ICIC Express Letters
Volume **18**, Number **1**, January **2024**

ICIC International ©2024 ISSN 1881-803X
pp. **97–104**

MACHINE LEARNING-BASED AUTHORSHIP ATTRIBUTION TOWARDS CITIZEN JOURNALISM ARTICLES IN THE INDONESIAN LANGUAGE: A PRELIMINARY STUDY

Retno Kusumaningrum[1,*], Selvi Fitria Khoerunnisa[2], Aflah Tazakka[1]
Khadijah[1] and Prajanto Wahyu Adi[1]

[1]Department of Informatics
Universitas Diponegoro
Jl. Prof. Soedarto, SH, Tembalang, Semarang 50275, Indonesia
aflahtazakka@students.undip.ac.id; {khadijah; prajanto}@live.undip.ac.id
*Corresponding author: retno@live.undip.ac.id

[2]School of Postgraduates Study
Universitas Diponegoro
Jl. Imam Bardjo SH No. 5, Semarang 50241, Indonesia
selvifitria@student.undip.ac.id

Received January 2023; accepted April 2023

Abstract. *Author attribution is an important application area of natural language that can be implemented as an application for plagiarism detection. Despite the success of various methods applied to solving the problem of authorship attribution towards various corpus in various languages, solving the problem of authorship attribution in the Indonesian corpus is still a big challenge. This problem is due to the stiffness of finding research on authorship attribution for Indonesian language documents and the availability of benchmark datasets for this task. Therefore, this study compares various classical machine learning algorithms, such as logistic regression, support vector machine, Naïve Bayes, random forest, and gradient boosting. This study was applied to the Indonesian language corpus crawled from kompasiana.com as an example of citizen journalism. In addition, we implemented term frequency-inverse document frequency as a feature and experimented by using three different ratios of training and testing data (90 : 10, 80 : 20, and 70 : 30) to improve model performance. The results show that a ratio of 90 : 10 has the best average accuracy of 0.86. Regarding the classifier used, it shows that SVM has the best average accuracy for all ratios with a value of 0.93, while the logistic regression method, Naïve Bayes, random forest, and gradient boosting, have average accuracy respectively of 0.87, 0.77, 0.83, and 0.85.*

Keywords: Author attribution, Classical machine learning, Support vector machine, Term frequency-inverse document frequency

1. **Introduction.** Plagiarism is taking someone else's work and making it appear as if it were their work without including the source. In today's digital era, plagiarism is a common condition due to the ease of accessing data and information online. It is easier for everyone to copy and paste other people's works. Furthermore, copying and pasting seem normal and fair, so many people are unaware that they have committed plagiarism. Therefore, plagiarism becomes very detrimental to the creator of a work and is also dangerous for the generation that usually does this copy-paste activity. Furthermore, it can have an impact in the form of a decrease in creativity and critical thinking skills.

On the other hand, the development of Internet technology and freedom of the press means that more and more citizens are playing an active role in reporting information and news without needing journalism qualifications and education. This form is usually

DOI: 10.24507/icicel.18.01.97

referred to as citizen journalism. The process of publishing citizen journalism is generally carried out without checking by the publisher's editor, so the problems that usually arise lie in credibility, information accuracy, and partisanship [1]. This condition certainly increases the chances of plagiarism being found in citizen journalism.

One of the tasks that can be applied as a solution to the plagiarism detection process for citizen journalism is to perform authorship attribution [2]. It is also known as authorship or author identification [3]. In addition, authorship attribution can also be used to determine a literary work or document whose author is unknown [4]. Various methods have been applied to solving the problem of authorship attribution. In addition, research on authorship attribution has also been developed for various language domains.

The first research is an authorship identification for Japanese Twitter users [5]. Another research is an authorship attribution for news articles in the Arabic language [6]. Both of those researches were conducted based on the stylometric approach, which focuses on quantitative analysis of the individuality of the author's style and technique. The first research employed a combination of character-n-gram frequency ($n = 1$, 2, and 3) as stylometric features and subsequently implemented cosine similarity for similarity ranking [5]. Since it uses cosine similarity ranking, it will work better if the stylometric features are represented as a vector focusing on orientation judgment, not magnitude. The second research investigates three different stemming methods in stylometric authorship attribution for Arabic and implements Ward linkage and Euclidean distance as a clustering method [6]. In contrast to cosine similarity, the Euclidean distance will work better when the stylometric features are represented as a vector focusing on magnitude judgment, not orientation. Furthermore, various studies have been investigated for authorship attribution for English corpus, such as Reuters Corpus Volume I [2,4,7,8], BBC News Dataset [4], IMDb62 dataset [2,9], Twitter dataset [9], ISOT database [10], Social Media Forensics database [10], PAN 2012 dataset [11], and Judgment dataset (i.e., consists of legal judgments from three Australian High Court judges) [2].

Despite the success of various methods applied to solving the problem of authorship attribution towards various corpus in various languages, solving the problem of authorship attribution in the Indonesian corpus is still a big challenge. This problem is due to the stiffness of finding research on authorship attribution for Indonesian language documents and the availability of benchmark datasets for this task. In addition, the success of applying a method in a particular language will sometimes give different results when applied in another language [6], in this case, the Indonesian language. The Indonesian language differs from English because of differences in language roots, i.e., Indonesian comes from Austronesian, while English comes from Germany. Therefore, this research aims to address these literature gaps by applying various classical machine learning algorithms to solving the authorship attribution problem for Indonesian-language corpus.

Because authorship attribution is a multi-class classification task, this study compares various classical machine learning algorithms widely applied to solving multi-class classification tasks. As preliminary studies, those algorithms are logistic regression, support vector machine (SVM), Naïve Bayes, random forest, and gradient boosting. This research was applied to the Indonesian language corpus crawled from kompasiana.com, where Kompasiana curated news from various authors. It is commonly referred to as user-generated content or another famous name, citizen journalism.

The rest of this paper is organized as follows. Section 2 describes the detailed methodology, including data collection, preprocessing, TF-IDF feature extraction, classification model generation, and evaluation. Section 3 gives experimental results and their analysis, followed by Section 4, in which conclusions are drawn.

2. **Methodology.** Several steps are employed in this research, as illustrated in Figure 1. The first step is data collection, implemented using a web crawling process. The collected data are subsequently fed into the text preprocessing step. The third step is the feature extraction process. Generating a classification model is the fourth step. In this study, we implement five algorithms: logistic regression, support vector machine (SVM), Naïve Bayes, random forest, and gradient boosting. The results are further evaluated to determine the best model.

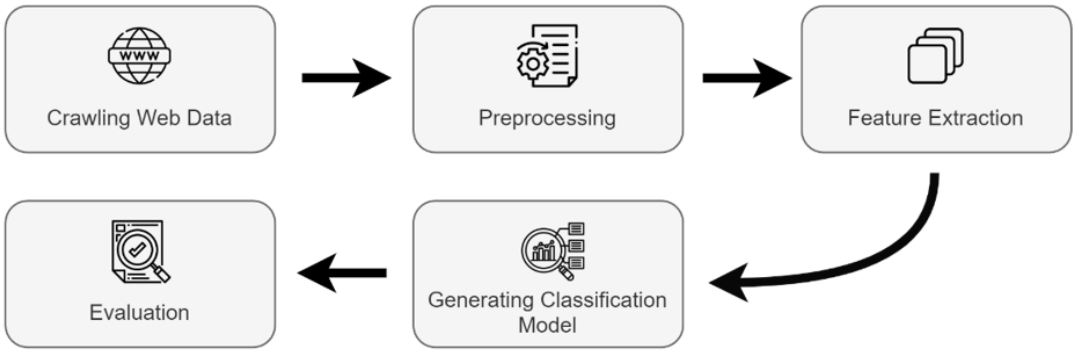


FIGURE 1. Research methodology

2.1. **Data collection.** This research used blog posts and articles from an Indonesian website (kompasiana.com) with content spanning 2018-2022. A total of 7,680 blog posts from various categories were retrieved, written by 15 randomly chosen valid authors (512 posts each). Note that site management data were used to ascertain the ground truth text-to-author configuration, which was used to evaluate the prediction accuracy of the various machine learning models tested.

2.2. **Preprocessing.** Preprocessing was conducted to remove noise and prepare the data for robust machine learning. Data obtained during the web crawling is preprocessed according to case folding, filtering, stop words removal, and tokenization stages.

Case folding was performed, changing random capital letters to lower-case forms and vice versa. Subsequently, words were ASCII encoded, unnecessary characters were discarded, and the words were recorded to UTF-8. Characters, words, and phrases unique to kompasiana.com and website presentation were removed, including URLs, ads, and metadata. This step is called filtering. Unnecessary stop words were also removed to reduce computing overhead. Lastly, the remaining dataset was tokenized to save space. Data volume reduction is known to improve performance, and the methods applied have already been determined to avoid biasing NLP task results [12].

2.3. **TF-IDF feature extraction.** We converted every article or post to a feature vector at this stage. Feature extraction and selection applied TF-IDF vectorization with a unigram model whose weight value per word was based on its frequency in the text. TF-IDF is commonly used for text mining and NLP [13]. TF-IDF was chosen because it can evaluate the relationship of each word in a set of documents [14]. This method can extract the essential words in the document and improve the search. Another advantage of TF-IDF is that it is easy, lightweight, and does not require heavy computing.

2.4. **Generating classification model.** The vector generated at the feature extraction stage is used for classification. The classification method chosen was logistic regression, SVM, Naïve Bayes, random forest, and gradient boosting, as explained in the following sub-subsections.

2.4.1. *Logistic regression.* The logistic regression technique learns the connection probabilities between independent (x) and dependent (y) variables [15]. This method will classify data into discrete classes. In this study, logistic regression will calculate the probability of an article being included in an existing category or list. Then, calculating the logistic regression uses the sigmoid function, which converts anything into a range of 0 and 1. Logistic regression is represented by an equation [16]:

$$p = \frac{e^{\left(\alpha+\sum_{i=1}^{N}\beta_i x_i\right)}}{1+e^{e^{\left(\alpha+\sum_{i=1}^{N}\beta_i x_i\right)}}} \tag{1}$$

where $\vec{x} = (x_1, x_2, \ldots, x_N)$ is feature vector of instance data, p is the predicted output probabilities, α is the y intercept, β_i is the regression coefficient, and $e = 2.71828$ is the base of the system of natural logarithms.

Subsequently, we calculate the logits or log odds to determine the predicted class as follows:

$$\log(odds) = \ln\left(\frac{p}{1-p}\right) \tag{2}$$

When the p equals 0.8 for binary classification (A and B), and the *odds* value equals 4. It means the probability that class is class A will be greater than class B. Since we implement logistic regression for multi-class classification tasks, several independent binary logistic regression models will be built, for example, when we have three classes (A, B, and C), the first model was built to separate class A and the rest of the class (classes B and C), and the subsequent model was built to separate classes B and C.

2.4.2. *Support vector machine.* SVM is one of the well-known classification algorithms that applies statistical learning to sample training data. In this case, text data were used for training [17]. SVM uses a simple mathematical model, as in $wx' + \gamma = 0$, to enable the division of linear domains [18]. SVM can handle linear and non-linear problems. This study uses linear SVM. Although linear SVM is generally intended for binary classification, multi-class classification can be handled using the One-to-Rest or One-to-One approaches. In this study, we implemented the One-to-Rest approach.

2.4.3. *Naïve Bayes.* Naïve Bayes classification applies a statistical method of predicting prospective members of categorical groups [19]. This method uses a simple probabilistic classifier that computes probabilities by counting the frequencies and combinations of values in a given dataset. Applying the Bayes concept, which assumes all independent variables and only considers the value of each class variable, causes the classification in this study only to consider the features of each class [20]. Even though it is called naive because it rarely applies in the real world, classification using Naïve Bayes can learn quickly. Equation $P(A|B) = \frac{P(A)P(B|A)}{P(B)}$ is used to determine the conditional probability. $P(A|B)$ is the probability of the occurrence of event A when event B occurs, $P(A)$ is the probability of the occurrence of A, $P(B|A)$ is the probability of the occurrence of event B when event A occurs, and $P(B)$ is the probability of the occurrence of B.

2.4.4. *Random forest.* The random forest model applies an ensemble learning technique to constructively regressing options using decision trees during training. A decision tree is used as a base in this development method; the random forest will build many trees, and then the best feature will be randomly selected. That is why it is called random [21]. Every tree in the forest will be used for classifying new objects from the input. Random forest will choose the classification with the most votes from each tree. The classes identified by most trees are then chosen [19]. Random forest can be described with a mathematical model $n_{ij} = w_i C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)}$. n_{ij} is the importance of node j, w_i is weighted number of samples reaching node j, C_j is the impurity value of node j, $left(j)$ is

the child node from the left split on node j, and $right(j)$ is the child node from the right split on node j. Random forest is considered to be able to handle high-dimensional data such as text. A random forest runs with subsets of data. Then the random forest is also considered to work faster because it only works on some of the features in the model.

2.4.5. *Gradient boosting.* The boosting algorithm has the main characteristic of converting weak learners into solid and robust classifiers. One of the well-known boosting families is gradient boosting. Gradient boosting [22] uses a gradient descent learning model that minimizes the calculated loss to improve accuracy. The gradient boosting learning process continuously adapts new models to provide more accurate estimates of class variables [23]. Each new model tends to correlate with and minimize the negative slope of the system's loss function. Although boosting can be done by adding a new model to the acid sequentially and in each iteration, the model will learn from the mistakes learned from all the sources, decreasing the gradient makes classification more effective.

2.5. **Evaluation.** The model that has been formed is then essential to check whether the model is working correctly. It is because machine learning aims to learn patterns that generalize well to unseen data and not just memorize the data seen during training. The model performance was evaluated using accuracy. In short, accuracy is the fraction of predictions our model got right.

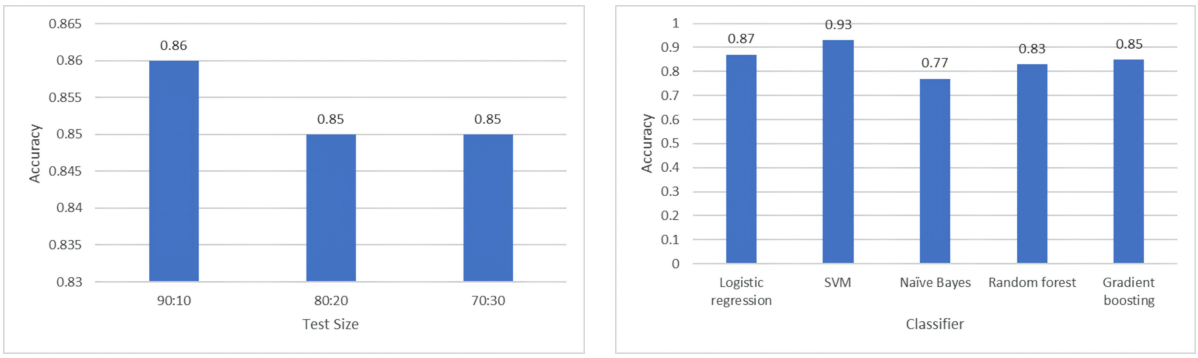
3. **Result and Discussion.** The model will be formed based on a predetermined classifier. The classifier consists of logistic regression, SVM, Naïve Bayes, random forest, and gradient boosting. Then it will be evaluated and compared to produce the best model. Typically, researchers test hyperparameter values based on the sizes of the training and testing datasets. Nguyen et al. [24] showed that training and testing dataset size combinations could significantly affect model performance, even improving it. Hence, we experimented with three different splitting ratios (90 : 10, 80 : 20, 70 : 30). Ratio 90 : 10 which means 90% of the data is for training and 10% for testing. The same intent applies to 80 : 20 and 70 : 30. Table 1 shows the evaluation results of each machine learning model and training-testing ratio.

Table 1. Comparison of evaluation results

	Training-testing dataset ratios			
	90 : 10	80 : 20	70 : 30	**Average**
Logistic regression	0.88	0.87	0.87	0.87
SVM	0.94	0.93	0.93	0.93
Naïve Bayes	0.77	0.77	0.76	0.77
Random forest	0.84	0.83	0.83	0.83
Gradient boosting	0.86	0.86	0.84	0.85
Average	0.86	0.85	0.85	

Based on Table 1, we can calculate the average of each splitting ratio for all methods. It serves to find out the best comparison for the data. The average comparison of each splitting ratio is illustrated in Figure 2(a). The average for each classification method is compared to determine which is best for all ratios. The average comparison for each method is illustrated in Figure 2(b).

Figure 2(a) shows that the splitting ratio with a value of 90 : 10 has the highest accuracy of 0.86. It is compared to the splitting ratio of 80 : 20 and 70 : 30, with both accuracy values of 0.85. It shows that this dataset's accuracy is directly proportional to the data training and the decrease in data testing. Generally, the 80 : 20 ratio is popularly used in dividing training and test data. A more extensive training data ratio causes more data



(a) Per test size (b) Per method

FIGURE 2. Graph of evaluation results

variations to be trained, so the model can learn the data distribution better and improve model performance. However, these conditions still depend on the characteristics of the data being trained. As has been explained, the research dataset is a dataset of articles from citizen journalism with a total of 15 classes, so the multi-class classification task that must be handled in this study is very complex. This problem must be balanced with a large number of data samples. Therefore, a 90 : 10 division is the most appropriate ratio.

Based on Figure 2(b), the SVM performs best compared to all other methods. The average accuracy value produced by SVM reaches 0.93. This study aligns with the previous study by Zhou et al. [25], which shows that SVM outperformed the other classical machine learning models, such as Naïve Bayes and random forest, for poetry Chinese authorship attribution. This result is due to the ability of the SVM model to handle both categorical and continuous variables. Previously defined TF-IDF converts a set of words into a vector. TF-IDF generates continuous values that can be handled with SVM. SVM divides data items into classes to fix the hyperplane and maximize marginal value. It causes the data to correlate better, even when the values are well outside the class mean. In particular, textual data is highly similar, even when written by different authors.

Nonetheless, SVM handles this task without issue. Even using linear SVM, multi-class assignments can be handled by combining several binary classifications. Moreover, SVM can use higher dimensional spaces with certain mapping functions that allow non-linear correlations.

Logistic regression, gradient enhancement, and random forest methods also show good accuracy values. Logistic regression does not require an assumption of normality between the independent variables; hence, there is no need to test this assumption. Improved gradients are good at handling data with complex patterns, which are common in textual data. The random forest method uses recursive decision trees for feature selection and is good at clarifying data with incomplete attributes or noisy features. Thus, it can also deal with non-linear problems. However, this method requires configuration at an early stage, and if the selected hyperparameter combination is correct, the resulting prediction will be optimized.

This research uses multinomial Naïve Bayes, which can handle multi-class classification problems. Naïve Bayes has the lowest accuracy value because it relies on probabilistic correlation data to make predictions. Therefore, multicollinearity can hinder good performance. Several types of Naïve Bayes are commonly used in classification. Multinomial Naïve Bayes uses Bayesian theory as a basis for classification by calculating the probability distribution of each class. However, unfortunately, multinomial Naïve Bayes can only work well with discrete data but do not work with continuous data, and it is known that the TF-IDF used produces continuous data.

4. **Conclusions.** In this study, we experimented with various machine-learning methods to match 15 authors with their respective text sections based on collected blog data. The collected data were preprocessed, and the TF-IDF feature vectorization was used for extraction and matching. This study compares the logistic regression model, SVM, Naïve Bayes, random forest, and gradient boosting. After preprocessing and vectorization, the data is classified by several methods (i.e., logistic regression, SVM, Naïve Bayes, random forest, and gradient boosting) to find the best model. The search for the best model also uses different dataset splitting. The results show that the splitting ratio of 90 : 10 has the highest accuracy value, which explains that the increase in training data and the decrease in data testing are directly proportional to the accuracy. Subsequently, the SVM method has the highest accuracy, with an average of 0.93, because SVM can handle categorical and continuous variables well and then divide the data items into classes to improve the hyperplane and maximize the marginal value. It causes better-correlated data.

As future research work, the deep learning model can be developed to generate an authorship attribution model. Our current model is limited only to the classical machine learning model. Generally, the Recurrent Neural Network (RNN) is outperformed in a task related to sequential data, such as citizen journalism articles. We want to experimentally test the performance of several architectures in RNN, such as Long Short-Term Memory (LSTM), bidirectional LSTM, LSTM with attention mechanism, and Gated Recurrent Unit (GRU). Another proposed research area uses different word vectorization techniques, such as Word2Vec or Bidirectional Encoder Representations from Transformers (BERT). Both Word2Vec and BERT use dense vector representation so that it is expected to be more able to understand the semantics of the word. Whereas this study still employed the TF-IDF.

Acknowledgment. This work is supported by Faculty of Science and Mathematics, Universitas Diponegoro under the research grant number 1264E/UN7.5.8/PP/2022. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] U. K. H. Ecker et al., The psychological drivers of misinformation belief and its resistance to correction, *Nat. Rev. Psychol.*, vol.1, no.1, pp.13-29, DOI: 10.1038/s44159-021-00006-y, 2022.

[2] Y. Sari, M. Stevenson and A. Vlachos, Topic or style? Exploring the most useful features for authorship attribution, *Proc. of the 27th International Conference on Computational Linguistics*, pp.343-353, 2018.

[3] A. Abbasi, A. R. Javed, F. Iqbal, Z. Jalil, T. R. Gadekallu and N. Kryvinska, Authorship identification using ensemble learning, *Sci. Rep.*, vol.12, no.1, pp.1-16, DOI: 10.1038/s41598-022-13690-4, 2022.

[4] S. T. P. Gupta, J. K. Sahoo and R. K. Roul, Authorship identification using recurrent neural networks, *ACM International Conference Proceeding Series*, pp.133-137, DOI: 10.1145/3325917.3325935, 2019.

[5] S. Okuno, H. Asai and H. Yamana, A challenge of authorship identification for ten-thousand-scale microblog users, *Proc. of the IEEE International Conference on Big Data*, pp.52-54, 2014.

[6] A. Omar and W. I. Hamouda, The effectiveness of stemming in the stylometric authorship attribution in Arabic, *International Journal of Advanced Computer Science and Applications*, vol.11, no.1, pp.116-121, DOI: 10.14569/ijacsa.2020.0110114, 2020.

[7] R. R. Iyer and C. P. Rose, A machine learning framework for authorship identification from texts, *arXiv Preprint*, arXiv: 1912.10204, 2019.

[8] A. M. Mohsen, N. M. El-Makky and N. Ghanem, Author identification using deep learning, *Proc. of 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA2016)*, Anaheim, CA, USA, pp.898-903, DOI: 10.1109/ICMLA.2016.45, 2016.

[9] O. Fourkioti, S. Symeonidis and A. Arampatzis, Language models and fusion for authorship attribution, *Inf. Process. Manag.*, vol.56, no.6, DOI: 10.1016/j.ipm.2019.102061, 2019.

[10] F. Alonso-Fernandez, N. M. S. Belvisi, K. Hernandez-Diaz, N. Muhammad and J. Bigun, Writer identification using microblogging texts for social media forensics, *IEEE Trans. Biom. Behav. Identity Sci.*, vol.3, no.3, pp.405-426, DOI: 10.1109/TBIOM.2021.3078073, 2021.

[11] N. E. Benzebouchi, N. Azizi, N. E. Hammami, D. Schwab, M. C. E. Khelaifia and M. Aldwairi, Authors' writing styles based authorship identification system using the text representation vector, *Proc. of the 16th International Multi-Conference on Systems, Signals & Devices*, pp.371-376, 2019.

[12] I. M. Rabbimov and S. S. Kobilov, Multi-class text classification of Uzbek news articles using machine learning, *Journal of Physics: Conference Series*, vol.1546, no.1, DOI: 10.1088/1742-6596/1546/1/012097, 2020.

[13] A. Ogden and M. El-Haj, Financial narrative summarisation using a hybrid TF-IDF and clustering summariser: AO-Lancs system at FNS 2022, *Proc. of the 4th Financial Narrative Processing Workshop*, vol.1958, no.6, pp.79-82, 2022.

[14] F. Shehzad, A. Rehman, K. Javed, K. A. Alnowibet, H. A. Babri and H. T. Rauf, Binned term count: An alternative to term frequency for text categorization, *Mathematics*, vol.10, no.21, pp.1-25, DOI: 10.3390/math10214124, 2022.

[15] P. Guleria and M. Sood, Artificial intelligence and machine learning for the healthcare sector: Performing predictions and metrics evaluation of ML classifiers on a diabetic diseases data set, in *Cognitive and Soft Computing Techniques for the Analysis of Healthcare Data*, A. K. Bhoi, V. H. C. de Albuquerque, P. N. Srinivasu and G. Marques (eds.), Academic Press, 2022.

[16] K. Yeturu, Machine learning algorithms, applications, and practices in data science, in *Principles and Methods for Data Science*, A. S. R. S. Rao and C. R. Rao (eds.), Elsevier, 2020.

[17] Z. Jun, The development and application of support vector machine, *Journal of Physics: Conference Series*, vol.1748, no.5, DOI: 10.1088/1742-6596/1748/5/052006, 2021.

[18] S. K. Mohapatra and M. N. Mohanty, Big data classification with IoT-based application for e-health care, in *Cognitive Big Data Intelligence with a Metaheuristic Approach*, S. Mishra, H. K. Tripathy, P. K. Mallick, A. K. Sangaiah and G.-S. Chae (eds.), Academic Press, 2022.

[19] A. Dharma, A. C. Naibaho, L. M. Bancin and A. Andrew, Classification of Indonesian slang using Naïve Bayes and decision tree methods on social media, *J. Mantik*, vol.6, no.36, pp.1792-1798, 2022.

[20] R. Blanquero, E. Carrizosa, P. Ramírez-Cobo and M. R. Sillero-Denamiel, Variable selection for Naïve Bayes classification, *Comput. Oper. Res.*, vol.135, 105456, DOI: 10.1016/j.cor.2021.105456, 2021.

[21] K. Shah, H. Patel, D. Sanghvi and M. Shah, A comparative analysis of logistic regression, random forest and KNN models for the text classification, *Augmented Human Research*, vol.5, no.1, DOI: 10.1007/s41133-020-00032-0, 2020.

[22] I. Abuqaddom, B. A. Mahafzah and H. Faris, Oriented stochastic loss descent algorithm to train very deep multi-layer neural networks without vanishing gradients, *Knowledge-Based Syst.*, vol.230, 107391, DOI: 10.1016/j.knosys.2021.107391, 2021.

[23] C. Griesbach, A. Mayr and E. Bergherr, Variable selection and allocation in joint models via gradient boosting techniques, *Mathematics*, vol.11, no.2, 411, DOI: 10.3390/math11020411, 2023.

[24] Q. H. Nguyen et al., Influence of data splitting on performance of machine learning models in prediction of shear strength of soil, *Math. Probl. Eng.*, vol.2021, DOI: 10.1155/2021/4832864, 2021.

[25] A. Zhou, Y. Zhang and M. Lu, C-Transformer model in Chinese poetry authorship attribution, *International Journal of Innovative Computing, Information and Control*, vol.18, no.3, pp.901-916, DOI: 10.24507/ijicic.18.03.901, 2022.