ICIC Express Letters
Volume **16**, Number **10**, October **2022**

ICIC International ©2022 ISSN 1881-803X
pp. 1079–1086

MODELING THE CLASSIFIER OF ONLINE LEARNING EFFECTIVENESS USING MACHINE LEARNING

Paulus Insap Santosa[1,*] and Rianto[2]

[1]Department of Electrical Engineering and Information Technology
Universitas Gadjah Mada
Grafika Street 2, Yogyakarta 55281, Indonesia
*Corresponding author: insap@ugm.ac.id

[2]Department of Data Science
Universitas Teknologi Yogyakarta
Siliwangi Street, Yogyakarta 55285, Indonesia
rianto@staff.uty.ac.id

Received January 2022; accepted March 2022

Abstract. *The COVID-19 pandemic that hits almost all countries in the world has caused extraordinary changes, including education. To minimize the spread of the virus, educational institutions are forced to conduct online teaching and learning processes. Online learning using certain media and technology has been initiated and implemented long before the pandemic. However, their effectiveness has not been fully understood. This study aims to model a classifier to determine the effectiveness of online learning based on eight determinants. It also compared the three algorithms, i.e., Support Vector Machine, Naïve Bayes, and Decision Tree. The size of the dataset was 400. It was collected via an online survey. The respondents were undergraduate students from several higher education in Indonesia. The result showed that Naïve Bayes and Decision Tree have the same accuracy, i.e., 93.3%, slightly higher than Support Vector Machine of 91.7%. Based on the precision and recall value, Naïve Bayes achieved the highest precision and recall value of 93.2% and 93.9%, respectively.*

Keywords: Academic performance, Classifier modeling, Ease of navigation, Effectiveness, Machine learning, Online learning

1. **Introduction.** The closure of the education sector, due to the COVID-19 pandemic, forced education sector managers to carry out educational activities to comply with government regulations, i.e., through online mode. Technologically, this is not a new approach because online teaching and learning have been developed long before the pandemic, although limited. During the current pandemic, the online teaching and learning process is implemented at every level of education, from kindergarten to higher education. This forced implementation affects the effectiveness of the teaching and learning process as indicated by several indicators, e.g., students' engagement and understanding of the course materials. Based on the survey, many institutions said that they were not ready to implement online education; however, 96% of education practitioners in Indonesia prefer to use online learning [1].

Indonesia is an archipelagic country consisting of thousands of geographically separated islands. This geographical condition results in uneven telecommunication infrastructure and Internet networks, resulting in unequal access to online learning services. Inequality in the availability of Internet bandwidth and the imbalance in the competence of human resources cause the effectiveness of online learning to be low. The low efficacy of teaching and learning is not only influenced by human resources and infrastructure. The difficulty of assessing student motivation due to the lack of direct interaction between teachers

DOI: 10.24507/icicel.16.10.1079

and students also affects online learning success [2]. This situation requires a different approach.

Online learning allows students to learn from the comfort of their homes. Virtually, they can attend the course while doing something else, whether related to the attended course. The students may treat online learning as having a holiday. This holiday attitude affects their physical and mental preparations differently. On the side of the lectures, they are required not only as subjects who deliver the course materials but also must build and improve the learning environment and motivate the students that online learning is equal to direct face-to-face lectures. To tackle this situation, synchronous and asynchronous modes are implemented. Synchronous mode means that teachers and students meet face-to-face virtually using online media such as Zoom, Google Meet, and Microsoft Teams. In contrast, asynchronous mode only uses a learning management system without face-to-face meetings [3].

Students have different experiences when attending online courses, including bad experiences. Many students had experienced stressful online learning where their stress levels were mild to severe. Previous studies showed factors that have caused the above experience, e.g., mood, sleep hours, time of the day, and energy level [4], compassion, health problems, economic problems, communication problems, and stress management [5], demographic backgrounds, motivational measures, and ill-structured data [6], and usage behavior [7]. Thus, the preferred online learning mode is essential to find [8]. Previous studies seldom explore the influence of aspects related to the teaching and learning process such as delivery methods, lecture time, assignment feedback, and lecture modes (synchronous, asynchronous, blended) on the effectiveness of online learning. This study aims to develop a model for classifying the features that determine the effectiveness of online learning for undergraduate students at Indonesian universities. It compares three classification algorithms, i.e., Support Vector Machine, Naïve Bayes, and Decision Tree, to obtain the highest accuracy or other related measures. Model analysis was carried out using the confusion matrix to determine the accuracy, precision, recall, and error rate [9]. The Decision Tree is nonparametric with no weight parameters that affect the results; thus, it is easy to use and interpret. Support Vector Machine can handle complex data, making it able to perform multi-class classification. On the other hand, with its probabilistic model, Naïve Bayes is the most commonly used machine learning algorithm.

The current study combines several determinants of student academic performance originating from students, lecturers, and some features from the online learning applications. By using this approach, this article provides more holistic types of features used in the classification process than previous studies.

The structure of this article is as follows. Following this introduction, Section 2 explains the method and related works. The method section describes the survey instrument for data collection, survey period, respondent sampling method, survey, data cleansing, and analysis. Subsequently, the related works section discusses previous studies related to this current study. Section 3 discusses the result of the data analysis and the discussion of the findings. Section 4 provides the readers with the conclusion of the study.

2. Method and Related Works.

2.1. **Method.** This study comprises several steps, from survey questionnaires development to classifier modelling. The questionnaire was developed to determine the combination of several features that respondents rated as effective or ineffective online learning models. Table 1 shows all features used in this study and the number of options available for respondents to choose from when they fill out the survey. The choices for each feature provide a categorical data.

Table 1. The features used

Feature	Definition	#options
Preferred time of study	Best/worst time of the day to attend class	5
Preferred learning mode	The preferred learning mode	3
Synchronous application	The most liked/disliked application for synchronous learning	4
Asynchronous application	The most liked/disliked application for asynchronous learning	4
Course material format	The most liked/disliked course material format	3
Mode of delivery	The delivery mode of the course material	6
Assignment feedback	Whether there is feedback of the given assignment	2
Ease of navigation	Ease of navigation of the application used during the class	2

Data was collected using an online survey using Google Forms. The survey was conducted from July 11th to July 30th, 2021. The respondents were undergraduate students from several higher institutions in Indonesia, both state and private. The target higher institutions were chosen using convenient sampling. The students from the target institutions were asked to participate in this survey voluntarily through social media and email invitations. The total number of students who participated in the survey was 523. Raw data from Google Forms were collected as an Excel file.

The subsequent steps were data preparations comprising standardization and data cleansing. The purpose of this step is to ensure data validity. This process was carried out because some responses did not meet some criteria, such as the names of the institutions not found in Indonesia, incomplete answers, respondents gave 'neutral' choice to all questionnaires, and duplicate responses. After the data preparation step was completed, the coding and labeling step proceeded.

Coding transforms respondents' answers previously stated in a text form into numerical values according to the number of options in the questionnaires. The purpose of the coding process is to facilitate quantification during data analysis. The coding process was done using Microsoft Excel. Labeling refers to the data on the questionnaires filled out by respondents containing questions for effective and ineffective indicators. The labels used in the study were 'effective' and 'not-effective', denoted by 1 and 0, respectively.

The next step was model development. Like coding and labeling, model development was also conducted using Microsoft Excel. The dataset used in the model development was 400, which was evenly divided into 200 data labeled 'effective' and 200 data labeled 'not-effective'. The model proposed in this study was binary classification with 1 and 0 to represent 'effective' and 'not-effective', respectively [10].

The developed model was then analyzed using three classification algorithms, i.e., Support Vector Machine, Naïve Bayes, and Decision Tree. The dataset was split into 70% for training and 30% for testing the model. Python was used to analyze the developed model utilizing Google Collaboration. Subsequently, the data were analyzed using a confusion matrix [11] to determine the model goodness expressed in terms of accuracy, precision, recall, and error rate. Equation (1), Equation (2), and Equation (3) present the accuracy, precision, and recall, respectively [12]. In these equations, TP, TN, FP, and FN are True Positive, True Negative, False Positive, and False Negative, respectively.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$precision = \frac{TP}{TP + FP} \tag{2}$$

$$recall = \frac{TP}{TP + FN} \tag{3}$$

Another metric to measure the model goodness is F1score. This metric combines prediction and recall to assess the diagnostic performance of the prediction algorithm. Equation (4) presents the F1score formula [13]:

$$F1score = 2 * \frac{precision * recall}{precision + recall} \tag{4}$$

The error rate is calculated by subtracting one from the accuracy score, as shown in Equation (5) [14]:

$$error\ rate = 1 - accuracy \tag{5}$$

2.2. **Related works.** One of the well-known classification methods is the Decision Tree. This method can simplify complex problems into easier ones to interpret [15]. Maaliw III [16] conducted a study motivated by the failure of the virtual learning environment to recognize the needs of each student in learning. The respondents in this study were 507 students. The result of this study was a model for classifying student learning styles that were analyzed using the Decision Tree J48 algorithm. This model was then used to create a prototype of an adaptive virtual learning environment by considering students' preferences to develop an intelligent user interface.

Another focus on online learning research is predicting students' academic performance. Research on predicting academic performance aims to provide early intervention to students expected to get low performance. The prediction model was developed using data on student interest in one of the implemented teaching materials. The developed prediction model meets the requirements with 84% to 93% accuracy. However, this study only provides suggestions for creating the academic environment and has not discussed online learning methods in-depth [17]. A similar study was conducted by [7]. They found that the frequency of Internet use correlated positively with students' academic performance. On the other hand, the volume of Internet traffic correlated negatively. These findings showed that online learning is not only about the use of Internet resources but also how these resources are utilized.

Learning styles play an essential role in helping students master the course materials. The transition from offline to online learning is not just about technology, but students' appropriate learning style also has a significant role [18]. This study found that students prefer to use a case study model to understand the material that is difficult to understand. In addition to finding dominant features in learning styles, this study showed that the Support Vector Machine and the Decision Tree have the highest and the lowest accuracy, respectively. Understanding learning materials is related to the learning method and associated with student cognitive level. To understand the appropriate student cognitive level in determining the proper method in the teaching and learning process, a study conducted by [19] carried out a classification of student cognitive levels using a tree-based classifier expanded with a meta-algorithm called LogitBoost. This approach gives an accuracy of 97.169%.

A study used transfer learning to classify six types of vehicles in different study settings, i.e., crossover, sedan, hatchback, van, pickup, and minivan [20]. A testing dataset comprising 640 car images was used. In the actual experiment, a Stanford-based dataset was used. This dataset consists of 8000 car images. Seventy percent of the dataset was used as the training data, including 10% for validation, and another 30% was used as the test data. The use of transfer learning combined with the duplication-based data augmentation technique increases accuracy from 92.68% to 99.70%.

3. Result and Discussion.

3.1. **Result.** This study aims to model a classifier to determine the effectiveness of online learning. The developed model used a binary classification where 0 represents 'not-effective' and 1 represents 'effective'. It was generated using a dataset of 400 rows. The modeling process comprised coding and labeling as stated in the previous section.

The developed model was tested to assess its accuracy, precision, and recall. Evaluating the model accuracy is essential because the model will be used as a knowledge source to classify the effectiveness of online learning. The accuracy test was carried out using three classification algorithms, i.e., Support Vector Machine (SVM), Naïve Bayes (NB), and Decision Tree (DT). Each classification algorithm has its advantages and disadvantages. Figure 1 presents the result of the data analysis in terms of confusion matrix of SVM, NB, and DT, respectively. Based on the confusion matrix depicted in Figure 1, the comparison of the three algorithms in terms of their True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) is presented in Table 2. For each algorithm, the number of data test was 120 (30% of the total dataset).

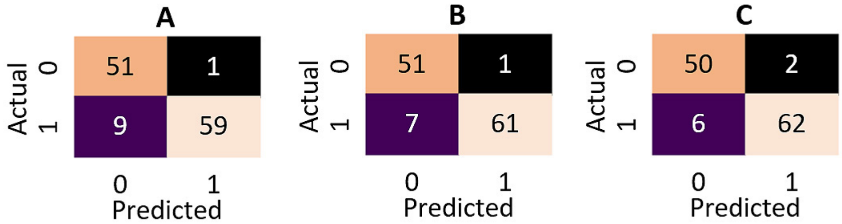

Figure 1. Confusion matrix for (A) SVM, (B) NB, and (C) DT

Table 2. Summary of the confusion matrix of SVM, NB, and DT

Algorithm	TP	TN	FP	FN
Support Vector Machine (SVM)	59	51	1	9
Naïve Bayes (NB)	61	51	1	7
Decision Tree (DT)	62	50	2	6

In order to have a deeper understanding of the model, three other parameters were tested. These three parameters were precision, recall, and F1score. Table 3 presents the comparison of the three algorithms in terms of their respective accuracy, precision, recall, F1score, and macro average.

Table 3. The result of the analysis

	SVM			NB			DT			
	P	R	F1	P	R	F1	P	R	F1	Sup
0	85.0%	98.1%	91.1%	87.9%	98.1%	92.7%	89.3%	96.2%	92.6%	52
1	98.3%	86.8%	92.2%	98.4%	89.7%	93.8%	96.9%	91.2%	93.9%	68
A			91.7%			93.3%			93.3%	120
MA	91.7%	92.5%	91.7%	93.2%	93.9%	93.3%	93.1%	93.7%	93.3%	120

Notes: SVM: Support Vector Machine, NB: Naïve Bayes, DT: Decision Tree

A: Accuracy, MA: Macro Average, P: Precision, R: Recall, F1: F1score, Sup: Support

3.2. **Discussion.** The comparison between the three classification methods, namely SVM, NB, and DT, aims to find the best algorithm based on the accuracy results. SVM has the advantage of high feature space accuracy [21], while NB is a simple classification algorithm for probabilistic classifiers [22]. The DT classification algorithm is more straightforward and accessible because it breaks the dataset into smaller subsets so that the object class differences are more natural [23].

SVM that has multiple kernels is more appropriate to implement high-space classification. The essential criterion in SVM classification is the maximum hyperplane margin. This characteristic makes SVM suitable and recommended for certain types of classification, especially text analysis [24]. The above reason provides one of the justifications why SVM was not appropriate in this study, although it has a high accuracy value. The model developed in this study uses a structured dataset that has gone through the coding process as needed, while SVM is more recommended for unstructured data. Data collection in this study was done using a survey, one of the most effective tools for drawing conclusions [25]. It employed the Likert scale to code categorical data; thus, the data is more structured.

In contrast to SVM, the NB and DT algorithms produce a slightly higher accuracy value even though the difference is only 1.6% as shown in Table 3. NB has several advantages as a classification algorithm, including working with not too much data with good accuracy results [26]. This advantage is often used as one of the considerations for researchers to use NB. On the other hand, the DT algorithm also has advantages in its simplicity and ease of predicting future events based on historical data [27]. The advantages possessed by NB and DT are closely related to the dataset used in this study, resulting in a reasonably high accuracy value.

Looking further at Table 3, the accuracy of both the NB and DT algorithms was 93.3%, so the error rate obtained was 6.7%. This value is considered sufficient accuracy for a model applied in supervised learning [28]. However, even though the accuracy values for NB and DT are the same, there are differences between these two algorithms when viewed from their respective confusion matrix.

Table 2 shows that in terms of True Positive (TP), DT was slightly better than NB, while in terms of True Negative (TN), NB was slightly better than DT. Based on the labeling system used in this study, i.e., 1 for 'effective' and 0 for 'not-effective', DT has a slightly greater chance of incorrect prediction on the positive label, while NB on the negative label. The slight difference between NB and DT is apparent from their respective precision and recall as shown in Table 3. Data in Table 3 indicates that NB has a better recall (sensitivity) rate than DT, although it is tiny. This finding is in line with several previous studies and recommends the application of the NB algorithm in the development of the classifier model [29, 30]. Subsequently, the model developed using the NB algorithm will be used to evaluate the online learning process that the student's performance can be predicted in terms of success and failure. If indications of failure were detected, learning methods should be changed to make sure that the learning outcomes can be achieved.

Subsequently, the model developed using the NB algorithm will be used to evaluate the online learning process that the student's performance can be predicted in terms of success and failure. If indications of failure were detected, learning methods should be changed to make sure that the learning outcomes can be achieved.

4. **Conclusions.** This study proposes a classifier model to determine the effectiveness of online learning. Three models were employed to test the model, i.e., Support Vector Machine, Naïve Bayes, and Decision Tree. The analysis results showed that the accuracy for these three algorithms is 91.7%, 93.3%, and 93.3%, respectively. The analysis results show that the accuracy for the Naïve Bayes and Decision Tree algorithms is the same. However, the Naïve Bayes' macro average recall value, which is 93.9%, is slightly higher

than the Decision Tree of 93.7%. These numbers show that Naïve Bayes is recommended to develop predictive classification models for online learning. Readers are advised to understand that the results obtained are in the context of the data collected and analyzed in this study. The limitations of the study include the features used in this study. Although in this study eight features were used, there must be other features that may affect the effectiveness of online learning. All features used in this research are external, i.e., features that came from "outside" the user. There are internal features that are also believed to affect the effectiveness of online learning. One of the suggestions for developing this study is to determine the dominant features, both internal and external, that may affect online learning success. The subsequent study should implement a classification model that is combined with selected features to be used to test the consistency of the model with the actual case.

Acknowledgment. This work is partially supported by the Graduate School of Universitas Gadjah Mada. The author(s) would like to acknowledge the respondents participated in the survey. The author(s) also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] D. F. Murad, R. Hassan, Y. Heryadi, B. D. Wijanarko and Titan, The impact of the COVID-19 pandemic in Indonesia (face to face versus online learning), *Proc. of 2020 3rd Int. Conf. Vocat. Educ. Electr. Eng. Strength. Framew. Soc. 5.0 through Innov. Educ. Electr. Eng. Informatics Eng. (ICVEE2020)*, pp.2020-2023, DOI: 10.1109/ICVEE50212.2020.9243202, 2020.

[2] D. L. Edy, Widiyanti and Basuki, Revisiting the impact of project-based learning on online learning in vocational education: Analysis of learning in pandemic COVID-19, *The 4th Int. Conf. Vocat. Educ. Training (ICOVET2020)*, pp.378-381, DOI: 10.1109/ICOVET50258.2020.9230137, 2020.

[3] M. Gazan, Synchronous and asynchronous online learning: Perceptions of students at a State University in Turkey, *FIRE Futur. Implementations Res. Educ.*, vol.1, no.2, pp.96-107, 2020.

[4] P. Kaur, H. Kumar and S. Kaushal, Affective state and learning environment based analysis of students' performance in online assessment, *Int. J. Cogn. Comput. Eng.*, vol.2, no.12, pp.12-20, DOI: 10.1016/j.ijcce.2020.12.003, 2021.

[5] X. Lin, College student employment data platform based on FPGA and machine learning, *Microprocess. Microsyst.*, no.11, DOI: 10.1016/j.micpro.2020.103471, 2020.

[6] J. Y. Wu, Learning analytics on structured and unstructured heterogeneous data sources: Perspectives from procrastination, help-seeking, and machine-learning defined cognitive engagement, *Comput. Educ.*, vol.163, no.11, DOI: 10.1016/j.compedu.2020.104066, 2021.

[7] X. Xu, J. Wang, H. Peng and R. Wu, Prediction of academic performance associated with Internet usage behaviors using machine learning algorithms, *Comput. Human Behav.*, vol.98, no.4, pp.166-173, DOI: 10.1016/j.chb.2019.04.015, 2019.

[8] A. Saha, A. Dutta and R. I. Sifat, The mental impact of digital divide due to COVID-19 pandemic induced emergency online learning at undergraduate level: Evidence from undergraduate students from Dhaka City, *J. Affect. Disord.*, vol.294, no.7, pp.170-179, DOI: 10.1016/j.jad.2021.07.045, 2021.

[9] F. Rahmad, Y. Suryanto and K. Ramli, Performance comparison of anti-spam technology using confusion matrix classification, *IOP Conf. Ser. Mater. Sci. Eng.*, vol.879, no.1, DOI: 10.1088/1757-899X/879/1/012076, 2020.

[10] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim and U. R. Acharya, Automated detection of COVID-19 cases using deep neural networks with X-ray images, *Comput. Biol. Med.*, vol.121, no.4, DOI: 10.1016/j.compbiomed.2020.103792, 2020.

[11] B. Sheng, O. M. Moosman, B. Del Pozo-Cruz, J. Del Pozo-Cruz, R. M. Alfonso-Rosa and Y. Zhang, A comparison of different machine learning algorithms, types and placements of activity monitors for physical activity classification, *Meas. J. Int. Meas. Confed.*, vol.154, DOI: 10.1016/j.measurement.2020.107480, 2020.

[12] M. Hasnain, M. F. Pasha, I. Ghani, M. Imran, M. Y. Alzahrani and R. Budiarto, Evaluating trust prediction and confusion matrix measures for web services ranking, *IEEE Access*, vol.8, pp.90847-90861, DOI: 10.1109/ACCESS.2020.2994222, 2020.

[13] Z. DeVries et al., Using a national surgical database to predict complications following posterior lumbar surgery and comparing the area under the curve and F1-score for the assessment of prognostic capability, *Spine J.*, vol.21, no.7, pp.1135-1142, DOI: 10.1016/j.spinee.2021.02.007, 2021.

[14] A. Priya, S. Garg and N. P. Tigga, Predicting anxiety, depression and stress in modern life using machine learning algorithms, *Procedia Comput. Sci.*, vol.167, pp.1258-1267, DOI: 10.1016/j.procs.2020.03.442, 2020.

[15] A. N. Priyanka and D. Kumar, Decision tree classifier: A detailed survey, *Int. J. Inf. Decis. Sci.*, vol.12, no.3, pp.246-260, DOI: 10.1504/ijids.2020.10029122, 2020.

[16] R. R. Maaliw III, Adaptive virtual learning environment based on learning styles for personalizing e-learning system: Design and implementation, *Int. J. Recent Technol. Eng.*, vol.8, no.6, pp.3398-3406, DOI: 10.35940/ijrte.f8901.038620, 2020.

[17] H. Waheed, S. U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani and R. Nawaz, Predicting academic performance of students from VLE big data using deep learning models, *Comput. Human Behav.*, vol.104, no.11, DOI: 10.1016/j.chb.2019.106189, 2020.

[18] F. Rasheed and A. Wahid, Learning style detection in e-learning systems using machine learning techniques, *Expert Syst. Appl.*, vol.174, no.9, DOI: 10.1016/j.eswa.2021.114774, 2021.

[19] Y. Yamasari, S. M. S. Nugroho, K. Yoshimoto, H. Takahashi and M. H. Purnomo, Expanding tree-based classifiers using meta-algorithm approach: An application for identifying students' cognitive level, *International Journal of Innovative Computing, Information and Control*, vol.15, no.6, pp.2085-2107, DOI: 10.24507/ijicic.15.06.2085, 2019.

[20] A. Harras, A. Tsuji, S. Karungaru and K. Terada, Enhanced vehicle classification using transfer learning and a novel duplication-based data augmentation technique, *International Journal of Innovative Computing, Information and Control*, vol.17, no.6, pp.2201-2216, DOI: 10.24507/ijicic.17.06.2201, 2021.

[21] Rianto, A. B. Mutiara, E. P. Wibowo and P. I. Santosa, Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation, *J. Big Data*, vol.8, no.1, DOI: 10.1186/s40537-021-00413-1, 2021.

[22] T. M. Ma, K. Yamamori and A. Thida, A comparative approach to Naïve Bayes classifier and support vector machine for email spam classification, *2020 IEEE 9th Global Conference on Consumer Electronics (GCCE)*, pp.324-326, DOI: 10.1109/GCCE50665.2020.9291921, 2020.

[23] M. A. Rahman and Y. A. Akter, Topic classification from text using decision tree, K-NN and multinomial Naïve Bayes, *The 1st Int. Conf. Adv. Sci. Eng. Robot. Technol. (ICASERT2019)*, DOI: 10.1109/ICASERT.2019.8934502, 2019.

[24] C. C. Aggarwal, *Data Classificaton, Algorithms and Applications*, CRC Press, A Chapman & Hill Book, Boca Raton, Florida, 2015.

[25] P. K. Kuiper and K. M. Hood, Examining sentiment analysis when evaluating survey responses, *2019 IEEE 13th International Conference on Semantic Computing (ICSC)*, pp.412-415, DOI: 10.1109/ICOSC.2019.8665581, 2019.

[26] V. Bahel, S. Pillai and M. Malhotra, A comparative study on various binary classification algorithms and their improved variant for optimal performance, *2020 IEEE Reg. 10 Symp. (TENSYMP2020)*, no.5, pp.495-498, DOI: 10.1109/TENSYMP50017.2020.9230877, 2020.

[27] K. Kim, A hybrid classification algorithm by subspace partitioning through semi-supervised decision tree, *Pattern Recognit.*, vol.60, pp.157-163, DOI: 10.1016/j.patcog.2016.04.016, 2016.

[28] R. Saravanan and P. Sujatha, Algorithms: A perspective of supervised learning approaches in data classification, *2018 2nd Int. Conf. Intell. Comput. Control Syst.*, pp.945-949, https://ieeexplore.ieee.org/abstract/document/8663155, 2018.

[29] I. G. A. Suciningsih, M. A. Hidayat and R. A. Hapsari, Comparation analysis of Naïve Bayes and decision tree C4.5 for Caesarean section prediction, *J. Soft Comput. Explor.*, vol.2, no.1, pp.46-52, DOI: 10.52465/joscex.v2i1.25, 2021.

[30] A. R. Dikananda, I. Ali, Fathurrohman, R. A. Rinaldi and Iin, Genre e-sport gaming tournament classification using machine learning technique based on decision tree, Naïve Bayes, and random forest algorithm, *IOP Conference Series: Materials Science and Engineering*, vol.1088, no.1, DOI: 10.1088/1757-899X/1088/1/012037, 2021.