Indonesian Journal of Electrical Engineering and Computer Science
Vol. 37, No. 1, January 2025, pp. 398~405
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v37.i1.pp398-405

Application of data mining for diagnosis of ENT diseases using the Naïve Bayes method with genetic algorithm feature selection

Linda Perdana Wanti, Nur Wachid Adi Prasetya, Ihza Awaludin,
Muhammad Bintang Aditya Saputra, Syamaidzar Nadifa Furi, Dimas Maulana Dwi Kumara
Department of Computer and Bussines, Politeknik Negeri Cilacap, Cilacap, Indonesia

Article Info

Article history:

Received May 25, 2024
Revised Sep 5, 2024
Accepted Sep 29, 2024

Keywords:

Data mining
Diagnosing
ENT
Implementation
Naïve Bayes

ABSTRACT

Ear, nose, and throat (ENT) disease is a disorder that occurs in the eustachian tube in one of the organs, be it the ear, nose, or throat. Early signs of ENT disease include sore throat, painful swallowing, swollen and red tonsils, runny nose, nosebleeds, blocked nose, discharge from the ears, and others. To determine the diagnosis, it is necessary to carry out a physical examination of the ears, nose, and throat as recommended by an expert, namely an ENT doctor. The research carried out was implementing data mining for the diagnosis of ENT diseases using the Naïve Bayes (NB) method. This method was chosen because it can increase the accuracy, efficiency, and accessibility of health services and is also easy to understand and apply to classify ENT disease symptom data. The NB method was used to build an ENT diagnosis classification model and the model performance was evaluated using accuracy, precision, and recall metrics. To increase the accuracy of the NB algorithm predictions, feature selection using a genetic algorithm can be used. Genetic algorithms can help select the most relevant and significant features, improving the accuracy of NB models by eliminating irrelevant or noisy features. By applying this method, predictions for ENT diseases can be produced with an accuracy of 95,67%.

Corresponding Author:

Linda Perdana Wanti
Department of Computer and Bussines, Politeknik Negeri Cilacap
Soetomo Street, No. 1, Sidakaya, South Cilacap, Cilacap Regency, Central Java, Indonesia
Email: linda_perdana@pnc.ac.id

1. INTRODUCTION

Ear, nose, and throat (ENT) disease is a disorder that occurs in the eustachian tube in one of the organs, be it the ear, nose, or throat [1]. Early signs of ENT disease include sore throat, painful swallowing, swollen and red tonsils, runny nose, nosebleeds, blocked nose, discharge from the ears, and others [2]. To determine the diagnosis, it is necessary to carry out a physical examination of the ears, nose, and throat as recommended by an expert, namely an ENT doctor. The ENT has important functions, such as hearing, breathing, smelling, speaking, and swallowing food and drink [3]. ENT disease is a type of disease or disorder related to these parts of the human body system. Examples include ear infections, rhinitis (inflammation of the nose), sinusitis (inflammation of the sinuses), tonsillitis (inflammation of the tonsils), and many more [4]. The causes of ENT disease can vary depending on the type of disease. ENT diseases are often caused by viral or bacterial infections [5]. The infection can affect the ears (otitis), nose (rhinitis), sinuses (sinusitis), or throat (pharyngitis). Allergies to pollen, dust, animal dander, or certain chemicals can cause ENT symptoms such as a blocked nose, sneezing, or itchy eyes [6]. Exposure to air pollution, such as cigarette smoke or industrial pollution, can increase the risk of infection or irritation of the upper respiratory tract. Abnormal anatomical structures, such as deviated septum (crooked nasal septum) or enlarged tonsils, can increase the risk of ENT disease [7]. Injury or trauma to the ear, nose, or throat can also cause health problems in the ENT system [8]. Exposure to harmful substances such as chemical fumes or toxic materials can also cause irritation or damage to the upper respiratory tract. Some ENT disorders can have a genetic component that contributes to the risk of developing the disease. Environmental factors such as cold or damp weather can also affect the ENT system and increase the risk of infection or other disorders [2].

The increasing number of ENT sufferers must be supported by the availability of adequate health facilities and facilities so that the treatment of patients can be carried out quickly [4]. The availability of experts in several health facilities still does not support rapid treatment of patients. By utilizing information technology, it can support services for treating ENT patients more effectively and efficiently. One way is to develop an independent diagnosis system for ENT patients, namely an expert system which is a branch of artificial intelligence development [9]. An expert system is a type of artificial intelligence system designed to imitate human problem-solving abilities in a certain domain [10]. They use preprogrammed knowledge and rules to analyze information, perform reasoning, and provide solutions or recommendations to users [11]. Expert systems usually consist of two main components, namely knowledge or knowledge base [12]. This component contains information and rules obtained from human experts in a specific domain. This knowledge is usually expressed in the form of “if-then” rules that relate problem conditions to appropriate actions or recommendations [13]. Next is the inference engine, in this section the expert system is responsible for carrying out reasoning based on available knowledge [14]. Inference engines apply existing rules to evaluate input information and produce conclusions or recommendations [15]. Expert systems in the health sector can assist in disease diagnosis or treatment planning by analyzing symptoms reported by patients and presenting recommendations based on available medical knowledge [16]. The main advantages of expert systems include the ability to provide consistent and fast recommendations, even in complex cases where many factors must be considered [14].

Several studies have been carried out by [17]–[21] who developed an expert system using several methods such as fuzzy Tsukamoto, forward chaining, dempster-shafer and certainty factor. The most recent research carried out compared to research that has already been carried out is that this research uses the Naïve Bayes (NB) method. The NB method is used to build an ENT diagnosis classification model and the model performance is evaluated using accuracy, precision, and recall metrics with the final result being the classification of data into predetermined categories or classes [22]. This method was chosen because it can increase the accuracy, efficiency, and accessibility of health services and is also easy to understand and apply to classify ENT disease symptom data [23]. The NB method was used to build an ENT diagnosis classification model and the model performance was evaluated using accuracy, precision, and recall metrics [24]. This research aims to develop an expert system for ENT diagnosis based on data mining and implement the NB method in diagnosing ENT diseases based on patient symptoms by selecting features using a genetic algorithm. Genetic algorithms can help select the most relevant and significant features, improving the accuracy of NB models by eliminating irrelevant or noisy features. By reducing the number of features through feature selection, the computation required to build and run a NB model becomes more efficient, both in terms of time and resources. Another novelty of this research compared to other research is that the NB method is relatively simple and fast to implement, and does not require a lot of parameter tuning [25]. Although more complex, many libraries and tools support the implementation of genetic algorithms, which makes their integration into the feature selection process easier. This research also has better generalization capabilities because this combination helps in reducing overfitting by eliminating irrelevant features, which can increase the model’s ability to generalize to data that has never been seen before.

2. METHOD

Data mining is a form of implementation that is applied to search for a model and pattern that can make predictions on data based on previous data in a certain period [26]. The NB classifier algorithm is one of the algorithms found in classification techniques [27]. NB is a classification using probability and statistical methods discovered by British scientist Thomas Bayes, namely predicting future opportunities based on experience, so it is known as Bayes’ theorem [28]. The NB method is a data classification technique for predicting the probability that data falls into a certain class [29]. NB uses Bayes’ theorem, which is a formula for calculating conditional probability which allows calculating the probability of event B occurring if event A has already occurred [30]. In classification problems, event A represents a certain class and event B represents a feature or characteristic of certain data [31]. In the NB method, the existence of one feature does not affect the existence of other features because of the class label [32].

The research was carried out by conducting experiments using a decision support system to predict patients with ENT disease, using data mining methods, namely the NB algorithm with the genetic algorithm selection feature. The data used to conduct research is primary data and secondary data. To measure the level of accuracy of predictions using the Rapid Miner tool 5. Feature selection using a genetic algorithm in the application of data mining for identifying ENT diseases using the NB method reduces the number of features that need to be processed, which can increase computational efficiency and reduce model complexity. Feature selection using genetic algorithms also selects the most relevant features, genetic algorithms can help increase the accuracy of predictive models and also reduce overfitting by eliminating irrelevant or redundant features. Genetic algorithms can find complex and non-linear combinations of features that may not be found with traditional feature selection methods and can be adapted to various types of problems and data, both linear and non-linear. This method is more robust to highly correlated features or noise in the data, due to the population-based selection process and iterative evaluation. Overall, the use of a combination of genetic algorithms for feature selection and NB for classification can provide a more optimal, efficient, and accurate solution in identifying ENT diseases in data mining applications.

Knowledge discovery in database (KDD) is a process of determining information that can be useful and producing patterns contained in the data [33]. The information contained in the database is large so that it cannot previously be known and is potentially useful. Data mining is a method or stages of a set of KDD iterative processes [34]: The KDD method process has five stages. The first is data selection or data selection by collecting data using secondary techniques obtained from an agreement between both parties between the communities that have been carried out. The dataset used is ENT patient data identified based on the symptoms experienced by the patient. Second, preprocessing or cleansing is the removal of duplicate data. Third, transformation or data selection with parameters 0.9 for training data and 0.1 for testing data. Fourth, data mining or building models using the NB algorithm [35]. Lastly is evaluating the model and looking for accuracy, precision, and recall values.

Stages of the NB method classification process, namely [36]:

a) Collect the dataset
Collect the dataset that will be used for training and testing the model. The dataset should include features (independent variables) and labels (dependent or target variables).
b) Data pre-processing
Eliminate missing data or outliers. Divide the dataset into training data and testing data. If there are categorical features, convert them to numeric because NB works better with numeric data.
c) Calculate the probability of each class
For each data involved, the probability for each class is calculated. NB implementation to combine a priori class probabilities with the probability of feature values from a given class [23].
d) Calculating posterior probabilities
Bayes' theorem is used to combine the prior probability of each class with the probability calculated from the previous step and to calculate the posterior probability, namely the probability that new data falls into a certain class based on its features.
NB method formula [37]:

$$P(h|D) = \frac{P(D|h) * P(h)}{P(D)} \tag{1}$$

Information:
h: data hypothesis with a certain class
D: data that does not yet have a class
P(h): hypothesis probability (prior probability)
P(D): probability of D
P(D|h): probability h based on condition D (posterior probability)
P(h|D): probability D based on the conditions in hypothesis h
e) Predict the most likely class
The performance of classification methods using NB methods is usually evaluated using metrics such as accuracy, precision, and recall, which measure how well the classifier correctly assigns data points to their original classes. The class with the highest probability generated becomes the prediction class for the new data[38].
f) Model evaluation
Test the model on test data to evaluate its performance using metrics such as accuracy, precision, recall, and F1-score [39].
g) Model refinement
Make improvements to the model by tuning parameters, or by improving data pre-processing.

For more details regarding the calculation of determining prior probability, likelihood calculation and posterior probability calculation, it is explained in Figure 1.

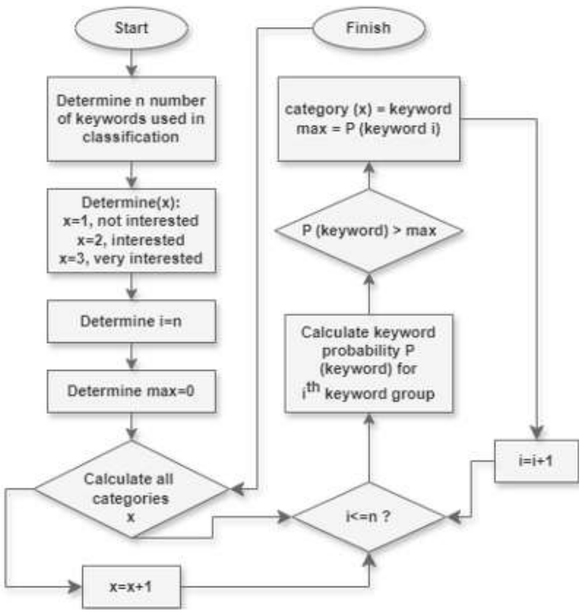

Figure 1. Stages of NB method

3. RESULTS AND DISCUSSION

The data used is primary data and secondary data where data is obtained directly from relevant sources related to the research carried out. By taking data from a reliable source, the data is ensured with reliable reality and validity. The dataset used is 208 medical data from patients who have symptoms leading to ENT during 2023. The method used is the NB algorithm to obtain accuracy, recall, and precision results [34]. The research process uses the Rapid Miner tool to process data. The following are the stages in using the rapid miner. Testing is carried out using the NB method with the genetic algorithm selection feature shown in Figure 2. This test allows the model to select the best features that are most relevant in predicting classes or categories, which are then used for classification [40].

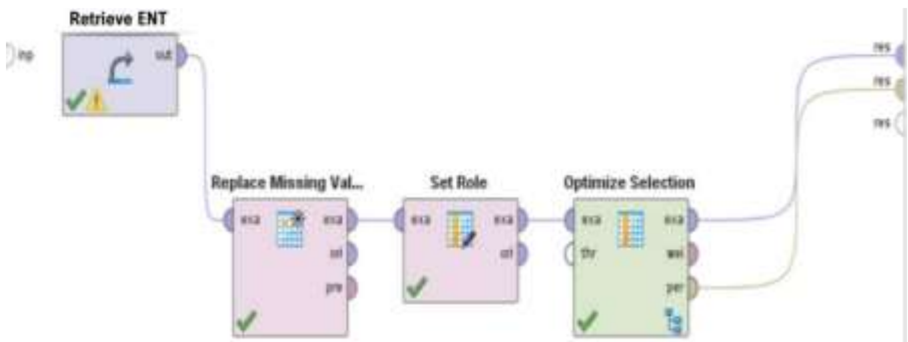

Figure 2. Data connection with optimaze selection

The ENT disease diagnosis database is connected to replace missing values to correct incomplete data, then connected to the set role attribute to make one of the attributes in the ENT disease diagnosis data into the ID or label of all the attributes in the dataset, which is used as a label for the dataset ENT disease diagnosis is a class attribute, then the set role is connected to an operator (evolutionary) to optimize the selection of attributes that are relevant to the process of predicting ENT disease diagnosis results [22]. In optimizing selection (evaluation) there is a cross-validation process as seen in Figure 3.

Figure 3. Cross operator validation for optimaze selection

Cross-validation used in this research is 10-fold validation. The dataset totaling 208 data with 20 attributes will be divided into 10 parts. Where each part will be formed randomly. The principle of 10-fold validation is 1:9, where 1 part becomes testing data and 9 parts become training data so that these 10 parts have the opportunity to become testing data [41]. After training and testing, the level of accuracy can be measured [39]. In cross-validation, there is a process of applying the NB algorithm as shown in Figure 4.

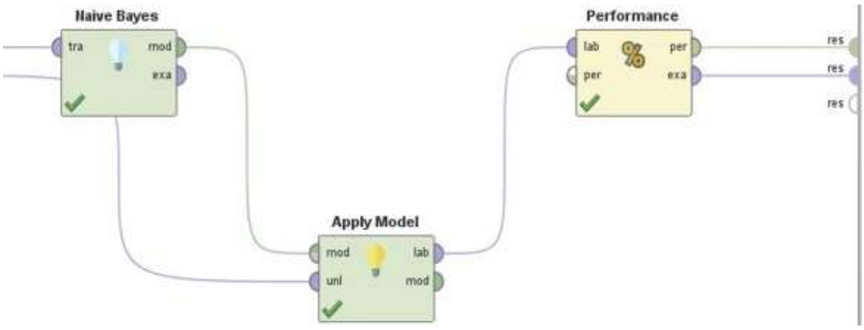

Figure 4. Use of the NB model

Testing the accuracy of this expert system uses the confusion matrix method [36]. Confusion matrix is a table that states the classification of the number of correct test data and the number of incorrect test data [42]. System accuracy calculations are carried out based on simulation data from experts [43]. The simulation data was compared with the results of the NB calculation method. This accuracy test was carried out with two scenarios. The first scenario is that the system displays only one disease with the highest total probability. Although the symptoms included can result in more than one disease. The results of the first scenario can be seen in Figure 4. The second scenario is that the system displays all diseases that can occur based on the symptoms entered. Through the second scenario which is implemented on 208 data, the data obtained from the confusion matrix results can determine the probability or opportunity using the NB Theorem as follows: the TP value is 181, the FP value is 18, the TN value is 42 and the FN value is 27. The calculation of values accuracy, which includes accuracy, precision, and recall values, is as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Accuracy = \frac{121+18}{121+18+42+27} = \frac{199}{208} = 0{,}9567$$

$$Precision = \frac{TP}{TP+TN} \quad (3)$$

$$Precision = \frac{181}{181+18} = \frac{181}{191} = 0{,}9476$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$Recall = \frac{181}{181+27} = \frac{181}{208} = 0{,}8702$$

The results of the Bayes theorem calculation show an accuracy value of 0.9567%, the precision value which is a metric that shows the proportion of positive predictions that are positive is 0.9476%, and the recall value which is the proportion of actual positive data that is predicted positively by the model is 0.8702%. From these results, it can be concluded that the NB algorithm can be used to diagnose ENT diseases because the accuracy value is quite high, namely 95.67%.

4. CONCLUSION

Based on the results of data mining research for diagnosing ENT diseases using the NB algorithm on patient medical record datasets with experts being ENT specialist doctors with KDD, the accuracy was tested using 208 data. From the tests that have been carried out, the NB algorithm can be used to diagnose ENT diseases. The test results obtained an accuracy of 95.67%, a precision of 94.76%, and a recall of 87.02%. From these results, it can be concluded that the NB algorithm can be used to diagnose ENT diseases because the accuracy value is quite high. Using NB with feature selection using a genetic algorithm can provide better accuracy results compared to using NB without feature selection. This is due to the genetic algorithm's ability to find the most relevant and significant feature combinations, thereby improving the performance of the classification model. This proves that the NB method can be used to build an ENT diagnosis classification model and the model performance is evaluated using accuracy, precision, and recall metrics.

ACKNOWLEDGEMENTS

This research received funding from the Director's Decree No. 268/PL43/HK.03.01/2024 through Cilacap State Polytechnic DIPA funds for the 2024 Lecturer Stimulus Research scheme with contract number 175//PL43/HK.07/2024, through the Cilacap State Polytechnic Research and Community Service Center.

REFERENCES

[1] A. Kaspar and S. Pifeleti, "A call to action for the inclusion of ENT/Audiology services in the public health approach to addressing non-communicable diseases in the Pacific Islands," *Public Health in Practice*, vol. 2, no. April, p. 100123, 2021, doi: 10.1016/j.puhip.2021.100123.

[2] S. Torretta *et al.*, "ENT management of children with adenotonsillar disease during COVID-19 pandemic. Ready to start again?," *International Journal of Pediatric Otorhinolaryngology*, vol. 138, no. May, 2020, doi: 10.1016/j.ijporl.2020.110145.

[3] L. Torres-García, R. M. Acosta, A. C. Martínez, I. I. Alcañiz, A. A. Velazquez, and M. A. Carceller, "Evolution in the incidence of infectious diseases in the pediatric ENT area during the COVID-19 pandemic," *Acta Otorrinolaringologica (English Edition)*, vol. 74, no. 4, pp. 232–238, 2023, doi: 10.1016/j.otoeng.2022.11.007.

[4] R. Cohen *et al.*, "Antimicrobial treatment of ENT infections," *Infectious Diseases Now*, vol. 53, no. 8, p. 104785, 2023, doi: 10.1016/j.idnow.2023.104785.

[5] I. A. Danevska, T. Jakjovska, S. Momchilovikj, and E. Gjinovska-Tasevska, "P174 Ear, nose and throat (ENT) disorders in cystic fibrosis patients in the cystic fibrosis centre at the institute for respiratory diseases in children in Skopje, Republic of North Macedonia," *Journal of Cystic Fibrosis*, vol. 19, p. S105, 2020, doi: 10.1016/s1569-1993(20)30509-9.

[6] D. Lu, H. Wang, R. Yu, H. Yang, and Y. Zhao, "Integrated infection control strategy to minimize nosocomial infection of coronavirus disease 2019 among ENT healthcare workers," *Journal of Hospital Infection*, vol. 104, no. 4, pp. 454–455, 2020, doi: 10.1016/j.jhin.2020.02.018.

[7] W. A. Massawe *et al.*, "Laryngopharyngeal reflux disease, prevalence and clinical characteristics in ENT department of a tertiary hospital Tanzania," *World Journal of Otorhinolaryngology - Head and Neck Surgery*, vol. 7, no. 1, pp. 28–33, 2021, doi: 10.1016/j.wjorl.2020.04.009.

[8] C. Eyermann, T. Raguin, D. Rohmer, E. Noel, and A. Charpiot, "Cochleovestibular manifestations in Fabry disease: Importance of screening and systematic ENT evaluation," *European Annals of Otorhinolaryngology, Head and Neck Diseases*, vol. 136, no. 4, pp. 273–279, 2019, doi: 10.1016/j.anorl.2019.04.014.

[9] B. K. Nuhu, I. Aliyu, M. A. Adegboye, J. K. Ryu, O. M. Olaniyi, and C. G. Lim, "Distributed network-based structural health monitoring expert system," *Building Research and Information*, vol. 49, no. 1, pp. 144–159, 2021, doi: 10.1080/09613218.2020.1854083.

[10] L. P. Wanti, O. Somantri, N. W. A. Prasetya, and L. Puspitasari, "Fuzzy expert system design for detecting stunting," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 34, no. 1, pp. 556–564, 2024, doi: 10.11591/ijeecs.v34.i1.pp556-564.

[11] A. Heiß, D. S. Paraforos, G. M. Sharipov, and H. W. Griepentrog, "Modeling and simulation of a multi-parametric fuzzy expert system for variable rate nitrogen application," *Computers and Electronics in Agriculture*, vol. 182, no. February, 2021, doi: 10.1016/j.compag.2021.106008.

[12] I. K. Mujawar and B. T. Jadhav, "Web-based fuzzy expert system for diabetes diagnosis," *International Journal of Computer Sciences and Engineering*, vol. 7, no. 2, pp. 995–1000, 2019, doi: 10.26438/ijcse/v7i2.9951000.

[13] R. Nourian, S. M. Mousavi, and S. Raissi, "A fuzzy expert system for mitigation of risks and effective control of gas pressure reduction stations with a real application," *Journal of Loss Prevention in the Process Industries*, vol. 59, pp. 77–90, 2019, doi: 10.1016/j.jlp.2019.03.003.

[14] L. P. Wanti and L. Puspitasari, "Optimization of the fuzzy logic method for autism spectrum disorder diagnosis," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 1, pp. 16–24, 2022, doi: 10.29207/resti.v6i1.3599.

[15] I. Bahroni, L. P. Wanti, N. W. Rahadi, A. A. Hartono, and R. Purwanto, "Implementation of forward chaining for diagnosis of dengue hemorrhagic fever," *Journal of Innovation Information Technology and Application (JINITA)*, vol. 4, no. 1, pp. 32–42, 2022, doi: 10.35970/jinita.v4i1.1204.

[16] O. Somantri, R. H. Maharrani, and L. P. Wanti, "An optimize weights Naïve Bayes model for early detection of diabetes," *Telematika*, vol. 15, no. 1, pp. 14–22, 2022, doi: 10.35671/telematika.v15i1.1307.

[17] R. Kusumaningrum, T. A. Indihatmoko, S. R. Juwita, A. F. Hanifah, K. Khadijah, and B. Surarso, "Benchmarking of multi-class algorithms for classifying documents related to stunting," *Applied Sciences (Switzerland)*, vol. 10, no. 23, pp. 1–13, 2020, doi: 10.3390/app10238621.

[18] S. Wang, J. Ren, and R. Bai, "A semi-supervised adaptive discriminative discretization method improving discrimination power of regularized naive Bayes," *Expert Systems with Applications*, vol. 225, no. November 2022, p. 120094, 2023, doi: 10.1016/j.eswa.2023.120094.

[19] A. Saibene, M. Assale, and M. Giltri, "Expert systems: definitions, advantages and issues in medical field applications," *Expert Systems with Applications*, vol. 177, no. November 2020, p. 114900, 2021, doi: 10.1016/j.eswa.2021.114900.

[20] O. Alhabashneh, "Fuzzy-based adaptive framework for module advising expert system," *Annals of Emerging Technologies in Computing*, vol. 5, no. 1, pp. 13–27, 2021, doi: 10.33166/AETiC.2021.01.002.

[21] H. T. Phan, N. T. Nguyen, V. C. Tran, and D. Hwang, "An approach for a decision-making support system based on measuring the user satisfaction level on Twitter," *Information Sciences*, vol. 561, pp. 243–273, 2021, doi: 10.1016/j.ins.2021.01.008.

[22] A. P. Wibawa *et al.*, "Naïve Bayes classifier for journal quartile classification," *International Journal of Recent Contributions from Engineering, Science & IT (iJES)*, vol. 7, no. 2, p. 91, 2019, doi: 10.3991/ijes.v7i2.10659.

[23] T. Wahyuningsih, D. Manongga, I. Sembiring, and S. Wijono, "Comparison of effectiveness of logistic regression, naive bayes, and random forest algorithms in predicting student arguments," *Procedia Computer Science*, vol. 234, pp. 349–356, 2024, doi: 10.1016/j.procs.2024.03.014.

[24] E. R. Arumi, Sumarno Adi Subrata, and Anisa Rahmawati, "Implementation of Naïve bayes method for predictor prevalence level for malnutrition toddlers in Magelang city," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 2, pp. 201–207, 2023, doi: 10.29207/resti.v7i2.4438.

[25] J. O. Onah, S. M. Abdulhamid, M. Abdullahi, I. H. Hassan, and A. Al-Ghusham, "Genetic algorithm based feature selection and Naïve Bayes for anomaly detection in fog computing environment," *Machine Learning with Applications*, vol. 6, no. April, p. 100156, 2021, doi: 10.1016/j.mlwa.2021.100156.

[26] Y. Saputra, N. Khoirunisa, and S. Arinal Haqq, "Classification of health and nutritional status of toddlers using the Naïve Bayes classification," *CoreID Journal*, vol. 1, no. 2, pp. 49–57, 2023, doi: 10.60005/coreid.v1i2.8.

[27] O. Peretz, M. Koren, and O. Koren, "Naive Bayes classifier – an ensemble procedure for recall and precision enrichment," *Engineering Applications of Artificial Intelligence*, vol. 136, no. PB, p. 108972, 2024, doi: 10.1016/j.engappai.2024.108972.

[28] C. J. Anderson *et al.*, "A novel naïve Bayes approach to identifying grooming behaviors in the force-plate actometric platform," *Journal of Neuroscience Methods*, vol. 403, no. July 2023, p. 110026, 2024, doi: 10.1016/j.jneumeth.2023.110026.

[29] W. Guo, G. Wang, C. Wang, and Y. Wang, "Distribution network topology identification based on gradient boosting decision tree and attribute weighted naive Bayes," *Energy Reports*, vol. 9, pp. 727–736, 2023, doi: 10.1016/j.egyr.2023.04.256.

[30] S. Wang, J. Ren, R. Bai, Y. Yao, and X. Jiang, "A max-relevance-min-divergence criterion for data discretization with applications on Naive Bayes," *Pattern Recognition*, vol. 149, no. September 2023, p. 110236, 2024, doi: 10.1016/j.patcog.2023.110236.

[31] M. Vishwakarma and N. Kesswani, "A new two-phase intrusion detection system with Naïve Bayes machine learning for data classification and elliptic envelop method for anomaly detection," *Decision Analytics Journal*, vol. 7, no. April, p. 100233, 2023, doi: 10.1016/j.dajour.2023.100233.

[32] M. F. A. Saputra, T. Widiyaningtyas, and A. P. Wibawa, "Illiteracy classification using K means-naïve bayes algorithm," *International Journal on Informatics Visualization*, vol. 2, no. 3, pp. 153–158, 2018, doi: 10.30630/joiv.2.3.129.

[33] W. A. van Eeden *et al.*, "Predicting the 9-year course of mood and anxiety disorders with automated machine learning: a comparison between auto-sklearn, naïve Bayes classifier, and traditional logistic regression," *Psychiatry Research*, vol. 299, p. 113823, May 2021, doi: 10.1016/j.psychres.2021.113823.

[34] A. Rozaq and A. J. Purnomo, "Classification of stunting status in toddlers using naive bayes method in the city of Madiun based on Website," *Jurnal Techno Nusa Mandiri*, vol. 19, no. 2, pp. 69–76, 2022, doi: 10.33480/techno.v19i2.3337.

[35] A. Tariq *et al.*, "Modelling, mapping and monitoring of forest cover changes, using support vector machine, kernel logistic regression and naive bayes tree models with optical remote sensing data," *Heliyon*, vol. 9, no. 2, p. e13212, 2023, doi: 10.1016/j.heliyon.2023.e13212.

[36] V. Balakrishnan and W. Kaur, "String-based multinomial naïve bayes for emotion detection among facebook diabetes community," *Procedia Computer Science*, vol. 159, pp. 30–37, 2019, doi: 10.1016/j.procs.2019.09.157.

[37] A. Saleh and F. Nasari, "Application of equal-width interval discretization in naive bayes method to improve the accuracy of student major selection prediction (in *Indonesian*)," *Masyarakat Telematika Dan Informasi : Jurnal Penelitian Teknologi Informasi dan Komunikasi*, vol. 9, no. 1, p. 1, 2018, doi: 10.17933/mti.v9i1.113.

[38] Y. Shang, "Prevention and detection of DDOS attack in virtual cloud computing environment using Naive Bayes algorithm of machine learning," *Measurement: Sensors*, vol. 31, no. July 2023, p. 100991, 2024, doi: 10.1016/j.measen.2023.100991.

[39] A. Ridwan and T. N. Sari, "The comparison of accuracy between naïve bayes clasifier and c4.5 algorithm in classifying toddler nutrition status based on anthropometry index," *Journal of Physics: Conference Series*, vol. 1764, no. 1, 2021, doi: 10.1088/1742-6596/1764/1/012047.

[40] T. Olsson, M. Ericsson, and A. Wingkvist, "To automatically map source code entities to architectural modules with Naive Bayes," *Journal of Systems and Software*, vol. 183, p. 111095, 2022, doi: 10.1016/j.jss.2021.111095.

[41] M. Libnao, M. Misula, C. Andres, J. Mariñas, and A. Fabregas, "Traffic incident prediction and classification system using naïve bayes algorithm," *Procedia Computer Science*, vol. 227, pp. 316–325, 2023, doi: 10.1016/j.procs.2023.10.530.

[42] S. Farhana, "Classification of academic performance for university research evaluation by implementing modified naive bayes algorithm," *Procedia Computer Science*, vol. 194, pp. 224–228, 2021, doi: 10.1016/j.procs.2021.10.077.

[43] S. Shastri *et al.*, "Development of a data mining based model for classification of child immunization data," *International Journal of Computational Engineering Research*, vol. 8, no. 6, pp. 41–49, 2018, [Online]. Available: www.ijceronline.com.

BIOGRAPHIES OF AUTHORS

Linda Perdana Wanti was born in Banyumas, October 10, 1988. Currently working as a lecturer at the Cilacap State Polytechnic since 2019. She is scientist and lecturer at college of the Department of Cyber Security Engineering, State Polytechnic of Cilacap, Cilacap Regency, Indonesia. She received the Master of Computer Science (M.Kom.) degree in Informatics Engineering from Universitas Amikom Yoyakarta in 2013. Her research interests include machine learning for decission support system, database and expert system. She can be contacted at email: linda_perdana@pnc.ac.id.

Nur Wachid Adi Prasetya was born in Wonogiri, November 15, 1988. Currently working as a lecturer at the Cilacap State Polytechnic since 2019. He graduated from Informatics Engineering at UIN Sunan Kalijaga Yogyakarta in 2011 and earned a Master of Computer Science (M. Kom) from the Master of Information Engineering at the University of Islam Indonesia in 2018. He has research interests in the field of image processing and information systems design. He can be contacted at email: nwap.pnc@pnc.ac.id.

Ihza Awaludin is a student of D4 Cyber Security Engineering at the Cilacap State Polytechnic. He can be contacted at email: ihzaawaludin@gmail.com.

Muhammad Bintang Aditya Saputra is a student of D4 Cyber Security Engineering at the Cilacap State Polytechnic. He can be contacted at email: as7171110@gmail.com.

Syamaidzar Nadifa Furi is a student of D4 Cyber Security Engineering at the Cilacap State Polytechnic. He can be contacted at email: Syamaidzarnf@gmail.com.

Dimas Maulana Dwi Kumara is a student of D4 Cyber Security Engineering at the Cilacap State Polytechnic. He can be contacted at email: dimaskumara69@gmail.com.