



## ANALYZING STUDENTS' LEARNING STRATEGIES USING A NOVEL LINEAR DISCRIMINANT WITH VECTOR MACHINE (LDVM) CLASSIFICATION ALGORITHM

**Mr. Ch. Sasidhar Rao , Dr. A.S. Arunachalam**

Research Scholar, Department of Computer Science,  
Vels Institute of Science, Technology and Advanced Studies (VISTAS), Pallavaram, Chennai, Tamil Nadu, India- 600117.

E-Mail: [sdrteachings@gmail.com](mailto:sdrteachings@gmail.com)

Associate Professor, Vels Institute of Science, Technology and Advanced Studies (VISTAS),  
Pallavaram, Chennai, Tamil Nadu, India- 600117.

E-Mail: [arunachalam1976@gmail.com](mailto:arunachalam1976@gmail.com)

---

### Abstract

With the information gained through comprehensive education, the learning process is the most effective approach for moulding the future society. The majority of today's youngsters are habituated to a multitude of distractions that might impede learning and understanding. Many students use several ways to understand the information and grasp the notion. The pre-processing approach, which is effective for extracting the most relevant data from the acquired set, is used in the first step of educational data mining. The preceding part of this stage involves the use of framed questionnaires to analyze the learning approach used by pupils. The questions are designed in such a manner that each method used throughout the learning process is thoroughly reviewed. Queries are used to gather data utilizing visual (V), auditory (A), read/write (R), and kinaesthetic (K) technologies. This VARK information-gathering approach is frequently utilized in many fields of research to identify the present mentality of the people in a given place. The most common preprocessing approaches are principal component analyses (PCA) and the Histogram equation. The feature extraction procedure that follows preprocessing is useful for deciding what information is necessary. There are several learning models used in various research projects that are valuable for understanding the strategy used by students in academic work. Even though it studies learning approaches, it has some limitations in precisely determining the optimum learning strategy used by the learners. This study is extremely helpful in determining the most efficient learning analysis model and reliable technique for classifying approaches. The SMOTE labelling approach and the deployment of the K-mean algorithm for classification remain the basic steps in this study. In comparison to previous classification algorithms, the suggested LDVM classification technique for categorizing FLSM processed combinations analyses data with a 92 per cent accuracy.

**Keywords:** SMOTE, VARK, PCA, LDVM, FSLM.

---

**DOI: 10.48047/ecb/2023.12.6.303**

## 1. Introduction

The learning process is the most powerful process for shaping upcoming society with fulfilling knowledge gathered from holistic education. The young generations are mostly addicted to many kinds of disturbances, which may affect the process of learning and understanding the concept fully. Many learners follow many unique techniques in understanding the content and learning the concept through teachers. The strategy followed by an individual may be different from each other and the evaluation process followed for understanding the capability of a student remains a difficult task. Most of the strategy followed in classroom learning varies such as Flipped Instruction, Project-Based Learning, Play-Based Learning, Discovery Learning, Authentic Learning, Community Engagement learning, High Expectations, Modeled Teaching, Unconditional Positive Regard, Scaffolding, Direct Instruction, Guided Practice, Spaced Repetition, Repetition, Prompting, Manipulatives, Differentiation, Prior Knowledge Assessment, Peer Assisted Learning, Student-Teacher Conference, etc.

The techniques implemented in the evaluation process and understanding of the knowledge of students don't predict the technique the student used for learning. Educational data mining plays a major role in understanding the methodologies followed by the student in learning and understanding the concept. The categorization was made with the usage of a unique educational data mining technique to scrutinize the collected records from proper questionnaires. The stages followed in the educational data mining technique are mostly common in many research work, which differs only from the analyzing algorithm used for the research work. This research work focuses on giving a perfect solution to the student's learning process and gives an idea for developing a good teaching methodology.

The first stage of educational data mining starts with the pre-processing technique, which is useful in extracting the most relevant information from the collected data. The previous stage of this step is carried out with the usage of framed questionnaires for analyzing the strategy followed by the students in the learning process. The questions are framed in such a manner that every strategy followed in the learning process is examined thoroughly. The data collection using queries is carried out with the usage of VARK technology, which VARK stands for Visual (V), Auditory (A), Read/Write (R), and Kinaesthetic (K). This VARK technique for collecting information is widely used in many fields of research to identify the present mentality of the people in a specific area.

The principal component analyses (PCA) and Histogram equations are the most popular techniques used in the pre-processing technique. The feature extraction followed after the pre-processing technique is useful in identifying the necessary

The multiple intelligent-based learning techniques using neural networks proposed previously for identifying students' learning strategies have a few drawbacks such as accuracy problems, time extension problems, and problems in feature extraction. The discussed drawbacks in multiple intelligent-based techniques for acquiring students' information from the database can be solved with proper training and validation processes. The training set is collected from the overall student records acquired using an implementation of VARK technology.

The researchers are very keen on understanding the learning strategy followed by the students in academics. The problems that occur during the process of identification of the methodology followed in the learning process are very significantly distributed as follows. Few research works discussed about understandability of student learning strategy, which has few issues in the validation and testing part.

- The research work is also not much focused on acquiring accurate results and they also lack in explaining the precision, recall, and f measure for acquired student learning information.
- Most of the research work follows the FSLM learning style model in finding the learning strategy. Sometimes FSLM has a disadvantage in omitting two dimensions or one dimension while processing.
- Much of the learning research work follows FSLM and Kolb's learning style model rather than using Gardner's theory. Very few of the research works are useful in using Gardner's theory of multiple intelligences.

The data collected from various colleges by using the VARK methodology are very useful in understanding the learners' behavioural patterns followed in reading and understanding the concepts. Further, the objectives may lead to fulfilling the better prediction of the learner's mindset as follows.

To predict the earlier stages of Sensitivity, specificity, recall, precision, F1 score, and accuracy all the major measurements used in this research work for predicting the mind setup of the students in learning and predicting the amount of time spent in reading and understanding the logical concept.

To compare the proposed LDVM algorithm technique for the process of classifying the learning methods combinations is better than that of the other traditional method concerning certain performance analysis.

This paper gives a clear idea about learners' behaviour in learning, it also examines the best labelling methodology followed while pre-processing the collected data. The VARK methodology is followed in collecting the record from the student community and has many irrelevant data also. The questionnaire is framed accordingly to inspect the capacity of the students as well as the learning strategy followed by the learner.

The prediction process of learning strategy will give the best method in learning the concept as well as in selecting the perfect technology. The proposed LDVM is very useful in the classification of different combinations of methods with high accuracy levels. The comparisons made in this research work are very useful in finding out the significant difference between the existing and proposed methodologies.

## 2. Literature Survey

In recent years, the importance of e-learning has been recognized by experts in the learning process. E-learning tools enable learners to access learning materials from anywhere and at any time. This has led to the development of Learning Management Systems (LMS) which are integrated web-based learning environments and instructional tools. LMSs enable the creation and delivery of content, monitoring of student activity, and identification of student performance. LMSs also provide interactive learning capabilities such as conversations, video conferencing, and discussion forums. The Learning Management System consists of a Course Management System (CMS), Learning Content Management System (LCMS), Managed Learning Environment (MLE), Learning Support Systems, and Learning Platform. LMSs attempt to modify and customize the learning environment based on the learner's preferences.

Neil Fleming's VARK learning model proposes four modulator ways for determining an individual's learning style Visual Learning, Auditory Learning, Physical Learning, and Social Learning. Daoruang and Beesuda [6] investigated the impact of learning style prediction based on user attributes using VARK.

- Visual learners acquire knowledge through the use of real-time visual aids such as graphs, charts, diagrams, and symbols.
- Auditory learners like to learn through listening to lectures, debates, and tapes.
- Tactile/kinesthetic learners attempt to learn through real-world experiences such as

project work.

- Social learners like to learn through social skills such as reading and writing.

Classification is a well-known machine learning-based data mining approach.

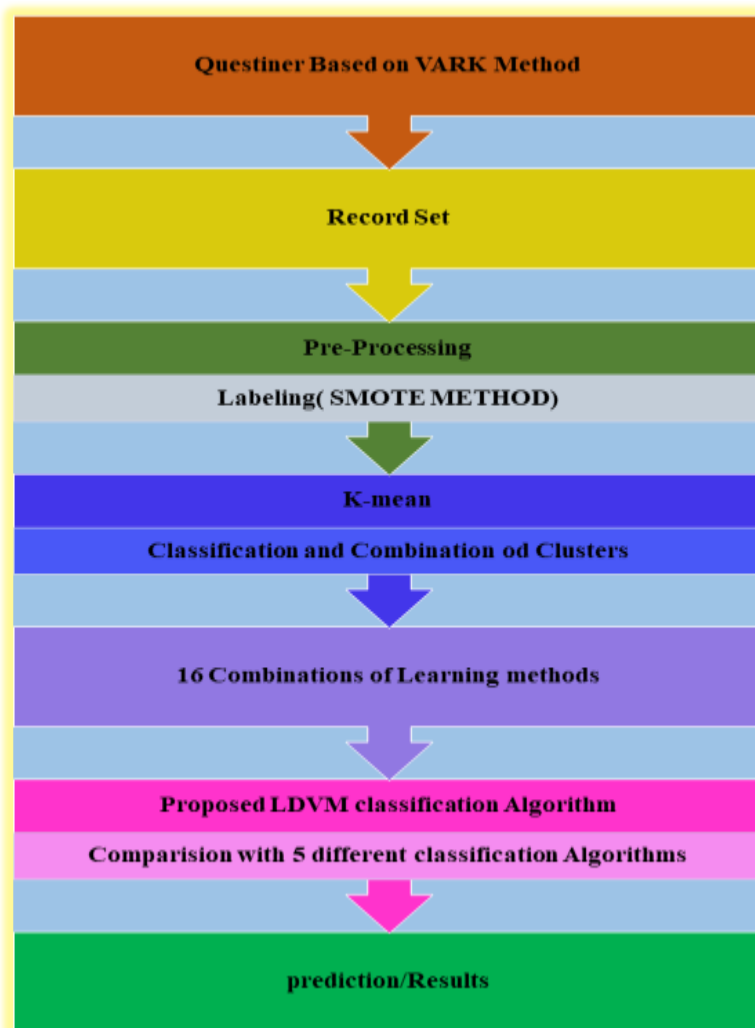
Classification is the process of categorizing each item in a collection of data into one of a preset set of classes or groupings. When applying data mining for customizing LMS, we must prioritize classification from big chunks of data submitted to the Learning Management System. J48 Decision trees, Nave Bayes, Support Vector Machines, and Random Forests are some of the most often used classification methods.

### 3. Methods and Materials

Student's learning techniques may differ from one another throughout their studies. In the learning process, data mining and other computational approaches are mainly employed to identify the optimum prediction strategy. Even though there are various ways to predict students' learning strategies, there are significant research gaps in accurate prediction and categorizing with a perfect classifier. The researchers also fall short of accurately analyzing the particular learning model used during the learning process.

Information on students from various streams and places is gathered. The required questionnaires are structured study techniques followed by students in learning viewpoint and evaluating the student's learning style and understanding of academic knowledge. The questionnaires are made up of various questions that are used to assess the student's learning processes.

The approach also included several supervisor-related questions for Learning techniques, technology, and learning models. The data set includes 12523 records of students who learned through several disciplines and took examinations based on their knowledge of the learning and teaching methodologies. The student record also investigates whether the student's Acceptance of Learning preference influences understandability and usability.



**Fig 2: Architecture of Proposed Methodology**

**The proposed methodology's architecture consists of the following steps:**

- The student record set collected from various departments is based on the VARK model.
- The information collected from various students is pre-processed using the SMOTE (Synthetic Minority Over-sampling Technique) technique to solve the unbalancing problem and improve the quality of the processed student record.
- The feature extraction process followed in this research work was selected based on the parameters that fully contribute to evaluating the performance of student learning strategy. The VARK model is very useful in collecting student information based on question setting.
- The parameter selection process followed in the research work is useful in extracting the most relevant information for the student learning aspect.

- The learning sessions are categorized based on the various learning objects accessed by the learner in the learning sequence. Every sequence followed in the learning session has a different sequential ID, learner ID, and learning object accessed during the learning process.
- The identification of learning sequences is followed by applying the classification process using FSLSM while assuming a separate learning style combination from other learning sequences.
- The following process is to give necessary sequential labelling for the training set for predicting the new learning style of the new sequence.
- The model training, testing, pattern identification, and evaluation of necessary results are the steps followed after the parameter selection process. The classification techniques are useful in building the training set. The testing set helps inspect the capacity of the built training set.
- The k-mean algorithm is used for ordering the assigned label to each sequence based on the FSLSM. Bayes classifier is used for the new learner identification process with existing learners according to the FSLSM.
- The obtained results are evaluated to identify the reason for learning strategy drawbacks.
- The efficiency of the prediction process followed in this research work is tested with different classification algorithms.
- K- Nearest Neighbour (KNN), Gaussian Naive Bayes (NB), Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), Classification and Regression Trees (CART), and Logistic Regression (LR) are the existing classification algorithms used for evaluating the proposed learning model.

### 3.1 Pre-Processing and Feature Extraction Process

The collected students learning records are taken for the pre-processing stage where all the unnecessary and irrelevant information is cleared. The records of collected learner information are categorized as follows

A: Active learners

R: Learners who use learning material for learning

Sen: Sensing Learners usually use materials with real-world applications and concrete factors of learning

Vi: Visual Learners like to visualize the learning materials with the representation of pictures, diagrams, and charts.

Ve: Verbal Learners use explanation of learning materials explained with words.

Seq: Sequential Learners follow a learning strategy of step by step learning process with better focus in a linear way.

G: Global Learners shapes the learning information holistically for better understanding.

The FLSM process followed in this research work consists of four dimensions where each dimension contains two opposite categories and each learner prefers a specific category for each dimension. The obtained combinations are as follows

Table 1. K- mean algorithm in different obtained combinations

S.No	Combinations	Cluster-ID	Number of sequences	Sensitivity	Specificity	Accuracy
1	(R, Sen, Ve, G)	C01	36	0.86500	0.97333	0.91917
2	(R,I,Ve,Seq)	C02	186	0.74000	1.09000	0.86500
3	(R, I, Vi, G)	C03	45	1.09000	0.97333	0.98167
4	(R,I,Vi,Seq)	C04	173	0.74000	0.98167	0.86083
5	(R, Sen, Ve, G)	C05	57	1.09000	0.97333	0.98167
6	(R,Sen,Ve,Seq)	C06	34	0.74000	0.98167	0.86083
7	(R, Sen, Vi, G)	C07	82	0.86500	0.98167	0.92333
8	(R,Sen,Vi,Seq)	C08	112	0.86500	1.09000	0.92750
9	(A, I, Ve, G)	C09	95	1.09000	0.96500	0.97750
10	(A,I,Ve, Seq)	C10	31	0.74000	1.09000	0.86500
11	(A, I, Vi, G)	C11	55	0.86500	1.09000	0.92750
12	(A,I,Vi,Seq)	C12	72	1.09000	1.09000	1.09000
13	(A, Sen, Ve, G)	C13	47	0.86500	0.98167	0.92333
14	(A,Sen,Ve,Seq)	C14	55	1.09000	1.09000	1.09000
15	(A, Sen, Vi, G)	C15	52	1.09000	1.98167	1.98583
16	(A,Sen,Vi,Seq)	C16	72	0.86500	1.09000	0.92750

The obtained combinations are undergone simulation for testing the accuracy with the K-mean clustering algorithm and the result is shown in Figure 2. The result is very clear to explain that cluster ID 15 has the highest accuracy than that of other clusters. The learning strategy followed by Active Learners (A), Sensing Learners (Sen), Visual Learners (Vi), and Global Learners (G) is mostly considered to be the best learning methodology.



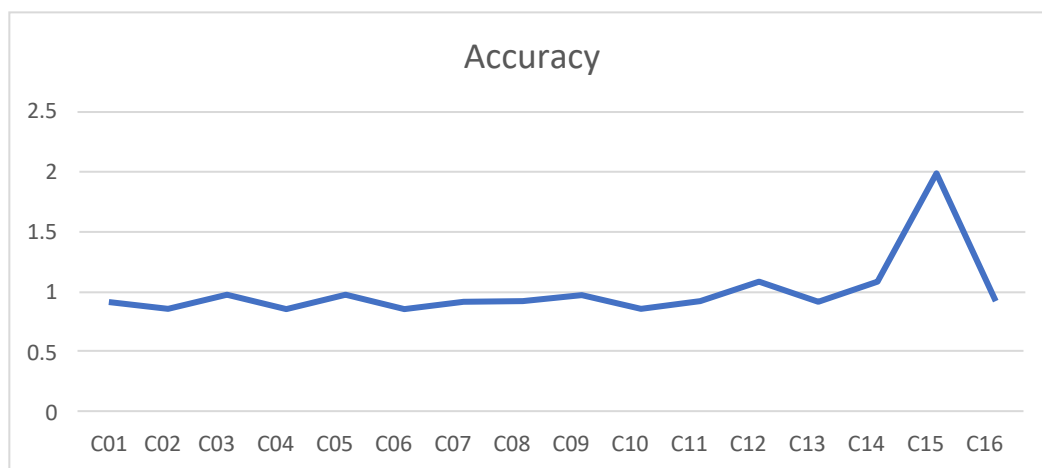


Figure 2. Accuracy of the obtained combinations

### 3.2 Proposed LDVM

The prediction process is tested with various classification algorithms with the proposed hybrid classification algorithm. K- Nearest Neighbour (KNN), Gaussian Naive Bayes (NB), Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), Classification and Regression Trees (CART) and Logistic Regression (LR) are the existing classification algorithms used for evaluating the proposed learning model. The proposed Linear Discriminant with Vector Machine (LDVM) is used for predicting the learning strategy followed by the learners.

The best model selection is carried out by simulating the obtained combinations with different classification algorithms and the proposed LDVM algorithm for selecting the best model for the learning strategy.

Table 2. LDVM algorithms measurement

Classification Algorithms	Precision	Recall	F1-Score	Kappa	Accuracy
LDVM	87.88	87.88	87.89	80.53	92

## 4. Experimental Results and Discussion

In this paper, different measurement methods were discussed to predict the learning strategy. The proposed LDVM was compared to other existing algorithms using standard measures such as Accuracy, Recall, Precision, Kappa, and F-measure. These measurements were used to test the quality of the classification model.

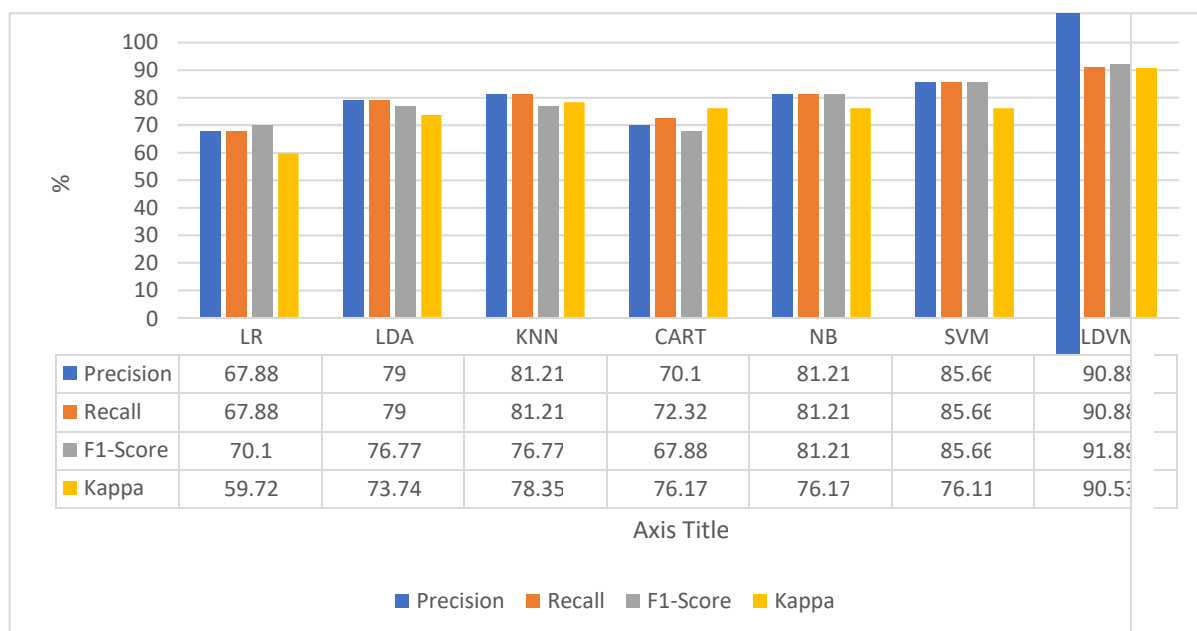
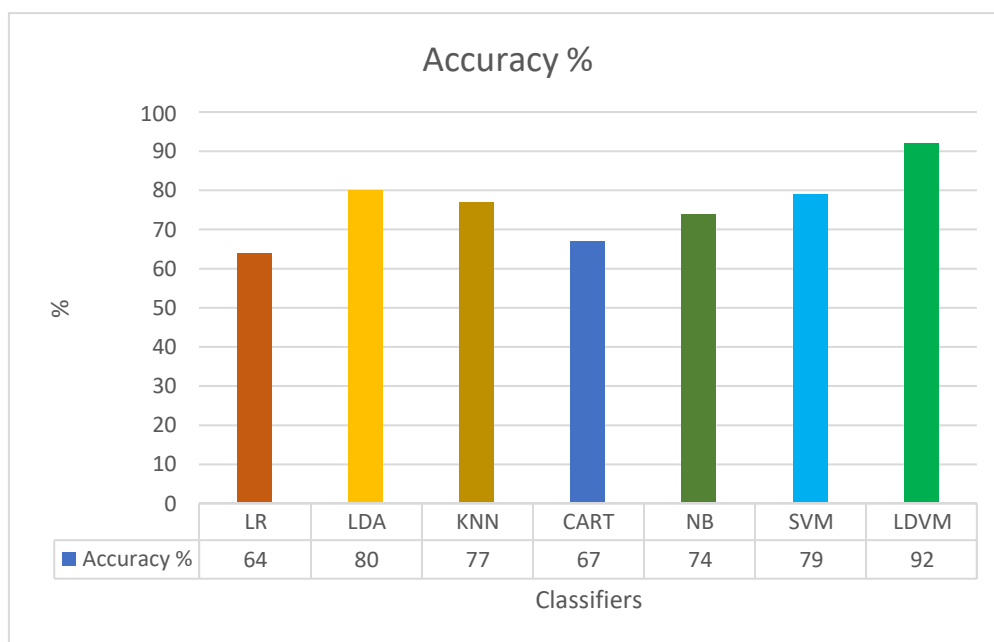


Figure 3 Classification Algorithms Measurement



The accuracy of the classification algorithms is also tested and stated in Figure 4. The proposed LDVM has a high accuracy compared with that of the other existing classification algorithms.

## 5. Conclusions

The prediction of the learning strategy followed by learners in academics always remains a challenging task. The research area dealing with academic performance also lacks in detecting the strategy followed by the learners in acquiring the knowledge. There are many learning models followed in many research works that are useful in understanding the technique followed by the learners in studies. Even though it inspects the learning techniques,

it has a few drawbacks in finding out the best learning strategy accurately followed by the learners. This research work is very helpful in finding out the best learning analyzing model with an accurate way of classifying the methods. The SMOTE technique involved in labelling and implementing the K-mean algorithm for the classification process remains the initial step involved in this research work. The proposed LDVM classification algorithm for classifying the FSLSM processed combinations processes the data accurately at 92% compared with other classification algorithms.

## References

- [1]. Aldahwan, N., and N. Alsaeed. "Use of artificial intelligence in Learning Management System (LMS): a systematic literature review." *International Journal of Computer Applications* 175, no. 13 (2020): 16-26.
- [2]. Da Silva, Nadia FF, Eduardo R. Hruschka, and Estevam R. Hruschka Jr. "Tweet sentiment analysis with classifier ensembles." *Decision Support Systems* 66 (2014): 170-179.
- [3]. Hasibuan, Muhammad Said, Lukito Edi Nugroho, and Paulus Insap Santosa. "Model Detecting Learning Styles with Artificial Neural Network." *Journal of Technology and Science Education* 9, no. 1 (2019): 85-95.
- [4]. Troussas, Christos, Akrivi Krouska, Cleo Sgouropoulou, and Ioannis Voyiatzis. "Ensemble learning using fuzzy weights to improve learning style identification for adapted instructional routines." *Entropy* 22, no. 7 (2020): 735.
- [5]. Daoruang, Beesuda, Krich Sintanakul, and Anirach Mingkhwan. "The study of learning achievement of learners classified VARK learning style in blended learning." In *Proceedings of the 2019 3rd International Conference on Natural Language Processing and Information Retrieval*, pp. 34-38. 2019.
- [6]. Daoruang, Beesuda, Krich Sintanakul, and Anirach Mingkhwan. "The study of learning achievement of learners classified VARK learning style in blended learning." In *Proceedings of the 2019 3rd International Conference on Natural Language Processing and Information Retrieval*, pp. 34-38. 2019.
- [7]. Rasheed, Fareeha, and Abdul Wahid. "Learning style detection in E-learning systems using machine learning techniques." *Expert Systems with Applications* 174 (2021): 114774.
- [8]. Singh, Asmita, and R. Lakshminathan. "Impact of different data types on classifier performance of random forest, naive bayes, and k-nearest neighbors algorithms." (2018).
- [9]. Francis, Bindhia K., and Suvanam Sasidhar Babu. "Predicting academic performance of students using a hybrid data mining approach." *Journal of Medical Systems* 43, no. 6 (2019): 1-15.
- [10]. Hasibuan, M. S., L. E. Nugroho, and P. I. Santosa. "Detecting Learning Style Based on Level of Knowledge." In *2018 Third International Conference on Informatics and Computing (ICIC)*, pp. 1-5. IEEE, 2018.
- [11]. Han, Bo, Paul Cook, and Timothy Baldwin. "A stacking-based approach to Twitter user geolocation prediction." In *Proceedings of the 51st annual meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 7-12. 2013.
- [12]. Marakas, George M. *Decision support systems in the 21st century*. Vol. 134. Upper Saddle River, NJ: Prentice Hall, 2003.
- [13]. Rao, Ch Sasidhar. "Personalization of Learning Management System using VARK." *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12, no. 7 (2021): 1538-1545.
- [14]. Dong, Xibin, Zhiwen Yu, Wenming Cao, Yifan Shi, and Qianli Ma. "A survey on ensemble learning." *Frontiers of Computer Science* 14, no. 2 (2020): 241-258.
- [15]. Liang, Guohua, Xingquan Zhu, and Chengqi Zhang. "An empirical study of bagging predictors for different learning algorithms." In *Twenty-Fifth AAAI Conference on Artificial Intelligence*. 2011.
- [16]. Džeroski, Saso, and Bernard Ženko. "Is combining classifiers with stacking better than selecting the best one?." *Machine learning* 54, no. 3 (2004): 255-273.
- [17]. Schapire, Robert E. "A brief introduction to boosting." In *Ijcai*, vol. 99, pp. 1401-1406. 1999.

- [18]. Aleksandrova, Yanka. "Predicting Students Performance in Moodle Platforms Using Machine Learning Algorithms." In *Conferences of the Department Informatics*, no. 1, pp. 177-187. Publishing house Science and Economics Varna, 2019.
- [19]. Kondo, Nobuhiko, Midori Okubo, and Toshiharu Hatanaka. "Early detection of at-risk students using machine learning based on LMS log data." In *2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, pp. 198-201. IEEE, 2017.
- [20]. Chen, Tianqi, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, and Hyunsu Cho. "Xgboost: extreme gradient boosting." *R package version 0.4-2 1*, no. 4 (2015): 1-4.
- [21]. Alhassan, Amal, Bassam Zafar, and Ahmed Mueen. "Predict Students' Academic Performance based on their Assessment Grades and Online Activity Data." *International Journal of Advanced Computer Science and Applications (IJACSA)* 11, no. 4 (2020).
- [22]. Liu, Yun-Fu, Jing-Ming Guo, and Jiann-Der Lee. "Halftone image classification using LMS algorithm and naive Bayes." *IEEE Transactions on Image Processing* 20, no. 10 (2011): 2837-2847.
- [23]. Rajper, Samina, Noor A. Shaikh, Zubair A. Shaikh, and G. Ali Mallah. Automatic detection of learning styles on learning management systems using data mining technique. *Indian Journal of Science and Technology* 9, no.1-5, 15, 2016.
- [24]. Baran, Evrim. Connect, participate and learn: Transforming pedagogies in higher education. *Bulletin of the IEEE Technical Committee on Learning Technology* 15, no. 1, 9-12, 2013.
- [25]. Dinçol, Sinem, SenarTemel,zgeOskay, Ümit I. Erdoğan, and Ayhan Yılmaz. The effect of matching learning styles with teaching styles on success. *Procedia-Social and Behavioral Sciences* 15,854-858, 2011.
- [26]. Solaskar, Supriya, and A. K. Sen. Improving learning management system using data mining. *International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 4 Issue 02*, 2015.
- [27]. Evale, Digna. Learning management system with prediction model and course-content recommendation module. *Journal of Information Technology Education: Research* 16, no. 1 437-45, 2016.
- [28]. Blagojević, Marija, and ŽivadinMicić. A web-based intelligent report e-learning system using data mining techniques. *Computers & Electrical Engineering* 39, no. 2, 465-474, 2013.
- [29]. Hong, Wonjoon, and M. Bernacki. A prediction and early alert model using learning management system data and grounded in learning science theory. In *Workshop and Tutorials Chairs*, no. 182, p. 358. 2016.
- [30]. Saroha, Kriti, and Phalit Mehta. Analysis and evaluation of learning management system using data mining techniques. In *2016 International Conference on Recent Trends in Information Technology (ICRTIT)*, pp. 1-5. IEEE, 2016.
- [31]. Hasibuan, M. S., Lukito Edi Nugroho, Paulus InsapSantosa, and S. S. Kusumawardani. A Proposed Model for Detecting Learning Styles Based on Agent Learning. *International Journal of Emerging Technologies in Learning* 11, no. 10 2016.