Application of Naive Bayes and Support Vector Machine Algorithms to Classify Exclusive Breastfeeding in Indonesia

Muh Zarkawi Yahya[a], Sri Astuti Thamrin[b*], Erna Tri Herdiani[c]

[a]zarkawi2206@gmail.com, [b]tuti@unhas.ac.id, [c]herdiani.erna@unhas.ac.id
[a,b,c] Department of Statistics, Faculty of Mathematics and Natural Sciences, Hasanuddin University, Makassar, Indonesia

Abstract

The World Health Organization (WHO) and the United Nations International Children's Emergency Fund (UNICEF) both advocate for initiation of breastfeeding within one hour of a baby's birth. The Indonesian government supported the WHO and UNICEF programs by implementing exclusive breastfeeding programs in 2012. Exclusive breastfeeding is recommended for the first 6 months of a child's life, followed by the introduction of appropriate and safe complementary nutrition while continuing breastfeeding for up to 2 years or more. There are 18 factors that can influence the success of the exclusive breastfeeding program, and these factors will be taken into consideration. To analyze the impact of these factors, the Naive Bayes algorithm and Support Vector Machine (SVM) using JASP were employed. The results indicate that SVM achieved a higher accuracy rate of 99.57% compared to the Naive Bayes method, which yielded an accuracy rate of 99.31%, when applied to exclusive breastfeeding data in Indonesia.

Keywords: Naïve Bayes, Support Vector Machine, Exclusive Breastfeeding.

1. Introduction

Breast milk, a divine gift from Allah, serves as nourishment for infants and offers protection against illnesses (Hasniati et al., 2015). It is considered the ideal nourishment for babies due to its complete blend of essential nutrients crucial for their growth and overall development (Putri et al., 2021). According to a report by the WHO, the worldwide average rate of exclusive breastfeeding in 2017 was merely 38%. The WHO aims to raise this rate by at least 50% by the year 2025 (WHO, 2017). In 2017, the Indonesian Ministry of Health reported that a significant proportion of women in Indonesia, approximately 96%, breastfeed their children. However, only 42% of these women adhere to the exclusive provision of breastfeeding for the recommended period of 6 months (Riskesdas, 2018).

To address the aforementioned challenges, a data processing system is required, specifically classification. Classification enables the grouping of data into distinct categories, facilitating ease of processing and analysis. Classification is a technique used to group data into distinct categories, making it easier to process and analyze (Wibawa et al., 2018). Commonly used classification methods in the field of statistics include discriminant analysis and logistic regression (Abdulazeez, 2021). The growing prevalence of the data era has led to a rapid increase in the volume of data, resulting in large data sets commonly referred to as "big data". To extract valuable insights from these large data sets and transform them into organized knowledge, there is an urgent need for a powerful and versatile analytical tool (Barstugan et al., 2021).

The swift progress of artificial intelligence technology has led to the emergence of machine learning methods, enabling machines to autonomously learn without explicit user guidance. These advancements draw upon various disciplines such as statistics, mathematics, and data mining (J. Han & Pei, 2012). One of the

tasks that can be accomplished through data mining is classification (Hendrian, 2018). Classification methods commonly used in data mining include classification and regression trees (CART), random forests, Naive Bayes, support vector machines (SVM), K-nearest neighbors (KNN), and neural networks (Sihombing & Yuliati, 2021).

The Naive Bayes method stands out as a highly effective and efficient inductive learning algorithm within the domains of machine learning and data mining (Muin et al., 2016). In comparison to other classification models, Naive Bayes exhibits exceptional performance. This assertion is supported by Xhemali and Hinde Stone in their publication titled "Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages". They concluded that Naive Bayes exhibits a higher level of accuracy compared to other classification models (Widianto, 2019). In their research on diabetes, Kumari and Chitra (Kumari & R.Chitra, 2014) aimed to classify the disease using the SVM method. The data used in this study consisted of 460 data points from the Pima Indian diabetes dataset. The classification outcomes yielded an average accuracy rate of 78%.

To assess the continuity of exclusive breastfeeding in Indonesia, researchers commonly employ the Naive Bayes algorithm and Support Vector Machine (SVM) for classification. Hence, in this paper, we adopt both the Naive Bayes algorithm and SVM to classify the continuity of exclusive breastfeeding in Indonesia.

2. Materials And Methods

2.1 Data Source

The data utilized in this study was sourced from the Indonesia Family Life Survey Batch 5 (IFLS-5), obtained from https://www.rand.org. Subsequently, a pre-processing phase was conducted to identify and eliminate irrelevant attributes, ensuring suitability for the subsequent stage of classification using the Naive Bayes method and SVM.

2.2 Classification using Naive Bayes

The approach from Naïve Bayes theorem, according to Kusumadewi (Kusumadewi, 2003) is as follows:

$$P\left(c_i|\boldsymbol{x}\right) = \frac{P(\boldsymbol{x}|c_i)\,P(c_i)}{P(\boldsymbol{x})} \tag{1}$$

Information:
$P(c_i|\boldsymbol{x})$: Probability of document X in category C_i
$P(\boldsymbol{x}|c_i)$: Opportunity in category C_i where the word in document X appears in that category.
$P(c_i)$: The odds of a given category, compared to the other categories analyzed
$P(\boldsymbol{x})$: Opportunities from the document specifically.

In its development, $P(\boldsymbol{x})$ can be removed because its value is fixed, so when compared with each category, this value can be removed. The Naive Bayes approach has a simple assumption, namely the assumption that all attributes are mutually independent or do not influence each other:

$$\begin{aligned} P(\boldsymbol{x}|c_i) &= P(x_1, x_2, \dots, x_d|c_i) \\ &= P(x_1|c_i)\,P(x_2|c_i) \dots P(x_d|c_i) \\ &= \prod_{j=1}^{d} P(x_j|c_i) \end{aligned} \tag{2}$$

In the case of numeric attributes, we adopt the default assumption that each atrribute follows a normal

distribution within each class, denoted as c_i. The mean and variance of attribute x_j for class c_i are represented as μ_{ij} and σ_{ij}^2, respectively. The probabilities for class c_i on dimension x_j, are calculated as follows:

$$P(x_j|c_i) \propto f(x_j|\mu_{ij},\sigma_{ij}^2) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \exp\left\{-\frac{(x_j-\mu_{ij})^2}{\sigma_{ij}^2}\right\} \quad (3)$$

For Naïve Bayes classification, the sample mean $\hat{\mu}_i = (\hat{\mu}_{i1},\dots,\hat{\mu}_{id})^T$ and sample covariance matrix diagonal $\hat{\sum}_i = diag(\sigma_{i1}^2,\dots,\sigma_{id}^2)$ are employed for each class c_i. The algorithm presents pseudocode outlining the steps for performing naive Bayes classification. Given an input dataset $\boldsymbol{D}$, this method estimates the prior and mean probabilities for each class.

There are attribute category assumptions of independence leading to a simplification of the size of the probability mass function (PMF) as in equation (4).

$$P(\boldsymbol{x}|c_i) = \prod_{j=1}^{d} P(x_j|c_i) = \prod_{j=1}^{d} f(\boldsymbol{X}_j = \boldsymbol{e}_{jrj}|c_i) \quad (4)$$

Where $f(\boldsymbol{x}_j = \boldsymbol{e}_{jrj}|c_i)$ is the PMF for $\boldsymbol{x}_j$, which can be estimated from $\boldsymbol{D}_i$.

$$\hat{f}(\boldsymbol{v}_j|c_i) = \frac{n_i(v_j)}{n_i} \quad (5)$$

Where $n_i(v_j)$ is the observed frequency of the value $v_j = e_j r_j$ corresponding to r_j the value of the category a_{jrj} for attribute $\boldsymbol{X}_j$ in class c_i. As in the case of Bayes' theorem, if the count is zero, we can utilize the pseudo-count method to determine the prior probability. Estimating with the provided pseudo-count is as follows:

$$\hat{f}(\boldsymbol{v}_j|c_i) = \frac{n_i(v_j)+1}{n_i+m_j} \quad (6)$$

Where $m_i = |dom(\boldsymbol{X}_i)|$.

2.3 Classification Using SVM

Support Vector Machine (SVM) is a powerful technique used for making predictions in both classification and regression cases (Santosa, 2007). The SVM technique is utilized to obtain the optimal hyperplane function for separating observations with distinct target variable values. SVM operates on the basic principle of a linear classifier, which involves classifying cases that can be separated linearly. However, SVM has been further developed to handle non-linear problems by incorporating the kernel concept in high-dimensional workspaces.

In high-dimensional space, a hyperplane is sought to maximize the distance (margin) between data classes. According to Santosa (Santosa, 2007), the hyperplane for linear classification in SVM is denoted as follows:

$$f(x) = \boldsymbol{w}^T\boldsymbol{x} + b \quad (7)$$

So according to Vapnik and Cortes(Cortes & Vapnik, 1995), an equation is obtained

$$\left[\left(\boldsymbol{w}^T.\boldsymbol{x}_i\right) + b\right] \geq 1 \text{ for } y_i = +1 \quad (8)$$

$$\left[\left(\boldsymbol{w}^T.\boldsymbol{x}_i\right)+b\right] \leq 1 \text{ for } y_i = -1$$

where $\boldsymbol{x}_i$ = training data set, $i = 1,2,\ldots n$ and y_1 = class label of $\boldsymbol{x}_i$. To obtain the optimal hyperplane, locate the hyperplane that lies equidistant between the two class boundary fields. To achieve this, maximize the margin or distance between the two sets of objects belonging to different class (Santosa, 2007). The margin can be calculated by $\frac{2}{\|\boldsymbol{w}\|}$.

Finding the optimal hyperplane can be achieved by using the quadratic programming (QP) problem method, which involves minimizing

$$\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} \tag{9}$$

With the provision of $y_1(\boldsymbol{w}^T.\boldsymbol{x}_i + b) \geq 1, \quad i = 1,2,3\ldots n.$

The optimization solution proposed by Cortes and Vapnik (Cortes & Vapnik, 1995) is solved by utilizing the Lagrange function in the following manner:

$$L(\boldsymbol{w}, b, a) = \frac{1}{2}\boldsymbol{w}^T - \sum_{i=1}^{n} a_i\left\{y_1\left[\left(\boldsymbol{w}^T.\boldsymbol{x}_i\right)+b\right]-1\right\} \tag{10}$$

where α_i represents the Lagrange multiplierss and i varies from 1 to n,the optimal value can be determined by maximizing L with respect to α_i and minimizing L with respect to $\boldsymbol{w}$ and b. This is like a dual-problem case (Gunn, 1998).

$$max_{\alpha}W(\alpha) = max_{\alpha}\left(min_{w,b}L(w,b,a)\right) \tag{11}$$

The minimum value of the Lagrange function can be expressed as follows:

$$\begin{aligned}\frac{\partial L}{\partial b} = 0 \implies \sum_{i=1}^{n}\alpha_i y_i = 0\\ \frac{\partial L}{\partial b} = 0 \implies \boldsymbol{w} = \sum_{i=1}^{n}\alpha_i \boldsymbol{x}_i y_i\end{aligned} \tag{12}$$

According to Santosa (Santosa, 2007), in order to simplify Eq. (12), it is necessary to convert it into the Lagrange multiplier function, resulting the following form.

$$(\boldsymbol{w}, b, a) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} - \sum_{i=1}^{n} a_i y_1(\boldsymbol{w}^T.\boldsymbol{x}_i) - b\sum_{i=1}^{n} a_i y_1 + \sum_{i=1}^{n} a_i \tag{13}$$

Then we have L_d as follows (Hastie et al., 2001):

$$L_d = \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n} a_i a_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j \tag{14}$$

and obtained a dual problem

$$max_{\alpha}L_d = \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n} a_i a_j y_i y_j x_i^T x_j \tag{15}$$

with limitations, $\alpha_i \geq 0, i = 1,2,\ldots,n$, and $\sum_{i=1}^{n} a_i y_1 = 0$.

The training data that falls on the hyperplane with $\alpha_i \geq 0$ is referred to as a support vector, while the training data that does not lie on the hyperplane has $a_i = 0$. After a solution to the quadratic programming

problem is found, the class of data to be predicted or tested can be ascertained by evaluating the ensuing function:

$$f(x_t) = \sum_{s=1}^{ns} \alpha_s y_s x_s . x_t + b \quad (16)$$

Information
x_t = data whose class will be predicted (testing data)
x_s = data support vector, $s = 1,2, \ldots ns$
ns = many data support vectors

2.4 Clasification performance measurement

The evaluation of classification performance is conducted by comparing the original data and the resulting data obtained from the classification model. This assessment is performed using a cross-tabulation, commonly known as a confusion matrix. The confusion matrix provides valuable information as it presents the original data classes along the matrix rows and the predicted data classes produced by the algorithm along the matrix columns (Geetha et al., 2020). The following is an example of a confusion matrix.

Table 1. Confusion matrix of three classes.

	F_{gh}	**Predicton class (h)**		
		Class 1	Class 2	Class 3
Original Class(g)	Class 1	F_{11}	F_{12}	F_{13}
	Class 2	F_{21}	F_{22}	F_{23}
	Class 3	F_{31}	F_{32}	F_{33}

Before classifying the data, testing is carried out by adjusting the distribution of training data and testing data to determine the performance of the Support Vector Machine model, namely the values of Accuracy, Precision, and recall. This test is carried out to find out how much influence the amount of training and testing data has on the classification carried out by the Support Vector Machine. Testing is done using the Confusion Matrix method. The testing model is shown in Table 2

Tabel 2. Skenario Pengujian Pada Support Vector Machine

Pengujian	**Training %**	**Testing %**
1	25	75
2	35	65
3	45	55
4	55	45
5	65	35
6	75	25
7	85	15
8	90	10

3. Results and Discussion

3.1 Classification of SVM

In the SVM classification model, training data is employed to predict the continuity classification of exclusive breastfeeding, categorizing it as poor, good, or very Good. To develop the model, training data is utilized, while testing data is employed to assess the performance of the trained model. Table 2 presents the evaluation of the SVM model, wherein a test is conducted by varying the percentage of training and testing data during the data splitting process. Table 2 demonstrates the successful classification of data on the continuity of exclusive breastfeeding by the SVM model. Through 8 iterations of adjusting the training and testing data, the model achieves an impressive average accuracy of 99.57%, precision of 99.94%, and recall of 99.94%.

Table 3. Average value of accuracy, precision and recall of SVM model

Trial	Training (%)	Testing (%)	Accuracy (%)	Precision (%)	Recall (%)
1	25	75	98.8	98.20	98.20
2	35	65	99.60	99.40	99.40
3	45	55	99.80	99.70	99.70
4	55	45	99.70	99.60	99.60
5	65	35	99.40	99.10	99.10
6	75	25	99.50	99.30	99.30
7	85	15	99.70	99.50	99.50
8	90	10	99.80	99.70	99.70
Average			99.54	99.31	99.31

3.2 Classification of Naïve Bayes

The Naïve Bayes classification model employs training data to predict the continuity classification of exclusive breastfeeding, categorizing it into the poor, good, or very good categories. The production of the model requires training data, while testing data is utilized to evaluate the model's performance. Subsequently, the accuracy of the classification results is assessed using the confusion matrix. During this stage, testing is performed by adjusting the distribution of training and testing data to gauge the performance of the Naive Bayes model, specifically examining accuracy, precision, and recall values.

Table 4. Average value of accuracy, precision and recall of Naïve Bayes model.

Trial	Training (%)	Testing (%)	Accuracy (%)	Precision (%)	Recall (%)
1	25	75	98.8	98.20	98.20
2	35	65	99.60	99.40	99.40
3	45	55	99.80	99.70	99.70
4	55	45	99.70	99.60	99.60
5	65	35	99.40	99.10	99.10
6	75	25	99.50	99.30	99.30
7	85	15	99.70	99.50	99.50
8	90	10	99.80	99.70	99.70
Average			99.54	99.31	99.31

Demonstrates the successful classification of exclusive breastfeeding categorization data by the Naïve Bayes method through 8 trials. Adjustments made to the training and testing data yield an average accuracy of 99.54%, precision of 99.31%, and recall of 99.31%.

From the results of testing each method, which was tested eight times for each method, a comparison was made between the Support Vector Machine and the Nave Bayes method, as shown in Table 5.

Table 5. Comparison of SVM and Naïve Bayes testing results.

Trial	Training (%)	Testing (%)	Support vector machine			Naïve Bayes		
			Accuracy (%)	Precision (%)	Recall (%)	Accuracy (%)	Precision (%)	Recall (%)
1	25	75	99.97	99.50	99.50	98.8	98.20	98.20
2	35	65	99.70	99.60	99.60	99.60	99.40	99.40
3	45	55	100	100	100	99.80	99.70	99.70
4	55	45	100	100	100	99.70	99.60	99.60
5	65	35	100	100	100	99.40	99.10	99.10
6	75	25	100	100	100	99.50	99.30	99.30
7	85	15	100	100	100	99.70	99.50	99.50
8	90	10	100	100	100	99.80	99.70	99.70
Average			99.57	99.94	99.94	99.54	99.31	99.31

Based on Table 4, it shows that the comparison of training data greatly affects the accuracy, precision, and recall of the two methods. From the results of trials conducted, the level of accuracy for the classification of continuity of exclusive breastfeeding using Indonesia Family Life Survey Batch 5 (IFLS-5) data between the SVM and Naive Bayes algorithms shows that the average value of accuracy, precision, and recall of the SVM outperforms Naive Bayes by 99.57%, 99.94%, and 99.94%, while the average value of accuracy, precision, and recall of Naive Bayes is 99.54%, 99.31%, and the average precision value of the Naive Bayes algorithm outperforms the SVM by 90.44%, while the SVM is 87.05%.

4. Conclusions

In a comparative analysis of the classification methods of Naive Bayes and SVM in classifying continuity of exclusive breastfeeding, it can be concluded that the accuracy of the Naive Bayes results is lower than SVM. Support Vector Machine (SVM) obtains an average percentage of 99.57% accuracy, precision of 99.94%, and recall of 99.94%, while the Naive Bayes method can only obtain an average accuracy of 99.54%, precision of 99.31%, and recall of 99.31%.

5. Conflicts of Interest

The authors declare that there is no conflict of interest regardng the publication of this paper.

6. Acknowledgement

We express our gratitude to the editor and reviewers for their valuable contributions towards enhancing the quality of the manuscript.

7. References

Abdulazeez, A. M. (2021). Data Mining Classification Algorithms for Analyzing Soil Data. Asian Journal of Research in Computer Science, 8(2), 18–28. https://doi.org/10.9734/AJRCOS/2021/v8i230196

Barstugan, M., Ozkaya, U., & Ozturk, S. (2021). Coronavirus (COVID-19) Classification using CT Images by Machine Learning Methods. CEUR Workshop Proceedings, 5, 1–10.

Cortes, C., & Vapnik, V. (1995). Support Vector Networks. Machine Learning. Kluwer Academic Publishers, 20, 273–297.

Geetha, P., Kannan, S., & Muthukumaravel, A. (2020). Sentiment classification on twitter data using support vector machine. Malaya Journal of Matematik, S(2), 4552–4555.

Gunn, S. R. (1998). Support vector machines for classification and regression. University of Southampton.

Hasniati, Y., Farika Indah, M., Asrinawaty, & Kasman. (2015). Determinant Exclusive Breastfeeding in Barito Kuala South Kalimantan. Jurnal MKMI, 11(1), 39–43.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). The Elements of Statistical Learning Data Mining, Inference, and Prediction. Springer series in Statistics.

Hendrian, S. (2018). Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan. Faktor Exacta, 11(3), 266–274. https://doi.org/10.30998/faktorexacta.v11i3.2777

J. Han, M. K., & Pei, J. (2012). Data Mining Concepts and Techniques. Elsevier.

Kumari, V. A., & R.Chitra. (2014). Classification of diabetes disease using support vector machine. Microcomput. Dev. 3, 1797–1801. 3(2), 1797–1801.

Kusumadewi, S. (2003). Artificial Intelligence (Teknik dan Aplikasinyua). Graha Ilmu.

Muin, Ashari, A., & Syarli. (2016). Metode Naive Bayes Untuk Prediksi Kelulusan. Jurnal Ilmiah Ilmu Komputer, 2(1), 22–26. https://media.neliti.com/media/publications/283828-metode-naive-bayes-untuk-prediksi-kelulu-139fcfea.pdf

Putri, F. P., Katmawanti, S., & Fanani, E. (2021). Correlation Between Use of The Contraception and Exclusive Breastfeeding In Indonesia in 2017 (2017 IDHS Analysis Data). RSF Conference Series: Medical and Health Science, 1(1), 42–52.

Riskesdas, kementrian K. R. I. (2018). Hasil Utama Riset Kesehatan Dasar.

Santosa, B. (2007). Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis. Graha Ilmu.

Sihombing, P. R., & Yuliati, I. F. (2021). Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia. MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer, 20(2), 417–426. https://doi.org/10.30812/matrik.v20i2.1174

WHO. (2017). Exclusive Breastfeeding For Optimal Growth, Development And Health Of Infants (2020th ed.). WHO.

Wibawa, A. P., Muhammad Guntur Aji Purnama, M. F. A., & Dwiyanto, F. A. (2018). Metode-Metode Klasifikasi. Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi, 3(1), 134.

Widianto, M. haldi. (2019). Algoritma Naive Bayes. Binus University.