Scientific Journal of Informatics
Vol. 10, No. 2, May 2023
p-ISSN 2407-7658 http://journal.unnes.ac.id/nju/index.php/sji e-ISSN 2460-0040

Hybrid Top-K Feature Selection to Improve High-Dimensional Data Classification Using Naïve Bayes Algorithm

Riska Wibowo[1*], M. Arief Soeleman[2], Affandy[3]
[1,2,3]Department of Informatics, Faculty of Computer Science, Universitas Dian Nuswantoro, Indonesia

Abstract.
Purpose: The naive bayes algorithm is one of the most popular machine learning algorithms, because it is simple, has high computational efficiency and has good accuracy. The naive bayes method assumes each attribute contributes to determining the classification result that may exist between attributes, this can interfere with the classification performance of naive bayes. The naïve bayes algorithm is sensitive to many features so this can reduce the performance of naïve bayes. Efforts to improve the performance of the naïve bayes algorithm by using a hybrid top-k feature selection method that aims to handle high-dimensional data using the naïve bayes algorithm to produce better accuracy.
Methods: This research proposes a hybrid top-k feature selection method with stages 1 Prepare the dataset, 2. Replace the missing value with the average value of each attribute, 3. Calculate the weight of the attribute value using the weight information gain method, 4. Determine the best k value by using the top-k weight relation with k =2 to 20 (according to the number of attributes), 5. Select attributes using the top-k feature selection method, 6. Backward Elimination with the naïve bayes algorithm, 7. Datasets that have been selected new attributes, then validated using 10 fold-cross validation where the data is divided into training data and testing data, 8. Calculate the accuracy value based on the confusion matrix table.
Result: Based on the experimental results of performance and performance comparison of several methods that have been presented (Naïve Bayes, deep feature weighting naïve bayes, top-k feature selection, and hybrid top-k feature selection). The experimental results in this study show that from 5 datasets from UCI Repository that have been tested, the accuracy value of the hybrid top-k feature selection method increases from the previous method. The accuracy comparison results show that the proposed hybrid top-k feature selection method is ranked as the best method.
Novelty: Thus, the Hybrid top-k feature selection method can be used to handle dimensional data in the Naïve Bayes algorithm.

Keywords: Naïve Bayes, Weight Information Gain, Top-K Feature Selection, Backward Elimination, Hybrid Feature Selection
Received February 2023 / **Revised** March 2023 / **Accepted** March 2023

INTRODUCTION

In agencies (companies), data in the form of text, sound, images and videos will increase over time, including applications in the field of medicine [1]. If the data is not utilized properly, it will become useless, because it needs more storage space. So that the data can be utilized again properly, it needs to be processed. The method of turning data into information is called data mining [2]. Data mining is the process of analyzing data from different perspectives and breaking it down/extracting it into useful information [3], [4]. The bigger and more data there is, the more data needs to be analyzed. Specialized techniques in Machine Learning are required for feature selection techniques where the data is complex, large in size, and prone to noise [5]. The problem is that most of the features from such a large data set are irrelevant or redundant, which usually affects the performance of Machine Learning [4]. To mitigate this, an effective way is to reduce the dimension of the feature space using feature selection techniques [4], [6]–[9].

Naive bayes algorithm is an algorithm used to solve classification problems based on the probability method [10]–[14]. Naive Bayes algorithm is a popular Machine Learning method used for classification algorithms because it works well, is efficient, and is simple. As one of the classification algorithms, Naive Bayes algorithm is good to use because it is efficient, simple and very sensitive to feature selection [15]. In this

*Corresponding author.
Email addresses: riskawibowokaka@gmail.com (Wibowo), m.arief.soeleman@dsn.dinus.ac.id (Soeleman), affandy@dsn.dinus.ac.id (Affandy)
DOI: 10.15294/sji.v10i2.42818

case, Naive Bayes shows good accuracy when the classification is considered balanced [16]. In [17] conducted a comparison of Naive Bayes, Neural Network, Decision Tree, Logistic Regression, and C4.5 algorithms. The results showed that the two best methods for solving classification problems were Naive Bayes and Logistic Regression. In research [18]conducted a comparison of the Naive Bayes and k-NN algorithms. The results showed that the accuracy of Naive Bayes was superior to k-NN. Naïve Bayes algorithm is included in the top 10 best algorithms, especially classification algorithms, so it was chosen in this study [19].

The Naïve Bayes algorithm has the disadvantage that it is very sensitive to many features resulting in low accuracy [15]. In order to increase the efficiency of the Naive Bayes algorithm on high-dimensional data using several methods, one of the methods used is feature selection [20]–[23]. The feature selection method uses three techniques, namely wrapper, filter and hybrid techniques [24]. Filter technology is used to build relationships between attributes based on the attributes inherent in the data. The wrapper technique selects features based on classification performance values. The hybrid technique is a combination of filter and wrapper techniques. In this research, the Weighted Information Gain method is used. The weighted information gain method is included in the filter technique. Determine the best k value by using the top-k weight relation with k =2 to 20 (according to the number of attributes). And the backward elimination method is included in the wrapper technique.

This research proposes high-dimensional data to calculate the weight value of each feature using weighted information gain. After calculating the weight value, attribute selection is then carried out using the top-k feature selection method. Then processed using backward elimination where the classification algorithm used in this research is the Naive Bayes algorithm to get the best accuracy. This research can be used as material for consideration and decision making in fields related to the weight information gain method, top-k feature selection, backward elimination and the naive bayes algorithm.

METHODS

The research methodology used in this study uses the experimental method. The experimental research method is a trial that is controlled by the researcher himself to investigate causal relationships (cause-and-effect relationships). The research steps include: 1. Problem analysis and literature review. This research begins by collecting technical papers and survey papers on the topic of high-dimensional data on the Naïve Bayes algorithm. 2. Dataset collection, the datasets used in this research are public UCI machine learning repository datasets, namely breast cancer, hepatitis, ionosphere, iris, and zoo datasets. 3. Data processing, in this study, datasets containing missing values (breast cancer and hepatitis datasets) are replaced with the average value of each attribute. 4. Proposed method, at this stage the breast cancer, hepatitis, ionosphere, iris, and zoo datasets are first selected for high-dimensional attributes. In the initial stage to determine the attributes that will be used in the calculation of the classification algorithm, namely the weight information gain method. Attributes are selected using the top-k method, and backward elimination in the Naïve Bayes classification method. 5. Stages of the experiment, determine the best k value by using the top-k weight relation with k =2 to 20 (according to the number of attributes). 6. Evaluation of results, after experimenting with all datasets with the proposed method, it produces accuracy values which are then evaluated and validated. From these results, conclusions can be drawn from this experimental research.

Problem Analysis and Literature Review

This research begins by collecting journal surveys, and technical papers related to the proposed method then knowing the state-of-the-art methods of research on the topic of high-dimensional data on the Naïve Bayes algorithm. This research uses datasets with high-dimensional data. The weakness of the Naive Bayes algorithm is that it is very sensitive to too many features, resulting in low accuracy. In this research, the dataset contains high-dimensional data and the weight value for each feature will be calculated using the weight information gain method. After the weight value is calculated, feature selection is performed and the features are selected using the top-k method. The classification algorithm used is Naive Bayes which gives the best accuracy value.

Data Collection

The dataset used in this research is a public dataset obtained from the UCI Repository which can be downloaded on the page https://archive.ics.uci.edu/ml/index.php. The following is a list of datasets used in this study that refers to previous research 1. breast cancer (286 records and 10 attributes), 2. hepatitis (155

records and 20 attributes), 3. ionosphere (351 records and 35 attributes), 4. iris (150 records and 5 attributes), and 5. zoo (101 records and 18 attributes).

Data Processing

The dataset used in this research is a dataset obtained from the UCI repository. In this study, there are datasets with missing values, datasets that have missing values will be replaced with the average value of each attribute [2].

The Proposed Method

At this stage, the proposed method in this study is explained, 1. Prepare the dataset, 2. Replace the missing value with the average value of each attribute, 3. Calculate the weight of the attribute value using the weight information gain method, 4. Select attributes using the top-k feature selection method, 5. Backward Elimination with the naïve bayes algorithm, 6. Datasets that have been selected for new attributes, then validated using 10 fold-cross validation where the data is divided into training data and testing data, 7. Calculate the accuracy value based on the confusion matrix table. The stages of the proposed method can be seen in Figure 1.

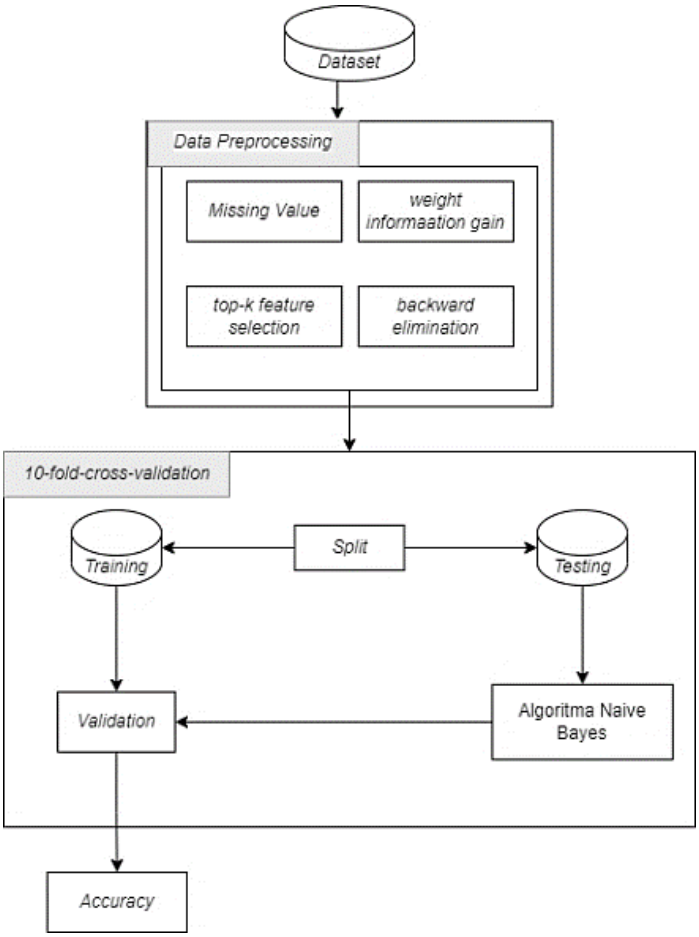

Figure 1. Stages of the proposed method

Stages of the experiment

Experiments were conducted using computers and supporting applications such as Microsoft Excel 365 and RapidMiner Studio 10.0.000. Several algorithms were used in this study as a comparison of the WIG and Backward Elimination methods, Naïve Bayes, Deep Feature Weighting Naive Bayes (DFWNB) [23], and Top-k Feature Selection Naïve Bayes to determine the best k value by using the top-k weight relation with k =2 to 20 (according to the number of attributes) [25].

Evaluation
At this stage, an evaluation of the research results is carried out. Evaluation is a measuring tool that can be used to evaluate or measure how good the proposed method is and whether the proposed method makes a significant difference in results. Evaluation or verification of research results is carried out by calculating the performance of the Naive Bayes classification algorithm using a confusion matrix. Confusion matrix is a table containing a two-dimensional matrix with one dimension showing the predicted value of the classification and the other dimension showing the true value of the classification [26].

Naïve Bayes
The Naïve Bayes algorithm is a popular machine learning technique for text classification, as it performs well, is efficient and very simple. As a classifier, Naïve Bayes is very efficient and simple and is very sensitive to feature selection [15]. The Naïve Bayes algorithm is a classification algorithm that uses probability and statistical methods. This algorithm was proposed by the English scientist Reverend Thomas Bayes. This algorithm predicts future opportunities based on previous experience, hence it is known as Bayes' Theorem. The theorem is combined with Naïve [27]. The use of Naïve Bayes algorithm has been introduced for a very long time, namely since 1702 - 1761. In Naïve Bayes algorithm, the presence or absence of certain characteristics of a class is not related to the characteristics of other classes. The Naïve Bayes algorithm is a classification algorithm that is simple, effective, fast, has a high level of accuracy and is the most widely used algorithm in the classification process compared to other classification algorithms. The Naïve Bayes algorithm is a popular machine learning technique for text classification, as it is simple, efficient and performs well in many domains. Naïve Bayes algorithm is widely used for text classification in machine learning which is based on probability features [28].

The steps of the Naïve Bayes method are as follows:
1. Prepare the dataset,
2. Count the number of classes in the dataset,
3. Calculate the amount of data that is the same as the same class,
4. Calculate by multiplying all the result values with the data for which the class is sought,
5. Calculate the classification result of Naïve Bayes algorithm using confusion matrix.

Weight Information Gain
The most popular way to weight each variable in an evaluation attribute is to use weight information gain [29]. In order to calculate information gain, we must first comprehend the concept of entropy. Entropy is a parameter that is frequently used in the field of information theory to assess the heterogeneity of a data sample set. If the data sample set is more heterogeneous, then the entropy value is greater [30]. Entropy is formulated in mathematics using Equations (1) and (2).

$$Entropy = \sum_i^C - p_i \, log_2 \, p_i \tag{1}$$

$$Gain(S, A) = Entropy(S) - \sum_{v \epsilon Value(A)} \frac{|S_v|}{|S|} \, Entropy(S_v) \tag{2}$$

Where C is the target attribute's number of values (number of classes). The number of samples for class i is represented by Pi. After collecting the entropy values for the data sample set, we can assess the effectiveness of the attributes. Information advantage is the name given to this efficiency metric. Mathematically, equations represent the acquisition of information from attribute A.

Where A is an attribute, V represents a potential value for A, and Value(A) represents the set of possible values for A. |Sv| is how many samples there are for value v. |S| represents the total amount of data, while Entropy (Sv) represents the entropy for samples with value v.

Backward Elimination
Backward Elimination is one of several computer-based repeated variable selection procedures. It starts with a model containing all independent variables of interest. Then, at each step, the variable with the smallest F-statistic is removed (if the F is not higher than the selected cutoff level). Here, it is usual to start

with the variable that has the lowest t-stat value. If the new model is better than the baseline, it becomes the new baseline and the search continues to remove the variable with the lowest t-stat value until some stopping criteria are met [31].

Rapidminer

RapidMiner is the most popular open source software that supports the design and documentation of the entire data mining process. Rapid Miner is widely used for data mining, machine learning, test mining and predictive analytics. The advantage and flexibility of RapidMiner are that it provides data mining and machine learning procedures including: ELT (extraction, transformation, loading), data preprocessing, visualization, modeling and evaluation. The data mining process is composed of nestable operators, described with XML (Extensible Mark-up Language), and created with a GUI (Graphical User Interface) that is easy for all to use[30].

Deep Feature Weighting Naïve Bayes (DFWNB)

In 2016 Jiang, Li proposed the Deep Feature Weight For Naïve Bayes (DFWNB) method. Where the DFWNB method is used to improve the accuracy of the Naïve Bayes algorithm by reducing high-dimensional data. The study used 36 datasets from the UCI repository, and used 10-fold cross validation as a validation method. The steps of the Deep Feature Weight For Naïve Bayes (DFWNB) method are as follows: 1. Prepare the dataset, 2. Use Correlation-based Feature Selection (CFS) to select features, 3. For each selected feature or attribute is given a weight value of w = 2, otherwise set the weight to w = 1, 4. Enter the weight value (w) obtained in the Naive Bayes algorithm using the weighted Naive Bayes formula, 5. Calculate the classification results of the Naive Bayes algorithm using Confusion Matrix and then measure the evaluation results [21].

Top-k Feature Selection

In 2020 Wibowo and Henny proposed the top-k feature selection method to determine the accuracy of whether a patient is detected with hepatitis disease or not. In this study using feature selection in data processing, and proposing a top-k feature selection method. The study used the hepatitis dataset from the UCI Repository, and used 10-fold cross validation as a validation method. The steps of the top-k feature selection method are as follows: 1. Prepare the dataset, 2. Calculate attribute weights using the weight information gain method, 3. Attribute selection using the top-k method, 4. Dataset after selecting new attributes, 5. Classification using the Naïve Bayes algorithm, 6. Calculate the accuracy value based on the confusion matrix table [23].

RESULT AND DISCUSSION

Experiments were conducted by testing 5 UCI repository datasets (breast cancer, hepatitis, ionosphere, iris, and zoo). The tested methods are naïve bayes classification method, naïve bayes algorithm attribute experiment based on deep feature weighting, naïve bayes algorithm attribute experiment based on top-k feature selection, and naïve bayes algorithm attribute experiment based on hybrid top-k feature selection. The experimental results of the method can be seen in Table 1 for the naïve bayes method, and the naïve bayes algorithm experiment based on deep feature weighting, Table 2 for the naïve bayes algorithm experiment method top-k feature selection, and in Table 3 can be seen a recap of the experimental results of the hybrid naïve bayes algorithm top-k feature selection.

Table 1. Accuracy results of naive bayes and deep feature weighting naive bayes (DFWNB)

Dataset	Number of attributes	Accuracy Naïve Bayes	Accuracy Deep Feature Weight Naïve Bayes (DFWNB)
breast_cancer	286	72.73%	71.68%
hepatitis	155	83.87%	85.16%
ionosphere	351	89.46%	91.20%
iris	150	95.33%	94.47%
zoo	101	95.05%	91.30%

Table 2. Accuracy result of top-k feature selection

Dataset	Number of attributes (Top-k)	Accuracy Top-k Feature Selection
breast_cancer	k=5	73.77%
hepatitis	k=11	85.81%
ionosphere	k=18	92.88%
iris	k=3	96.67%
zoo	k=9	97.03%

Table 3. Accuracy result of hybrid top-k feature selection

Dataset	Number of attributes (Top-k)	Accuracy Hybrid top-k feature selection
breast_cancer	k=9	75.17%
hepatitis	k=17	89.03%
ionosphere	k=18	93.73%
iris	k=4	96.67%
zoo	k=13	99.01%

Based on the experimental results of performance and performance comparison of several methods that have been presented (Naive Bayes, Deep Feature Weighting Naive Bayes (DFWNB), top-k feature selection, hybrid top-k feature selection) used in this study, it can be explained that the naïve bayes algorithm used to solve classification problems based on probability methods is very influential on high-dimensional data. From the accuracy comparison results in Figure 2, it can be seen that the proposed method is ranked the first best method.

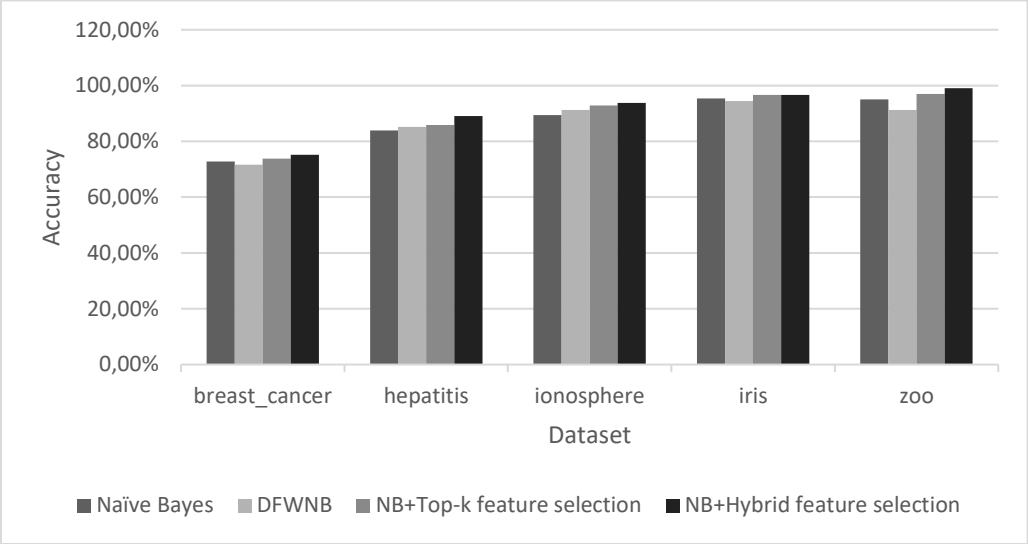

Figure 2. Comparison results of accuracy of each experiment method

CONCLUSION

A method is proposed to improve the performance of the Naïve Bayes algorithm, namely attribute weighting using the weight information gain method. After calculating the weight value, attribute selection is carried out, attributes are selected using the top-k method. Then processed again with the backward elimination method. The proposed method is then compared with previous related research, namely Naïve Bayes, Deep Feature Weighting Naïve Bayes (DFWNB) and top-k feature selection. The experimental results in this study show that from 5 datasets from UCI Repository that have been tested, the accuracy value increases from the previous method. This research has contributed, namely attribute weighting, the attribute selected is the attribute with the best weight value so that it can overcome high-dimensional data, namely data that has many attributes and each attribute is irrelevant.

REFERENCES

[1] H. Rao *et al.*, "Feature Selection Based On Artificial Bee Colony and Gradient Boosting Decision Tree," *Appl. Soft Comput.*, vol. 74, pp. 634–642, 2019, doi: 10.1016/j.asoc.2018.10.036.

[2] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems)*, 3rd Ed. USA: Morgan Kaufmann Publishers, 2012.

[3] Q. Guo and M. Zhang, "Implement Web Learning Environment Based On Data Mining," *Knowledge-Based Syst.*, vol. 22, no. 6, pp. 439–442, 2009, doi: 10.1016/j.knosys.2009.06.001.

[4] X. Sun, Y. Liu, M. Xu, H. Chen, J. Han, and K. Wang, "Feature Selection Using Dynamic Weights for Classification," *Knowledge-Based Syst.*, vol. 37, pp. 541–549, 2013, doi: 10.1016/j.knosys.2012.10.001.

[5] J. Apolloni, G. Leguizamón, and E. Alba, "Two Hybrid Wrapper-Filter Feature Selection Algorithms Applied to High-Dimensional Microarray Experiments," *Appl. Soft Comput.*, vol. 38, pp. 922–932, 2016, doi: 10.1016/j.asoc.2015.10.037.
[6] C.-F. Tsai and Y.-T. Sung, "Ensemble Feature Selection in High Dimension, Low Sample Size Datasets: Parallel and Serial Combination Approaches," *Knowledge-Based Syst.*, vol. 203, p. 106097, 2020, doi: 10.1016/j.knosys.2020.106097.
[7] S. B and M. K, "Firefly Algorithm Based Feature Selection for Network Intrusion Detection," *Comput. Secur.*, vol. 81, pp. 148–155, 2019, doi: 10.1016/j.cose.2018.11.005.
[8] M. R. Wijaya, R. Saptono, and A. Doewes, "The Effect of Best First and Spreadsubsample on Selection of a Feature Wrapper With Naïve Bayes Classifier for The Classification of the Ratio of Inpatients," *Sci. J. Inform.*, vol. 3, no. 2, pp. 139–148, 2016, doi: 10.15294/sji.v3i2.7910.
[9] N. P. Ririanti and A. Purwinarko, "Implementation of Support Vector Machine Algorithm with Correlation-Based Feature Selection and Term Frequency Inverse Document Frequency for Sentiment Analysis Review Hotel," *Sci. J. Inform.*, vol. 8, no. 2, pp. 297–303, 2021, doi: 10.15294/sji.v8i2.29992.
[10] D. M. Singh, N. Harbi, and M. Zahidur Rahman, "Combining Naive Bayes and Decision Tree for Adaptive Intrusion Detection," *Int. J. Netw. Secur. Its Appl.*, vol. 2, no. 2, pp. 12–25, 2010, doi: 10.5121/ijnsa.2010.2202.
[11] D. M. Farid *et al.*, "An Adaptive Ensemble Classifier for Mining Concept Drifting Data Streams," *Expert Syst. Appl.*, vol. 40, no. 15, pp. 5895–5906, 2013, doi: 10.1016/j.eswa.2013.05.001.
[12] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian Network Classifiers," *Mach. Learn.*, vol. 29, no. 2, pp. 131–163, 1997, doi: 10.1023/A:1007465528199.
[13] R. Zhang, W. Li, N. Liu, and D. Gao, "Coherent Narrative Summarization with A Cognitive Model," *Comput. Speech Lang.*, vol. 35, pp. 134–160, 2016, doi: 10.1016/j.csl.2015.07.004.
[14] L. Jiang, H. Zhang, Z. Cai, and D. Wang, "Weighted Average of One-Dependence Estimators," *J. Exp. Theor. Artif. Intell.*, vol. 24, no. 2, pp. 219–230, 2012, doi: 10.1080/0952813X.2011.639092.
[15] J. Chen, H. Huang, S. Tian, and Y. Qu, "Feature Selection for Text Classification with Naïve Bayes," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 5432–5435, 2009, doi: 10.1016/j.eswa.2008.06.054.
[16] H. Kang, S. J. Yoo, and D. Han, "Senti-lexicon and Improved Naïve Bayes Algorithms for Sentiment Analysis of Restaurant Reviews," *Expert Syst. Appl.*, vol. 39, no. 5, pp. 6000–6010, 2012, doi: 10.1016/j.eswa.2011.11.107.
[17] T. Hall, S. Beecham, D. Bowes, D. Gray, and S. Counsell, "A Systematic Literature Review on Fault Prediction Performance in Software Engineering," *IEEE Trans. Softw. Eng.*, vol. 38, no. 6, pp. 1276–1304, 2012, doi: 10.1109/TSE.2011.103.
[18] L. Dey, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari, "Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifier," *Int. J. Inf. Eng. Electron. Bus.*, vol. 8, no. 4, pp. 54–62, 2016, doi: 10.5815/ijieeb.2016.04.07.
[19] X. Wu *et al.*, "Top 10 Algorithms in Data Mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, 2008, doi: 10.1007/s10115-007-0114-2.
[20] J. Wu, S. Pan, X. Zhu, Z. Cai, P. Zhang, and C. Zhang, "Self-adaptive Attribute Weighting for Naive Bayes Classification," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1487–1502, 2015, doi: 10.1016/j.eswa.2014.09.019.
[21] L. Jiang, Z. Cai, and D. Wang, "Improving Naive Bayes for Classification," *Int. J. Comput. Appl.*, vol. 32, no. 3, 2010, doi: 10.2316/Journal.202.2010.3.202-2747.
[22] L. Jiang, D. Wang, Z. Cai, and X. Yan, "Survey of Improving Naive Bayes for Classification," in *Adv. Data Min. Appl.*, 2007, pp. 134–145, doi: 10.1007/978-3-540-73871-8_14.
[23] L. Jiang, C. Li, S. Wang, and L. Zhang, "Deep Feature Weighting for Naive Bayes and Its Application to Text Classification," *Eng. Appl. Artif. Intell.*, vol. 52, pp. 26–39, 2016, doi: 10.1016/j.engappai.2016.02.002.
[24] G. Chen and J. Chen, "A Novel Wrapper Method for Feature Selection and Its Applications," *Neurocomputing*, vol. 159, pp. 219–226, 2015, doi: 10.1016/j.neucom.2015.01.070.
[25] R. Wibowo and H. Indriyawati, "Top-k Feature Selection Untuk Deteksi Penyakit Hepatitis Menggunakan Algoritme Naïve Bayes," *J. Buana Inform.*, vol. 11, no. 1, p. 1, 2020, doi: 10.24002/jbi.v11i1.2456.
[26] X. Deng, Q. Liu, Y. Deng, and S. Mahadevan, "An Improved Method to Construct Basic Probability Assignment Based On The Confusion Matrix for Classification Problem," *Inf. Sci. (Ny).*, vol. 340–341, pp. 250–261, 2016, doi: 10.1016/j.ins.2016.01.033.
[27] Bustami, "Penerapan Algoritma Naive Bayes untuk Mengklasifikasi Data Nasabah Asuransi," *J.*

Inform., vol. 8, no. 1, pp. 884–898, 2014, doi: http://dx.doi.org/10.26555/jifo.v8i1.a2086.
[28] W. Zhang and F. Gao, "An Improvement to Naive Bayes for Text Classification," *Procedia Eng.*, vol. 15, pp. 2160–2164, 2011, doi: 10.1016/j.proeng.2011.08.404.
[29] V. Bolón-Canedo, I. Porto-Díaz, N. Sánchez-Maroño, and A. Alonso-Betanzos, "A Framework for Cost-Based Feature Selection," *Pattern Recognit.*, vol. 47, no. 7, pp. 2481–2489, Jul. 2014, doi: 10.1016/j.patcog.2014.01.008.
[30] J. Suntoro and C. N. Indah, "Average Weight Information Gain Untuk Menangani Data Berdimensi Tinggi Menggunakan Algoritma C4.5," *J. Buana Inform.*, vol. 8, no. 3, Oct. 2017, doi: 10.24002/jbi.v8i3.1315.
[31] V. Kotu and B. Deshpande, "Feature Selection," in *Predictive Analytics and Data Mining*, Elsevier, 2015, pp. 347–370.