**International Journal of Electrical and Computer Engineering (IJECE)**
Vol. 12, No. 6, December 2022, pp. 6707~6715
ISSN: 2088-8708, DOI: 10.11591/ijece.v12i6.pp6707-6715 

# Implementation of the C4.5 algorithm for micro, small, and medium enterprises classification

**Sri Lestari[1], Yulmaini[1], Aswin[2], Sylvia[1], Yan Aditiya Pratama[1], Sulyono[1]**
[1]Faculty of Computer Science, Institute of Informatics and Business Darmajaya, Lampung, Indonesia
[2]Faculty of Economics and Business, Institute of Informatics and Business Darmajaya, Lampung, Indonesia

**Article Info**

***Article history:***

Received Sep 29, 2021
Revised Jun 14, 2022
Accepted Jul 10, 2022

***Keywords:***

Algorithm C4.5
Classification of micro, small, and medium enterprises
Data mining
Decision tree

**ABSTRACT**

The coronavirus disease-19 (COVID-19) pandemic has spread to various countries including Indonesia. Thus, implementing large-scale social restrictions *(Bahasa: Pembatasan Sosial Berskala Besar (PSBB))* has resulted in the paralysis of the economy in Indonesia. including micro, small, and medium enterprises (MSMEs) have decreased turnover and even went out of business. The Department of Cooperatives and Small and Medium Enterprises (SMEs) in Pesawaran Regency, Lampung, oversees 3,808 MSMEs, whose development should be monitored as a basis for determining policies. However, there are problems in classifying MSMEs according to their categories because they have to check the existing data one by one, so it takes a long time. Therefore, this study proposed the C4.5 algorithm to solve this problem. In addition, this research compared with the naïve Bayes algorithm to find out which algorithm had a good performance and is suitable for this case. The results showed that 91% of MSMEs were included in the micro category, 8% was in a small category, and 1% was in the medium category. Based on the results, it explained that the C4.5 algorithm was bigger than naïve Bayes with a difference in the value of 3.79%. It had an accuracy value of 99.2%. Meanwhile, naive Bayes was 95.41%.



***Corresponding Author:***

Sri Lestari
Faculty of Computer Science, Institute of Informatics and Business Darmajaya
ZA. Pagar Alam St., No. 93 Gedong Meneng, Bandar Lampung, Indonesia
Email: srilestari@darmajaya.ac.id

## 1. INTRODUCTION

According to the head of economics of the international monetary fund (IMF), it is estimated that the global economy will decline by 3% due to the coronavirus disease (COVID-19) outbreak in the world one year ago. The statistics agency stated that Indonesia was in a recession in the third quarter with minus 3.49% per year. The decline in business was also impacted by micro, small, and medium enterprises (MSMEs). It was one of the pillars of the national economy. According to the Presidential Decree No. 99 of 1998, the definition of MSMEs are small-scale people's economic activities in business fields and they need to be protected to prevent unfair business competition. Head of Cooperatives and small, and medium enterprises (SMEs) Office in Lampung, Agus Nompitu, said that based on the Decree of the Ministry of Cooperatives and SMEs of the Republic of Indonesia, there were 245,136 MSMEs spread across 15 regencies/cities in Lampung Province. In Pesawaran Regency, there are 3808 MSMEs spread across 11 sub-districts. This data can make it easier for the Cooperatives and MSMEs Office of Pesawaran Regency to analyze the characteristics of MSMEs so that policies, policy revisions, assistance, marketing training can be appropriate on target. These policies are expected to increase the creativity and quality of MSMEs with economic value for the realization of prosperity for the people of Pesawaran Regency. This study will help realize the mission

of Pesawaran Regency (Mission VI). The mission of Pesawaran Regency (Mission VI) is "to realize superior and creative human resources and strengthen the regional economy".

MSMEs in Pesawaran Regency continue to develop so the number is increasing. It must be accompanied by rapid and appropriate administration. MSMEs need to be classified according to the criteria based on the Law of the Republic of Indonesia Number 28 of 2008 concerning MSMEs. However, it faces obstacles given a large amount of data. The data collection is carried out with a simple application. Administrative staff must check MSME data one by one in conducting classifications. A model is needed that makes it easier to do the classification. One of the methods used is to implement the C4.5 algorithm to predict the classification so that the process will be faster and more accurate. It is done to make it easier for the Cooperatives and SMEs Office in determining policies so that they are more targeted, and the results will be better for the development of existing MSMEs.

The narrative science survey in 2016 stated that 38% of large companies already used artificial intelligence (AI) technology. This figure continues to increase to 62% in 2018. The development of AI was moving faster and running in rapid progress in every area of human life. There are several methods developed in artificial intelligence, including support vector machine (SVM) and convolution neural network (CNN) used in the study of Qasmieh *et al.* [1] to classify and segment the occlusive iris. CNN algorithm is used for the classification of gangrene disease through high-resolution graphic images. This disease is very deadly because of the lack of blood supply to the body [2]. A similar algorithm, namely custom CNN, to classify images of female faces and male faces, was proposed by Zaman [3]. In addition, there is the k-nearest neighbor (KNN) and ID3 algorithm used in Sudarma and Harsemadi's research to classify music based on mood [4]. Next is naïve Bayes which is used to detect spam emails in a study conducted by Jaiswal *et al.* [5]. Furthermore, the C4.5 algorithm is used to predict the risk of rock burst in coal mines [6]. In addition, the C4.5 algorithm is used to classify parental involvement during the school from home (SFH) period, especially for kindergarten and elementary school children [7]. Furthermore, the decision tree (DT) algorithm, C4.5, was used in the study of Lei and Zeng to evaluate the relationship between perceived social support and exercise behavior. This is done with different interventions to detect the effect of the heterogeneous intervention [8]. The C4.5 algorithm is also used to predict new students who resign so that the results of this prediction can be used by the management to make strategic plans [9].

In line with previous research, our research also uses the DT algorithm. This method is very suitable for approximating reasoning, especially for systems that deal with problems that are difficult to define. The advantage of using the DT algorithm has a classification ability that is similar to the ability of human reasoning. The C4.5 DT algorithm has the advantage of using memory and computing more efficiently, information is used for classification and creates trees with multiple branches emerging from each node, it can work with missing or continuous data and others [10]. Based on those statements, this study will use the C4.5 DT algorithm to classify MSMEs in Pesawaran Regency to be three classes, namely MSMEs in considering the appropriate regulation so that the performance can be improved.

## 2. COMPREHENSIVE THEORETICAL BASE

This study will use a classification approach. One of the classification algorithms that will be applied is C4.5. In detail, the explanation regarding the classification approach and the working stages of the C4.5 algorithm is as follows:

### 2.1. Classification

One of the techniques in data mining for this study was classification. This technique was used to analyze grouped data dan take an instance. Furthermore, it considered to particular class so that failed classification could be minimalized. Besides, it was used to extract an accurate model for defining a data class from grouped data. The classification consisted of two steps. The first step was creating the model by implementing a classification algorithm on training data. The second step was the model which was extracted in the previous step. It was tested using the prepared data to measure the performance and accuracy of the model. This classification was the process of determining the class label on a non-dataset [11], [12].

### 2.2. Algorithm C4.5 decision tree

The C4.5 decision tree (DT) algorithm is supervised learning that builds a model from training data with known categories, and classification of test data with unknown categories [13], [14]. The C4.5 algorithm was used to create a decision tree. DT were a very powerful and well-known method of classification and prediction. This method turned a very large fact into a decision tree that represents the rules so that it was easy to understand in natural language. In addition, it was expressed in a database language such as structured query language (SQL) to search for records with certain categories. The decision tree can explore

the data between several input variables and a target variable so that a hidden relationship is found [15]. The C4.5 algorithm in building a DT consists of several stages, [16]: i) selecting attribute as root, ii) creating a branch for each value, iii) dividing cases into branches, and iv) repeating the process for each branch until all cases in the branch have the same class. The root attribute was selected based on the highest gain value of the existing attributes. Before calculating the gain, it had calculated the entropy. Entropy calculation used (1). Meanwhile, (2) was used to calculate gain:

$$Entropy\ (S) = \sum_{i-1}^{n} - pi * \log_2 pi \quad (1)$$

where $S$ is set of case, $n$ is total of partition for $S$, and $pi$ is proportion from $S_i$ of $S$.

$$Gain\ (S, A) = Entropy\ (S) - \sum_{i-1}^{n} \frac{|S_i|}{|S|} * Entropy\ (S_i) \quad (2)$$

Where $S$ is set of case, $A$ is attribute, $n$ is total of partition for attribute $A$, $|S_i|$ is number of cases on partition $i$, and $|S|$ is number of cases in $S$.

## 3. RESEARCH METHOD

This study classified MSMEs using the C45 algorithm. Before the classification stage, there were several stages as shown in Figure 1. The initial stage is problem identification. It was about the problems faced by the Cooperatives and SMEs Office in Pesawaran Regency related to MSMEs. Furthermore, it was supported by a related literature study to strengthen the foundation of the study based on the previous studies. Furthermore, it was continued with a data collection total of 3,808 MSMEs data spread across 11 sub-districts in Pesawaran Regency. The data for MSMEs per each sub-district can be seen in Table 1. Next, the data was done by preprocessing. After that, it was done by using the C4.5 Algorithm and evacuated. In detail, Figure 1 explained:

a. Identification of problem

At this step, we conducted observations and interviews with the Pesawaran Cooperatives and SMEs Office to find out the process of data collection and management of MSMEs. The result was found that there were so many problems in the MSMEs data collection process to classify the existing MSMEs into their categories, namely micro, small and medium. This is because a lot of incomplete data is found and must be synchronized with the data asset and turnover.

b. Literature study, previous research, data mining dan classification

The next step is a literature study by looking for related references from various sources, both from books, the internet, journal articles, and proceedings. The results of previous studies were used as a reference in solving problems faced by the Department of Cooperatives and SMEs in Pesawaran Regency. The approach used is one of the data mining techniques, namely classification. The classification algorithm that will be applied is the C4.5 algorithm.

c. Data collection

The next step is to collect data on SMEs. The Pesawaran Regency Cooperatives and SMEs service have 3,808 SMEs spread across 11 sub-districts. The data consists of various types of businesses with diverse assets and turnover. Based on these assets and turnover, the classification of MSMEs will be carried out.

d. Data preprocessing

After the data was obtained, it was continued with data preprocessing, namely by cleaning the data. For uncomplete data or empty attributes, it was able to be replaced with dominant data for any data with the same attributes that had missing values for the data they had. It found that data was with more than one column but it should be able to be used as one column. It will be transformed into data. Normalization of data used aimed to make complex data easier to process. For example, the criteria in MSMEs which were previously divided into 3 columns, namely micro, small and medium, can be used as one attribute, namely business criteria. Furthermore, the gender column which was previously split into two columns, male and female, can be used as one column with the gender attribute.

e. Implementation of decision tree method and naïve Bayes

After preprocessing the data, proceed with the implementation of the algorithm. The algorithm used is the C4.5 algorithm. Meanwhile, the comparison is the naive Bayes algorithm. C4.5 and naive Bayes algorithms are both classification algorithms, so this comparison is equivalent (apple to apple).

f. Evaluation

Evaluation will be carried out in this study to see the performance of the two algorithms (C4.5 and naive Bayes). The assessment used is by looking at the accuracy. This is done to ensure that the classification prediction results from the C4.5 and naive Bayes algorithms have good quality.

This study took the data from the Cooperatives and SMEs Office in Pesawaran Regency with 3,808 MSMEs, spread over 11 sub-districts as shown in Table 1. The next step is data preprocessing [17]. Preprocessing was done with data cleaning and data transformation. The data cleaned noisy, inconsistent data, and data that did not have complete or empty attributes. This study also removed the attributes of mobile phone numbers, education, sub-districts, and length of business. It was because there were many data vacancies in these attributes. Besides, these attributes did not affect the classification results.

The next step was to transform the data. In this stage, it set the alignment of the column with more than one column from data transformation. In addition, data normalization was used to change complex data to be easier to process. For example, the business criteria in MSMEs which were previously broken down into 3 micro, small and medium data can be used as one attribute. it stated that business criteria and the gender column were previously split into two columns, moreover, males and females were in one column, as well as the transformation of turnover and income asset data. The merging of the business criteria column referred to the Law of the Republic of Indonesia Number 28 of 2008 concerning micro, small, and medium enterprises. In chapter IV, it stated the criteria for MSMEs in article 6. Law Number 20 of 2008 is the author's reference for data transformation in the business criteria column. So, it was found that the attributes used in the classification process of MSMEs in Pesawaran Regency were business name, owner's name, type of business, product name, license owned, assets, turnover, and criteria.

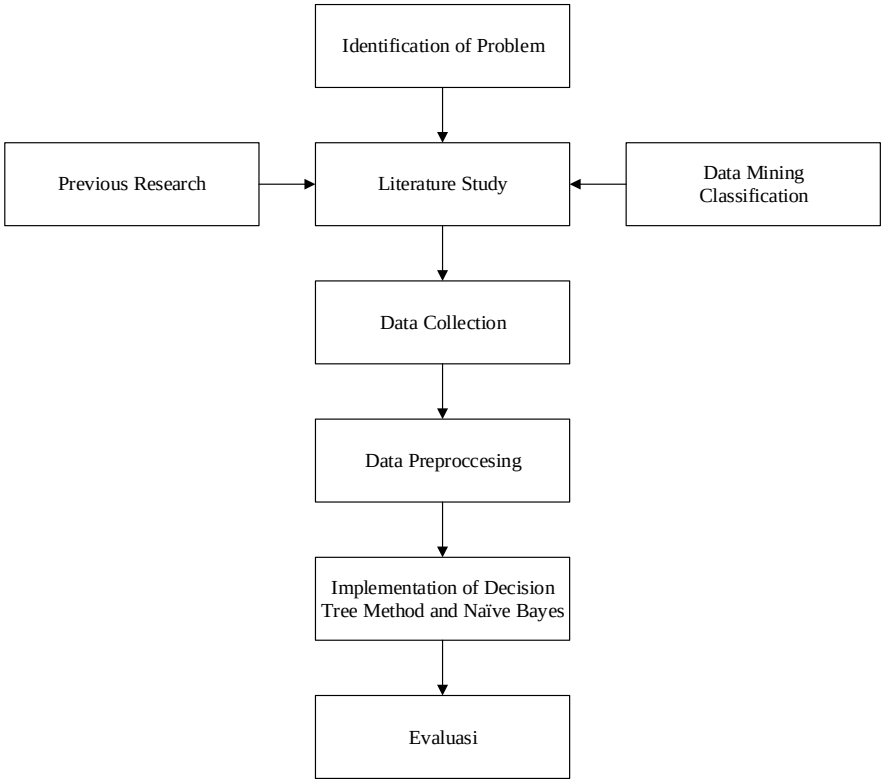


Figure 1. Research stages

Table 1. MSMEs from each district

| Districts | Total MSMEs |
|---|---|
| Gedong Tataan | 390 |
| Tegineneng | 599 |
| Negeri Katon | 126 |
| Kedondong | 572 |
| Waylima | 252 |
| Way Khilau | 144 |
| Punduh Pedada | 301 |
| Marga Punduh | 841 |
| Padang Cermin | 181 |
| Teluk Pandan | 167 |
| Way Ratai | 235 |

## 4. RESULTS AND DISCUSSION

### 4.1. Results

This study classified MSMEs from each sub-district using the C4.5 decision tree algorithm. The tools used by RapidMiner refer to previous research [18]–[24]. It started by reading the data, replacing the missing value, and splitting the data. A comparison was 70% for training data and 30% for testing data in line with previous research [25]–[27]. The classification process used the decision tree algorithm and naïve Bayes, as shown in Figures 2 and 3. The result of MSMEs categories was shown in Table 2 and the result of accuracy evacuation was shown in Table 3. Figure 2, it showed that the implementation model of the C4.5 algorithm used RapidMiner. It was started by reading MSMEs data and continued by filling empty data using replace missing values. The next step was in using algorithm tree (C4.5) and it measured the performance for accuracy.

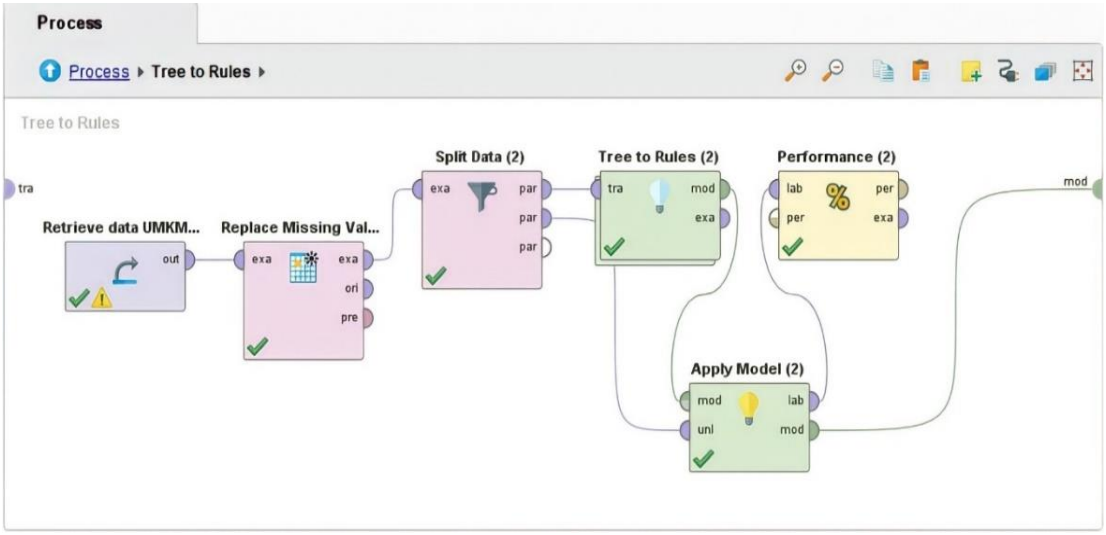


Figure 2. MSMEs classification model with C4.5 decision tree algorithm

Figure 3 showed that the implementation model of the naïve Bayes algorithm using RapidMiner. This step was almost the same as the creation model for C4.5 but it did not use to replace the missing value. It was because it used proper data. The next was doing split data, naïve Bayes Implementation, and performance measurement.

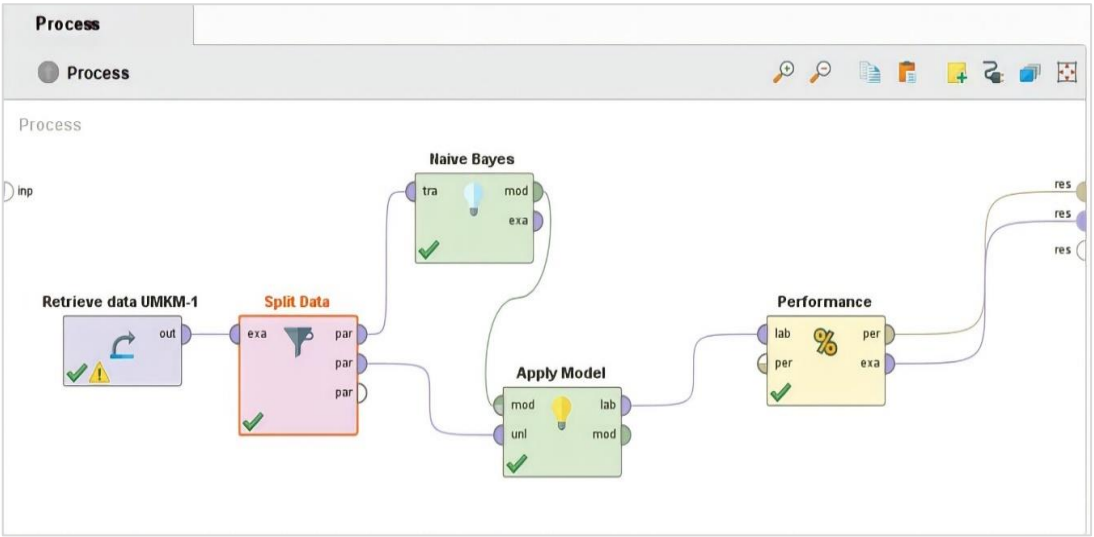


Figure 3. MSMEs classification model with naïve Bayes algorithm

Table 2 showed the results of the MSMEs categories, namely micro, small and medium from each sub-district in Pesawaran Regency. 2 sub-districts were categorized as micro-only SMEs, namely Negeri Katon and Way Khilau sub-districts. While in Table 3 showed the results of the performance evaluation for the C4.5 and naïve Bayes classification models was shown in Table 3. It showed that the average accuracy value of the C4.5 algorithm was 99.2% and naïve Bayes was 95.41%.

Table 2. Results of the MSMEs category in Pesawaran Regency

| No. | Districts | Micro | Small | Medium |
|---|---|---|---|---|
| 1 | Gedong Tataan | 93% | 7% | 0% |
| 2 | Tegineneng | 89% | 11% | 0% |
| 3 | Negeri Katon | 100% | 0% | 0% |
| 4 | Kedondong | 98% | 2% | 0% |
| 5 | Way Lima | 99% | 1% | 0% |
| 6 | Wai Khilau | 100% | 0% | 0% |
| 7 | Punduh Pedada | 89% | 6% | 5% |
| 8 | Marga Punduh | 92% | 6% | 2% |
| 9 | Padang Cermin | 87% | 12% | 1% |
| 10 | Teluk Pandan | 61% | 39% | 0% |
| 11 | Way Ratai | 79% | 21% | 0% |

Table 3. Evaluation of predictions for the MSMEs category in Pesawaran Regency

| No. | Districts | Prediction of Accuracy C45 | Prediction of Accuracy Naïve Bayes |
|---|---|---|---|
| 1 | Gedong Tataan | 98.97% | 97.94% |
| 2 | Tegineneng | 97.91% | 100% |
| 3 | Negeri Katon | 100% | ---- |
| 4 | Kedondong | 100% | 99.36% |
| 5 | Way Lima | 98.53% | 80.60% |
| 6 | Way Khilau | 100% | ---- |
| 7 | Punduh Pedada | 98.75% | 98.73% |
| 8 | Marga Punduh | 100% | 99.17% |
| 9 | Padang Cermin | 100% | 85% |
| 10 | Teluk Pandan | 97.22% | 100% |
| 11 | Way Ratai | 100% | 97.94% |
| | Average | 99.2% | 95.41% |

### 4.2. Discussion

Based on Table 2, MSMEs categories in Pesawaran Regency were Micro. There were only 3 sub-districts that had MSMEs in the Medium category, namely Punduh Pedada, Marga Punduh, and Padang Cermin. These percentages were 91% in micro category, 8% in small category, and 1% in medium category. It was a reference for the Pesawaran Regency Cooperatives and MSMEs Office in making policies to develop these MSMEs.

As for the results of the evaluation for the classification model with the C4.5 and naïve Bayes Algorithm, it showed that the average accuracy values obtained were 99.2% and 95.41%. In naïve Bayes, it was found that there were 2 sub-districts with undefined values, namely in Negeri Katon and Way Khilau sub-districts it was because there was only one attribute, namely only micro class. Meanwhile, it was concluded that the C4.5 algorithm was bigger than the naïve Bayes algorithm with a difference of 3.79% in value.

## 5. CONCLUSION

The results of this study indicate that 91% of MSMEs are included in the micro category, 8% in the small category, and 1% in the medium category. The majority of MSMEs in Pesawaran Regency is still included in the micro classification. Therefore, based on this information, the Department of Cooperatives Cooperatives and SMEs in Pesawaran can take various policies to develop existing MSMEs.

The results of the evaluation for the implementation of the algorithm showed that C4.5 was bigger than naïve Bayes with an accuracy value difference of 3.79%. It had an accuracy value of 99.2%. Meanwhile, naïve Bayes was 95.41%. It was recommended to use the C4.5 algorithm to facilitate the classification process on MSMEs data in Pesawaran Regency so that the process was more rapid and more precise.

**ACKNOWLEDGEMENTS**

This research was supported by a grant from the Directorate of Research and Community Service (DRPM) for The Excellent Applied Research for Higher Education (PTUPT) scheme number B/112/E3/RA.00/2021.

**REFERENCES**

[1] I. A. Qasmieh, H. Alquran, and A. M. Alqudah, "Occluded iris classification and segmentation using self-customized artificial intelligence models and iterative randomized Hough transform," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 5, pp. 4037–4049, Oct. 2021, doi: 10.11591/ijece.v11i5.pp4037-4049.

[2] P. S. Nair, T. A. Berihu, and V. Kumar, "An image-based gangrene disease classification," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 6, pp. 6001–6007, Dec. 2020, doi: 10.11591/ijece.v10i6.pp6001-6007.

[3] F. H. K. Zaman, "Gender classification using custom convolutional neural networks architecture," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 6, pp. 5758–5771, Dec. 2020, doi: 10.11591/ijece.v10i6.pp5758-5771.

[4] M. Sudarma and I. G. Harsemadi, "Design and analysis system of KNN and ID3 algorithm for music classification based on mood feature extraction," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7, no. 1, pp. 486–495, Feb. 2017, doi: 10.11591/ijece.v7i1.pp486-495.

[5] M. Jaiswal, S. Das, and K. Khushboo, "Detecting spam e-mails using stop word TF-IDF and stemming algorithm with naïve Bayes classifier on the multicore GPU," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 4, pp. 3168–3175, Aug. 2021, doi: 10.11591/ijece.v11i4.pp3168-3175.

[6] Y. Wang, "Prediction of rockburst risk in coal mines based on a locally weighted C4.5 algorithm," *IEEE Access*, vol. 9, pp. 15149–15155, 2021, doi: 10.1109/ACCESS.2021.3053001.

[7] I. A. E. Zaeni, D. Rifa Anzani, D. S. Putra, M. Devi, L. Hidayati, and I. Sudjono, "Classifying the parental involvement on school from home during covid-19 using c4.5 algorithm," in *4th International Conference on Vocational Education and Training, ICOVET 2020*, 2020, pp. 253–257, doi: 10.1109/ICOVET50258.2020.9230214.

[8] L. Lei and E. Zeng, "Research on the relationship between perceived social support and exercise behavior of user in social network," *IEEE Access*, vol. 8, pp. 75630–75645, 2020, doi: 10.1109/ACCESS.2020.2987073.

[9] E. Darmawan, "C4.5 algorithm application for prediction of self candidate new students in higher education," *Jurnal Online Informatika*, vol. 3, no. 1, Jun. 2018, doi: 10.15575/join.v3i1.171.

[10] Z. Çetinkaya and F. Horasan, "Decision trees in large data sets," *Uluslararası Muhendislik Arastirma ve Gelistirme Dergisi*, vol. 13, no. 1, pp. 140–151, Jan. 2021, doi: 10.29137/umagd.763490.

[11] S. S. Nikam, "A comparative study of classification techniques in data mining algorithms," *Oriental Journal of Computer Science and Technology*, vol. 8, no. 1, pp. 13–19, 2015.

[12] M. Sadikin and F. Alfiandi, "Comparative study of classification method on customer candidate data to predict its potential risk," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 6, pp. 4763–4771, Dec. 2018, doi: 10.11591/ijece.v8i6.pp4763-4771.

[13] Y. Song, X. Yao, Z. Liu, X. Shen, and J. Mao, "An improved C4.5 algorthm in bagging integration model," *IEEE Access*, vol. 8, pp. 206866–206875, 2020, doi: 10.1109/ACCESS.2020.3032291.

[14] A. R. Arellano, J. Bory-Reyes, and L. M. Hernandez-Simon, "Statistical entropy measures in C4.5 trees," *International Journal of Data Warehousing and Mining*, vol. 14, no. 1, pp. 1–14, Jan. 2018, doi: 10.4018/IJDWM.2018010101.

[15] Rusito and F. M. Taufany, "Implementation of decision tree method and C4.5 algorithm for classification of bank customer data (in Indonesian)," *Infokam*, vol. XII, no. 1, pp. 1–12, 2016.

[16] E. Elisa, "Analysis and application of C4.5 algorithm in data mining to identify factors causing accidents at PT. Arupadhatu Adisesanti Construction (in Indonesian)," *Jurnal Online Informatika*, vol. 2, no. 1, 2017, doi: 10.15575/join.v2i1.71.

[17] C. Wang, D. Chen, Y. Hu, Y. Ceng, J. Chen, and H. Li, "Automatic dialogue system of marriage law based on the parallel C4.5 decision tree," *IEEE Access*, vol. 8, pp. 36061–36069, 2020, doi: 10.1109/ACCESS.2020.2972586.

[18] I. Garcia-Magarino, G. Gray, R. Lacuesta, and J. Lloret, "Survivability strategies for emerging wireless networks with data mining techniques: a case study with NetLogo and RapidMiner," *IEEE Access*, vol. 6, pp. 27958–27970, 2018, doi: 10.1109/ACCESS.2018.2825954.

[19] R. Buchkremer *et al.*, "The application of artificial intelligence technologies as a substitute for reading and to support and enhance the authoring of scientific review articles," *IEEE Access*, vol. 7, pp. 65263–65276, 2019, doi: 10.1109/ACCESS.2019.2917719.

[20] A. Bolt, M. de Leoni, and W. M. P. van der Aalst, "Scientific workflows for process mining: building blocks, scenarios, and implementation," *International Journal on Software Tools for Technology Transfer*, vol. 18, no. 6, pp. 607–628, Nov. 2016, doi: 10.1007/s10009-015-0399-5.

[21] T. Pawar, P. Kalra, and D. Mehrotra, "Analysis of sentiments for sports data using RapidMiner," in *Proceedings of the 2nd International Conference on Green Computing and Internet of Things*, 2018, pp. 625–628, doi: 10.1109/ICGCIoT.2018.8752989.

[22] T. A. Mat, A. Lajis, and H. Nasir, "Text data preparation in rapidminer for short free text answer in assisted assessment," in *2018 IEEE 5th International Conference on Smart Instrumentation, Measurement and Application, ICSIMA 2018*, 2019, pp. 1–4, doi: 10.1109/ICSIMA.2018.8688806.

[23] M. M. Shabtari, V. Kumar Shukla, H. Singh, and I. Nanda, "Analyzing PIMA Indian diabetes dataset through data mining tool 'RapidMiner,'" in *2021 International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2021*, 2021, pp. 560–574, doi: 10.1109/ICACITE51222.2021.9404741.

[24] V. Kalra and R. Aggarwal, "Importance of text data preprocessing and implementation in RapidMiner," in *Proceedings of the First International Conference on Information Technology and Knowledge Management*, Jan. 2018, pp. 71–75, doi: 10.15439/2017KM46.

[25] A. A. Putra, R. Mahendra, I. Budi, and Q. Munajat, "Two-steps graph-based collaborative filtering using user and item similarities: Case study of E-commerce recommender systems," in *2017 International Conference on Data and Software Engineering (ICoDSE)*, Nov. 2017, pp. 1–6, doi: 10.1109/ICODSE.2017.8285891.

[26] Y. El Madani El Alami, E. H. Nfaoui, and O. El Beqqali, "Toward an effective hybrid collaborative filtering: A new approach based on matrix factorization and heuristic-based neighborhood," in *2015 Intelligent Systems and Computer Vision (ISCV)*, Mar. 2015, pp. 1–8, doi: 10.1109/ISACV.2015.7105543.

[27] G. Huo, Z. Wu, and J. Li, "Underwater object classification in sidescan sonar images using deep transfer learning and semisynthetic training data," *IEEE Access*, vol. 8, pp. 47407–47418, 2020, doi: 10.1109/ACCESS.2020.2978880.

## BIOGRAPHIES OF AUTHORS

**Sri Lestari** obtained her doctorate (Dr from the Electrical Engineering Doctor Program, Universitas Gadjah Mada, Yogyakarta, Indonesia in 2019. She is a lecturer in the Department of Computer Science, Institut Informatika and Bisnis Darmajaya, Bandar Lampung, Indonesia. Her research interests include artificial intelligence, recommendation system, collaborative filtering, data mining, decision support systems, and software engineering. Her representative published articles listed: Porat$^{Rank}$ to Improve Performance Recommendation System (Lecture Notes in Electrical Engineering, Springer 2021), Decision Support System for Service Quality Using SMART and Fuzzy ServQual Methods (JUITA: Jurnal Informatika, 2021), WP-Rank: Rank Aggregation based Collaborative Filtering Method in Recommender System (International Journal of Engineering and Technology (UAE), 2018), Performance Comparison of Rank Aggregation Using Borda and Copeland in Recommender System (International Workshop on Big Data and Information Security (IWBIS 2018)), NRF: Normalized Rating Frequency for Collaborative Filtering (The 2018 International Conference on Applied Information Technology and Innovation (ICAITI 2018)), Design and analysis model application system teaching online media. (Proceedings of the International Conference on Information Technology and Business (ICITB), 2016). Model of performance classification and selection doses achievement with algorithm C4.5 (Proceeds of Cysts, 2014). Application of Weighted Product Model for Selection of Prospective Employees (Journal of Information Systems, 2014). She can be contacted at email: srilestari@darmajaya.ac.id.

**Yulmaini** obtained her Master's degree from the postgraduate Program, at Universitas Gadjah Mada, Yogyakarta, Indonesia in 2011. She is a lecturer in the Department of Computer Science, Institut Informatika and Bisnis Darmajaya, Bandar Lampung, Indonesia. Her research interests include artificial intelligence, decision support systems, and software engineering. Her representative published articles lists: Application of Tsukamoto's Fuzzy Inference System in Determination of concentration for students' thesis topics (Journal of Critical Reviews, 2019), Formulating a Higher Education Competitiveness Model (Journal Of Talent Development And Excellence, 2020), Competitiveness Universities Strategy Development based on the Research and Information Technology (Journal Of Engineering and Technology Management, 2020), Improvement of One-Dimensional Fisherface Algorithm to extract the Features (Case study: Face Recognition) in Proceedings of International Conference on Science and Technology 2019, Implementation of Analytic Hierarchy Process For Determining Priority Criteria In Higher Education Competitiveness Development Strategy Based On RAISE++ Model (Proceedings of The 2nd Joint International Conference on Emerging Computing Technology and Sports (JICETS) 2019), Application of Tsukamoto's Fuzzy Inference System in Determination of concentration for students' thesis topics (Proceedings of the International Conference on Information Technology and Business (ICITB), 2019). She can be contacted at email: yulmaini@darmajaya.ac.id.

**Aswin** is a lecturer in the Department of Management specialization in marketing, Institut Informatika and Bisnis Darmajaya, Bandar Lampung, Indonesia. Her research interests include consumer behavior, customer satisfaction, service quality, brand equity, and social media marketing. Her representative published articles listed: Role of Social Media Marketing to Enhance The Supply Chain and Business Management (International Journal of Supply Chain Management (IJSCM), 2020), Analisis Brand Equity Perguruan Tinggi Swasta di Bandar Lampung (National Journal of Gentiaras dan Akuntansi (GEMA), 2021), Analysis of student and student interest in the Darmajaya IIB management study program (S1) Concentration in Bandar Lampung (National Journal of Applied Accounting and Business), 2019), Analysis of Community Influence on the Use of Family Planning Devices in Bandar Lampung City, (National Journal of GEMA Economic, 2017). She can be contacted at email: aswin@darmajaya.ac.id.

**Sylvia** obtained a bachelor's degree from the Informatics Engineering undergraduate program, IIB Darmajaya, Bandar Lampung, Indonesia in 2021. She was active in the field of organization and served as general treasurer at the IIB Darmajaya Student Association. Moreover, she was also active in participating in debate competitions at the faculty, provincial and national levels. She can be contacted at email: sylviamkmr@gmail.com.

**Yan Aditiya Pratama** is lecturer in Department of Management at Institute Informatics and Business Darmajaya, Bandarlampung. His specializations are Human Resource Development and Operation Management. Currently, he has concerned with micro, small, and medium enterprises development, digital concept operation, and entrepreneurship. Her representative published articles lists: Marketing Strategy through Swot Analysis on the Puncak Mas Tourist Attraction in Bandar Lampung (International Conference on Information Technology and Business (ICITB), 2021), The Recommendation System for Increasing the Independence of Micro, Small and Medium Enterprises (MSMEs) Using the Normalized Rating Frequency (NRF) Method (2021 4th International Conference on Information and Communications Technology (ICOIACT), 2021), Model Profit Economic on Female Workers In Indonesia (International Journal of Economics, Business and Accounting Research, 2021), The Role Of Positive Affect Mediators on Person Organization Fit And Job Satisfaction (JIM UPB (*Jurnal Ilmiah Manajemen Universitas Putera Batam*), 2020), The Strategic Development Of Pugung Raharjo Megalithic Park Using Swot Analysis (International Conference on Information Technology and Business (ICITB), 2020), Blockchain Technology for Tracking Chain Supply (International Conference on Information Technology and Business (ICITB), 2020). He can be contacted at email: yanaditiyapratama@darmajaya.ac.id.

**Sulyono** Graduated in the Informatics Engineering Study Program, Faculty of Computer Science, Darmajaya Institute of Informatics and Business in 2007, graduated Masters Program in Informatics Engineering with a specialization in Software Engineering, Darmajaya Institute of Informatics and Business in 2015. Became a programmer since sitting on the lecture bench, the application product has been widely used by users. Currently, he is a permanent lecturer in the Informatics Engineering Study Program, Faculty of Computer Science, Darmajaya Institute of Informatics and Business, Bandar Lampung City, Lampung Province. Capable of programming-based courses from the first semester to the final semester. Has attended several pieces of training that are appropriate in the field of teaching, especially programming. He can be contacted at email: sulyono@darmajaya.ac.id.