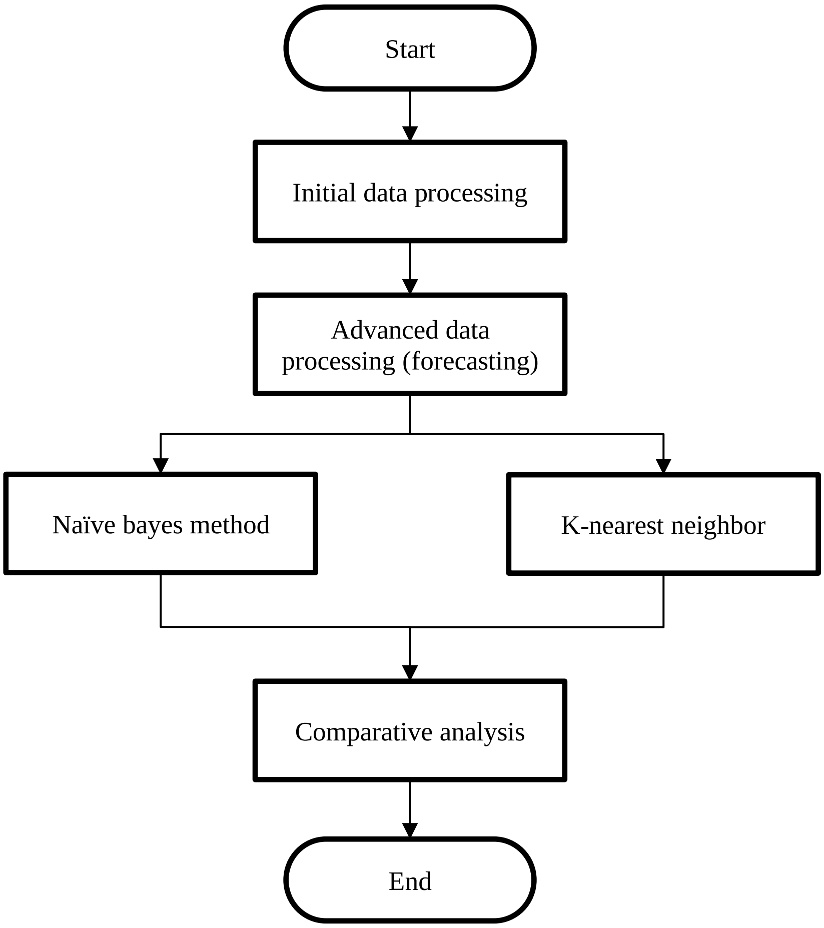

Figure 1. Research methodology

The research will be carried out according to the implementation stage plan that has been made. At the data collection stage, researchers collected data by visiting the East Java Provincial Health Office, the data used will be in the form of data on the prevalence of stunting cases 2018-2022 in each Regency / City, data on the prevalence of babies with Low Birth Weight (LBW) in 2018-2022 in each Regency / City, data on the prevalence of mothers who received Blood Added Tablets (BAT) during pregnancy in 2018-2022 in each Regency / City. The next stage is the initial data processing by cleaning the data, namely by deleting other data that is not needed, adding status, and organizing data by sorting or placing data based on the format that has been prepared. Furthermore, carry out advanced data processing, namely forecasting using the Naïve Bayes method and the KNN method using the RapidMiner tool. Advanced data processing consists of several stages including testing data, training data, related method, apply models, and performance with the flow shown in **Figure 2**.

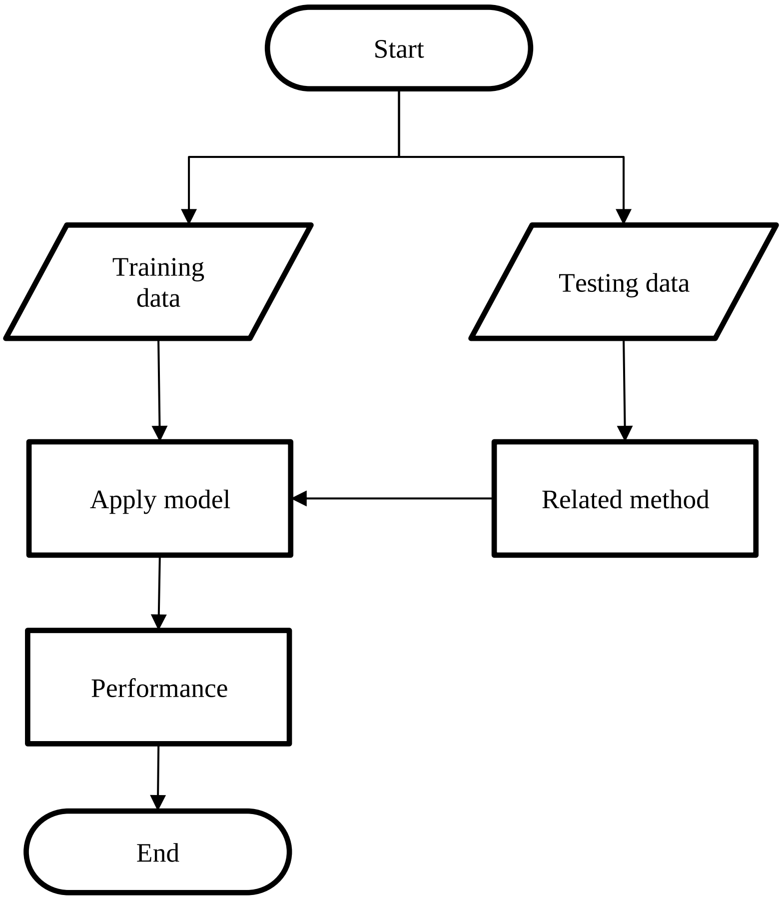

Figure 2. Advanced data processing flow

In the last stage, a comparative analysis of the results obtained from the two methods used was carried out. The results that appear in the form of accuracy percentages, where if the percentage of accuracy in one of the methods gets the highest level of accuracy, it can be known that the method is better at predicting the prevalence of cases.

3. RESULTS AND DISCUSSION

3.1 Data collection

The dataset used in this study was obtained from the East Java Health Office. The data spanning from 2019 to 2022 in East Java. It includes several variables, such as the prevalence of stunting cases, the prevalence of low birth weight (LBW) babies, and the prevalence of mothers receiving iron tablets during pregnancy. The descriptive statistics for this dataset are presented in **Table 2**.

Table 2. Descriptive Statistics of Stunting in East Java

	TTD	BBLR	Stunting
Valid	48	48	48
Missing	0	0	0
Median	89.700	3.800	10.350
Mean	89.696	3.875	10.300
Std. Deviation	3.499	0.393	2.257
Minimum	84.500	3.000	6.800
Maximum	101.100	4.700	15.800

3.2 Initial data processing

At this stage, data deletion was carried out in 2018, because in that year the stunting variable was incomplete and data was not available on the BBLR and TTD variables. The processed data was only taken from East Java Province and 5 regencies/cities, namely Pasuruan City, Probolinggo Regency, Batu City, Ponorogo Regency, and Pacitan Regency. This is done because the five regencies/cities are areas with the highest prevalence in East Java based on input data from the Puskesmas and get special attention from the Health Office. Next add status to the data manually through microsoft excel. After that, organize the data according to the format that has been prepared, by placing data per region and separating between Training data and Testing data. Separation of 75% Training data and 25% Testing data with 87.5% Training data and 12.5% Testing data was carried out. This is done to find out which model is the best.

3.3 Advanced data processing (forecasting)

At this stage, forecasting is carried out using 2 methods, namely the Naïve Bayes method and K - Nearest Neighbor with the help of the RapidMiner tool. Researchers distinguish simulations based on the percentage of training and testing data, where the first simulation was carried out using 75% Training data and 25% Testing data, then the second simulation was carried out using 87.5% Training data and 12.5% Testing data. In the K-Nearest Neighbor method, simulations were carried out using K values of 3, 5, and 7.

Forecasting using the Naïve Bayes and K-Nearest Neighbor methods, calculations were carried out per region in East Java Province, Pasuruan City, Probolinggo Regency, Batu City, Ponorogo Regency, and Pacitan Regency. Data simulation is carried out as in **Figure 3** where the data is connected to the "Split Data" operator to separate the Training data and Testing data, after which it is connected to the Naïve Bayes algorithm or K-Nearest Neighbor, then connected to the "Apply Model" operator.

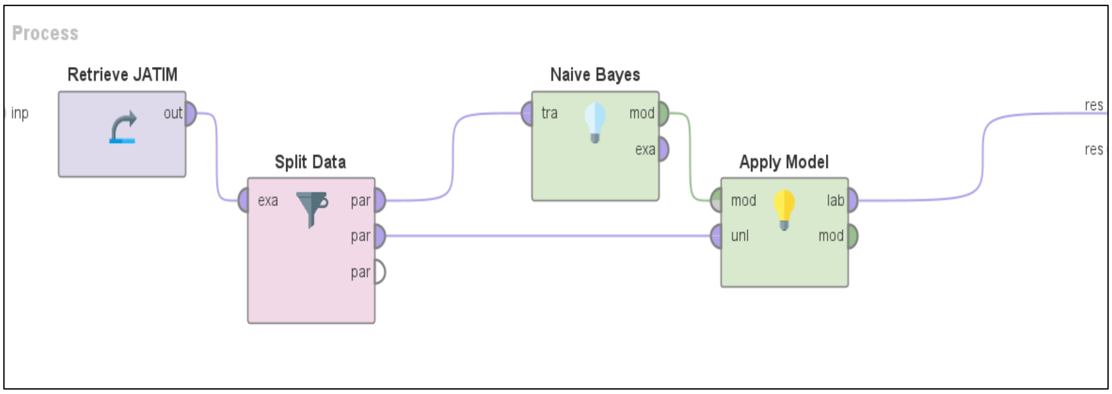

Figure 3. Data Simulation

3.3.1 Model Implementation

In this research, two predictive models were implemented: Naïve Bayes and K-Nearest Neighbor (KNN). These models were used to forecast the highest prevalence of stunting cases based on the provided dataset. The implementation process involved several stages, including data preprocessing, splitting the data into training and testing sets, applying the Naïve Bayes and KNN models, and evaluating their performance. The detailed flow of the advanced data processing, including the model application, is illustrated in **Figure 4** for Naïve Bayes model and **Figure 5** for KNN model. **Figure 4** the implementation of the Naïve Bayes model using RapidMiner for predicting stunting prevalence in East Java. The process begins with the Retrieve JATIM step, where the dataset from East Java is loaded for further processing. The data is then split into training and testing sets in the Split Data stage, which is crucial for validating the model's performance. Next, the Naïve Bayes model is trained using the training dataset. In this step, the Laplace correction parameter is enabled to handle potential zero-probability issues, ensuring more reliable probability estimates. Finally, in the Apply Model stage, the trained Naïve Bayes model is applied to the testing data to generate predictions. The outcomes of this process will be evaluated using accuracy and RMSE metrics to assess the effectiveness of the Naïve Bayes model in predicting the prevalence of stunting in East Java.

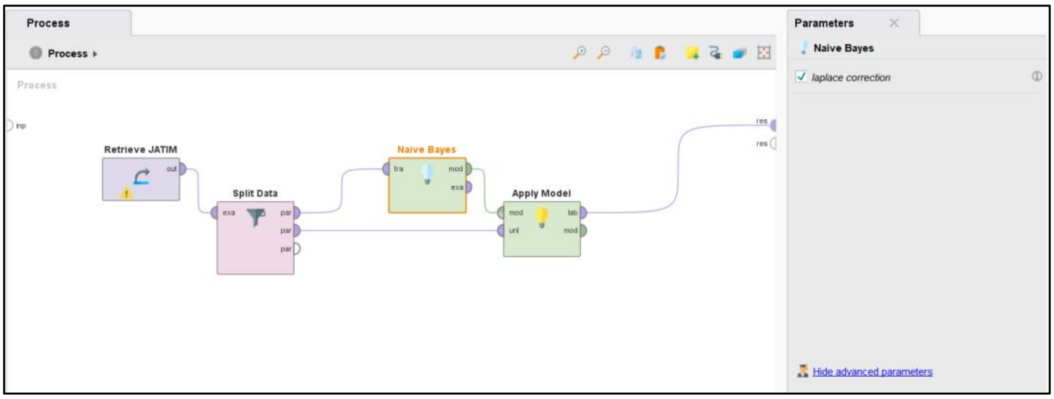

Figure 4. Naïve Bayes model

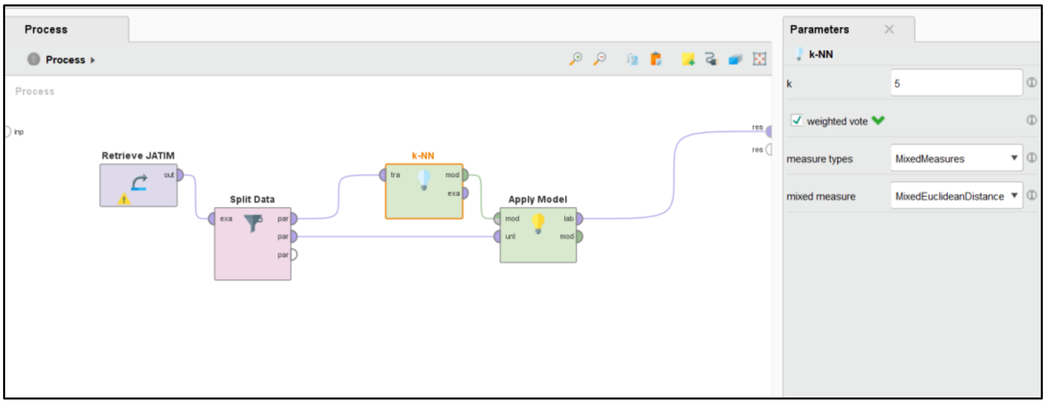

Figure 5. KNN Model

The image illustrates the implementation of the KNN model using RapidMiner for predicting stunting prevalence in East Java. The process begins with loading the dataset from East Java in the Retrieve JATIM step. The dataset is then split into training and testing sets during the Split Data stage, which is crucial for accurately evaluating the model's performance. In the KNN stage, the KNN model is trained using the training dataset with k set to 5 (in this research k values set 3, 5, and 7), meaning the model will consider the 5 nearest neighbors for making predictions. The weighted vote option is enabled, allowing the model to weigh neighbors differently based on their distance. The model employs a "MixedMeasures" type with "MixedEuclideanDistance" to handle various data types within the dataset. Finally, the trained model is applied to the testing data in the Apply Model step to generate predictions. The results from this process will be used to assess the KNN model's effectiveness in predicting the prevalence of stunting in East Java.

3.4 Comparative analysis

At this stage, a comparative analysis was carried out between the Naïve Bayes method and K - Nearest Neighbor based on the Accuracy value and the Root Mean Square Error (RMSE) value. The Accuracy Value is obtained from the calculation of the amount of data that has been in accordance with the group (TP) plus the number of correct negative data (TN) divided by the number of all data. While the RMSE value is generated from squaring errors (predictions - observations) divided by the amount of data (= average), then rooted. In **Figure 6** that the process in using Rapidminer by adding Performance Operators which serves to evaluate the performance of the model. So that the Accuracy and RMSE values per region are obtained by testing 2 times.

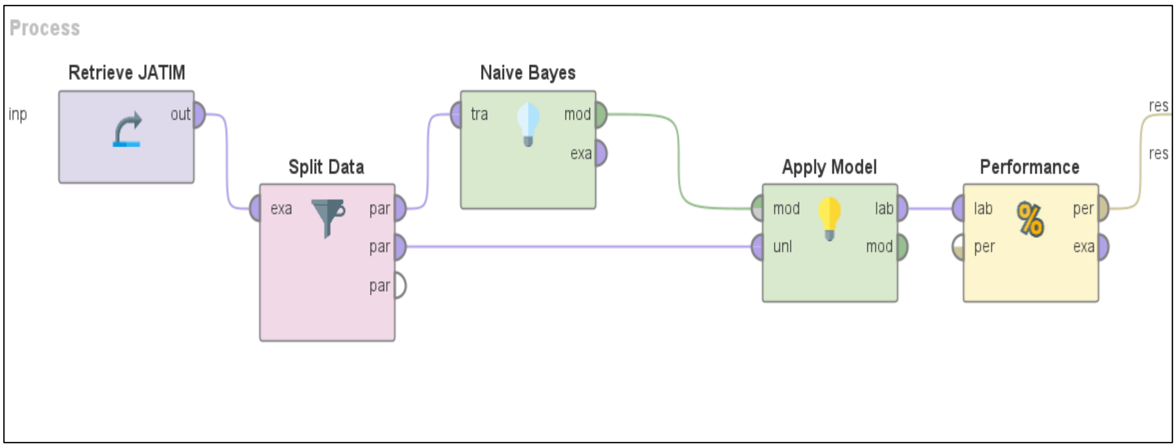

Figure 6. Algorithm Testing Process

Accuracy is used to determine the degree of similarity between the measurement results and the actual value measured. Accuracy measurement is set to find out how much the level of measurement error occurs in a measuring instrument. The accuracy results can be seen in **Table 3**.

Table 3. Comparison of the Accuracy of the East Java region

Algorithm	Simulation 1	Simulation 2
Naïve Bayes	83.33%	66.67%
K-Nearest Neighbor with K = 3	50.00%	33.33%
K-Nearest Neighbor with K = 5	58.33%	33.33%
K-Nearest Neighbor with K = 7	58.33%	33.33%

The accuracy value of forecasting results resulting from data processing simulation 1 in East Java province using the Naïve Bayes method has an optimal accuracy value of 83.33%. The accuracy value of forecasting results resulting from data processing simulation 2 in East Java province using the Naïve Bayes method has an optimal accuracy value of: 66.67%.

Table 4. East Java RMSE comparison

Algorithm	Simulation 1	Simulation 2
Naïve Bayes	0.382	0.469
K-Nearest Neighbor with K = 3	0.547	0.698
K-Nearest Neighbor with K = 5	0.504	0.653
K-Nearest Neighbor with K = 7	0.51	0.654

In **Table 4**, the RMSE results are presented. RMSE (Root Mean Square Error) is used to measure the error rate of the prediction results; a smaller RMSE value (closer to 0) indicates more accurate predictions.

The results demonstrate that the Naïve Bayes algorithm outperforms the K-NN method in both accuracy and RMSE, making it a more reliable model for predicting health indicators in East Java. These findings align with previous research, which often highlights Naïve Bayes robustness in handling categorical data and its simplicity in computation. In comparison, the K-NN method showed lower accuracy and higher RMSE values, possibly due to its sensitivity to the choice of K value and the distribution of the data. This sensitivity makes K-NN less reliable for the given dataset, where the high variability of health indicators requires a model that can generalize well from training data.

Applying these findings to real cases, the Naïve Bayes model can be effectively used for public health planning and intervention in regions with similar demographic and health profiles. Its ability to provide accurate predictions can assist health authorities in identifying areas at higher risk and allocating resources more efficiently. The comparative analysis confirms the superior performance of the Naïve Bayes method over the K-NN algorithm in predicting health-related outcomes in East Java. This study's findings contribute to the body of knowledge on predictive modeling in public health and offer practical implications for enhancing health surveillance and intervention strategies. Future research could explore the integration of additional variables and advanced machine learning techniques to further improve predictive accuracy and applicability in diverse health contexts.

4. CONCLUSIONS

Based on the results of research that has been carried out in the discussion above, the results of simulations carried out in 6 regions, the Naïve Bayes method obtained the highest accuracy value of 83.33% in simulation 1 and 66.67% in simulation 2. The smallest RMSE values were 0.382 in simulation 1 and 0.469 in simulation 2. This shows that the Naïve Bayes method can make good predictions.

REFERENCES

[1] R. Abdila, "Kominfo ajak masyarakat turunkan Prevalensi Stunting," Kominfo. Accessed: Dec. 03, 2023. [Online]. Available: https://www.kominfo.go.id/content/detail/17436/kominfo-ajak-masyarakat-turunkan-prevalensi-stunting/0/sorotan_media#:~:text=Upaya%20pemerintah%20mencegah%20stunting%20dilakukan,untuk%20meningkatkan%20status%20gizi%20anak

[2] D. F. Susanti, "Mengenal Apa Itu Stunting…," Kementerian Kesehatan Direktorat Jenderal Pelayanan Kesehatan. Accessed: Dec. 03, 2023. [Online]. Available: https://yankes.kemkes.go.id/view_artikel/1388/mengenal-apa-itu-stunting#:~:text=Sahabat%20sehat%2C%20definisi%20stunting%20sendiri,badannya%20berada%20di%20bawah%20standar

[3] Tim Medis Siloam Hospitals, "Mengenal Stunting: Pengertian, Penyebab, Serta Pencegahan," Siloam Hospitals. Accessed: Dec. 03, 2023. [Online]. Available: https://www.siloamhospitals.com/informasi-siloam/artikel/apa-itu-stunting

[4] Jatim Newsroom, "Targetkan Stunting Jatim Turun Hingga 13,5% Tahun 2024, Wagub Emil: Intervensi Harus Sesuai Data Riil Di Lapangan," Dinas Kominfo Provinsi Jatim. Accessed: Dec. 03, 2023. [Online]. Available: https://kominfo.jatimprov.go.id/berita/targetkan-stunting-jatim-turun-hingga-13-5-tahun-2024-wagub-emil-intervensi-harus-sesuai-data-riil-di-lapangan

[5] M. Y. Anshori *et al.*, "Estimation of closed hotels and restaurants in Jakarta as impact of corona virus disease spread using adaptive neuro fuzzy inference system," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 11, no. 2, pp. 462–472, Jun. 2022.

[6] F. S. Nugraha, M. J. Shidiq, and S. Rahayu, "Analisis Algoritma Klasifikasi Neural Network Untuk Diagnosis Penyakit Kanker Payudara," *Jurnal Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 149–156, Aug. 2019.

[7] F. A. Susanto *et al.*, "Estimation of Closed Hotels and Restaurants in Jakarta as Impact of Corona Virus Disease (Covid-19) Spread Using Backpropagation Neural Network," *Nonlinear Dynamics and Systems Theory*, vol. 22, no. 4, pp. 457–467, 2022.

[8] M. Y. Anshori, I. H. Santoso, T. Herlambang, D. Rahmalia, K. Oktafianto, and P. Katias, "Forecasting of Occupied Rooms in the Hotel Using Linear Support Vector Machine," 2023.

[9] F. S. Rini, T. D. Wulan, and T. Herlambang, "Forecasting the Number of Demam Berdarah Dengue (DBD) Patients Using the Fuzzy Method at the Siwalankerto Public Health Center," in *AIP Conference Proceedings*, American Institute of Physics Inc., Jan. 2023.

[10] A. Muhith, I. H. Susanto, D. Rahmalia, D. Adzkiya, and T. Herlambang, "The Analysis of Demand and Supply of Blood in Hospital in Surabaya City Using Panel Data Regression," *Nonlinear Dynamics and Systems Theory*, vol. 22, no. 5, pp. 550–560, 2022.

[11] M. Y. Anshori, T. Herlambang, P. Katias, F. A. Susanto, and R. R. Rasyid, "Profitability estimation of XYZ company using H-infinity and Ensemble Kalman Filter," in *The 5th International Conference of Combinatorics, Graph Theory, and Network Topology (ICCGANT 2021)*, IOP Publishing Ltd, Jan. 2021.

[12] D. F. Karya, P. Katias, and T. Herlambang, "Stock Price Estimation Using Ensemble Kalman Filter Square Root Method," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Apr. 2018. doi: 10.1088/1742-6596/1008/1/012017.

[13] C. N. Dengen, K. Kusrini, and E. T. Luthfi, "Implementasi Decision Tree Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu," *SISFOTENIKA*, vol. 10, no. 1, p. 1, Jan. 2020, doi: 10.30700/jst.v10i1.484.

[14] D. B. Magfira *et al.*, "Electronic Nose for Classifying Civet Coffee and Non-Civet Coffee," 2023.

[15] R. Setiawan and A. Triayudi, "Klasifikasi Status Gizi Balita Menggunakan Naïve Bayes dan K-Nearest Neighbor Berbasis Web," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 2, p. 777, Apr. 2022, doi: 10.30865/mib.v6i2.3566.

[16] Y. Findawati, I. R. I. Astutik, A. S. Fitroni, I. Indrawati, and N. Yuniasih, "Comparative analysis of Naïve Bayes, K Nearest Neighbor and C.45 method in weather forecast," *J Phys Conf Ser*, vol. 1402, no. 6, p. 066046, Dec. 2019, doi: 10.1088/1742-6596/1402/6/066046.

[17] R. Setiawan and A. Triayudi, "Klasifikasi Status Gizi Balita Menggunakan Naïve Bayes dan K-Nearest Neighbor Berbasis Web," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 2, p. 777, Apr. 2022, doi: 10.30865/mib.v6i2.3566.

[18] A. P. Permana, K. Ainiyah, and K. F. H. Holle, "Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 6, no. 3, pp. 178–188, Sep. 2021, doi: 10.14421/jiska.2021.6.3.178-188.

[19] V. Asy'ari, M. Y. Anshori, T. Herlambang, I. W. Farid, D. Fidita Karya, and M. Adinugroho, "Forecasting average room rate using k-nearest neighbor at Hotel S," in *2023 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA)*, IEEE, Nov. 2023, pp. 496–500.

[20] M. R. Romadhon and F. Kurniawan, “A Comparison of Naive Bayes Methods, Logistic Regression and KNN for Predicting Healing of Covid-19 Patients in Indonesia,” in *2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, IEEE, Apr. 2021, pp. 41–44. doi: 10.1109/EIConCIT50028.2021.9431845.

[21] W. Ananda, M. Safii, and M. Fauzan, “Prediksi Jumlah Hasil Panen Sawit Menggunakan Algoritma Naive Bayes,” *Terapan Informatika Nusantara*, vol. 1, no. 10, 2021.

[22] V. Jackins, S. Vimal, M. Kaliappan, and M. Y. Lee, “AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes,” *J Supercomput*, vol. 77, no. 5, pp. 5198–5219, May 2021, doi: 10.1007/s11227-020-03481-x.

[23] Trivusi, “Pengertian dan Contoh Algoritma Naive Bayes Classifier,” Trivusi. Accessed: Dec. 06, 2023. [Online]. Available: https://www.trivusi.web.id/2022/07/algoritma-naive-bayes.html

[24] R. Setiawan and A. Triayudi, “Klasifikasi Status Gizi Balita Menggunakan Naïve Bayes dan K-Nearest Neighbor Berbasis Web,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 2, p. 777, Apr. 2022, doi: 10.30865/mib.v6i2.3566.

[25] Trivusi, “Yuk Kenali Apa itu Algoritma K-Nearest Neighbors (KNN),” Trivusi. Accessed: Dec. 06, 2023. [Online]. Available: https://www.trivusi.web.id/2022/06/algoritma-knn.html

[26] U. Hidayah and A. Sifaunajah, *Cara Mudah Memahami Algortima K-Nearest Neighbor Studi Kasus Visual Basic 6.0*. Jombang: LPPM Universitas KH. A. Wahab Hasbullah, 2019.

[27] F. Gorunescu, *Data Mining: Concepts, Models and Techniques*, vol. 12. Berlin, Heidelberg: Springer Science & Business Media, 2011.

[28] V. Asy’ari, M. Y. Anshori, T. Herlambang, I. W. Farid, D. Fidita Karya, and M. Adinugroho, “Forecasting average room rate using k-nearest neighbor at Hotel S,” in *2023 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA)*, IEEE, Nov. 2023, pp. 496–500. doi: 10.1109/ICAMIMIA60881.2023.10427942.