Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

A Novel Data Mining Approach to Improve Learning Capacity and Academic Success in High-Risk Postgraduate Students in Distance Learning

Onkar Bagaria[1], Tarang Bhatnagar[2], Anushka Sharma[3], Dr. Madhavi R[4], Jagtej Singh[5], Dr. Sweta Kumari[6]

[1]Assistant Professor, Department of Management Studies, Vivekananda Global University, Jaipur, India, Email Id- bagaria.omkar@vgu.ac.in
[2]Centre of Research Impact and Outcome, Chitkara University, Rajpura- 140417, Punjab, India, Email Id- tarang.bhatnagar.rop@chitkara.edu.in
[3]Assistant Professor, Department of Computer Science & Engineering, Sanskriti University, Mathura, Uttar Pradesh, India, Email Id- anushka@sanskriti.edu.in
[4]Professor, Department of Finance, JAIN (Deemed-to-be Univesity), Bangalore, Karnataka, India, Email Id- dr.madhavi@cms.ac.in
[5]Chitkara Centre for Research and Development, Chitkara University, Himachal Pradesh-174103 India, Email Id- jagtej.singh.orp@chitkara.edu.in
[6]Assistant Professor, Department of ISME, ATLAS SkillTech University, Mumbai, Maharashta, India, Email Id- sweta.kumari@atlasuniversity.edu.in

Introduction: Data mining (DM) is valuable in numerous academic disciplines. Since student data can be used to recognize important concept associated to student education behavior, education study is increasing rapidly. Educational Data mining (EDM) can assist institution to distinguish students' successes by analyze through achievement.
Methods: This study address the classification of a DM methods, the Dynamic Particle Swarm Optimized- Discrete Support Vector Machine (DPSO-DSVM), to advance the knowledge ability and academic attainment of high-risk post-Graduate (PG) students in distance learning program. The Gujarat University PG student dataset was gathered. Clean, transform, reduce and Feature select the data.
Results: The research used the DPSO-DSVM algorithm to improve the model's prediction performance and its capacity to detect at-risk students in their academic careers. Performance indicators such as precision, recall, loss and accuracy were examined to determine the DPSO-DSVM method's superiority in predicting academic performance over conventional techniques.
Conclusions: This innovative DM method can improve the educational results of high-risk PG distance learning students.

Keywords: Data Mining (DM), Distance Learning, Postgraduate Students, Dynamic Particle Swarm Optimized-Discrete Support Vector Machine (DPSO-DSVM).

1. Introduction

Data mining benefits several academic fields. Since student data can provide significant notions about educational practices, education research is growing rapidly. EDM helps organizations to identify students' triumphs by assessing their performance. Classification is used to evaluate learning outcomes.

Distance-learning Post-Graduates

Universities provide distance education courses in online to improve educational settings via flexibility, independent study, security, creativity, dynamic input and instructional services. [(1)] Educational institution student accomplishment is a key indication of program academic results. Universities must utilize evaluating results to improve student learning and advancement. Exams, tasks and its activities measure the students' understanding and learning. [(2)] DM tools allow instructors to identify academic abilities and weaknesses as well as engagement, communication and study patterns that can be affecting student performance. [(3)] In education, evaluations are crucial to evaluate student performance. Students are tested to determine their comprehension and dedication to learning. [(4)] Academic studies have enhanced using DM approaches. This technology lets you analyze education data to find patterns that drive decision-making and strategy creation. [(5)]

Educational Data Mining (EDM)

Data mining approaches can potentially facilitate the development of Machine learning (ML) models by using the available data. Using techniques for data mining on such data is sometimes referred to as EDM. EDM examines new approaches, develops and implements sophisticated procedures on large datasets. This analytical technique seeks to uncover important patterns that clarify students' behavior and academic abilities. Thus, academic DM allows for novel solutions to higher education difficulties. [(6)] EDM has uncovered essential details and patterns in massive educational data via different methods. The process includes DM. Today, ML classification and trend analysis predict academic outcomes. Feature group, data depth and variation affect the predictability. [(7)] EDM creates a lot of data that can predict students' understanding. EDM uses statistics, Information Technology (IT), ML, DM, Artificial Intelligence (AI) and database management systems to study education.[(8)] Social network analysis (SNA), psychological pedagogy, neuroscience, psychological testing and visual DM are needed for EDM. EDM combines scientific computation, teaching and analytics. [(9)]

According to the author of, [(10)] investigated online educational data analysis and extraction approaches to forecast how students performed after the semester. In addition, the researchers used data conversion and cleaning procedures to decrease the number of characteristics. To examine the achievements of 3518 higher education learners who were enrolled in an educational management system and they were fully engaged in their studies [(11)]. The determined student involvement with educational panels as a predicted consequence of a web-based educational experience and to what degree these communication statistics can be utilized

[12] to forecast or give assistance via academic achievement.

To discuss the possible application of sophisticated ML algorithms [13] in academic contexts, considering the standpoint of hyper parameter optimization. Auto ML was used to assess pupil progress on educational domains in the study [14]. The explored and compared modern ML methods with supervised learning for student test-performance estimation, like identifying students at a "significant risk" of sending out and estimating their final test rankings. To indicate a supervised method of learning a Decision Tree (DT) [15] classification oriented on student outcomes projection mechanism and collaboration improved classifier efficiency.

Contribution of the study

- The work aims to improve academic performance and distant learning effectiveness for high-risk PG students.
- The Research emphasizes DM capacity to uncover academic behavior tendencies.
- The Research highlights importance of data cleaning, transformation, reduction and feature selection.
- The study compares DPSO-DSVM performance to traditional DM techniques, assessing precision, recall, loss and accuracy.

This study is divided into parts: Methodology (Part 2), Results (Part 3), Discussions (Part 4) and Conclusion (Part 5).

2. Methodology

This section describes the DM approach for improving learning capacity and academic success in high-risk PG distance learning students. Learning outcomes for, at risk PG distance learning students are boosted by DM. The technique employs systematic data analysis and data-driven learning to help high-risk learning. Figure 1 shows the proposed method flowchart.

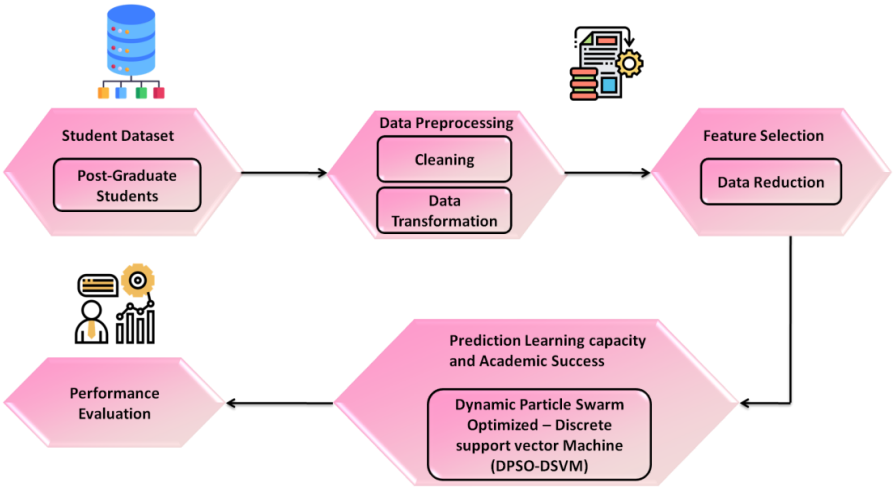

Figure 1. Flowchart of proposed method (Source: Author)

2.1. Dataset

Observational research [16] was conducted where 132 PG from Gujarat University's physical therapy departments were recruited utilizing convenient sampling. Learners who refuse to engage are barred. The institute's director granted the authorization to perform the research. The research is described to the individuals, who receive an inquiry form to complete. 100 students fill out the assessment.

2.2. Data Preprocessing

One of the most essential processes for maintaining the accuracy and reliability of data involves the purification of data, when examining academic performance and associated variables among PG students. When considering PG students, data cleaning includes numerous necessary methods to preserve data value.

Initially, it is crucial to detect and address any disparities, inaccuracies, or contradictions in data about academic achievements and scholarly endeavors, along with other pertinent factors. It's important to properly handle the missing data ratings, utilizing feature or, if necessary, removal.

2.2.1. Data Transformation

Data transformation allow academic to adjust and redistribute initial data to improved meet their study, making it significant for assess higher education result. This data alteration approach includes normalization, principles and connected element manufacture. These methods offer complete and precise depiction of instructive achievement to improve the general impact. Data is normalized to establish equal levels for dissimilar variables. This allows fair examination and avoids a single parameter level from disturbing the evaluation consistency gives statistics a mean of 0 and a standard deviation of 1. These upgrades expose hidden correlations the length of with example and make possible robust system that can dependably predict academic consequences as well as pressure evidence-based knowledge choice.

2.3. Feature Selection

The feature selection procedures perform a necessary constituent in investigative the educational achievement of PG learner, as it allows for a complete information of the main factor that impact their presentation. Selecting the major important and applicable fundamentals are critical, regardless of a diversity of influence, counting demographics, academic environment, and socioeconomic position and character behavior. Academics can choose during many candidate characters to find the major significant ones using metrics like association study, regression model and ML. This systematic method helps to explain the difficult mechanism behind student academic achievement that build better aid as well as encourage program to recover their presentation and knowledge.

2.3.1. Data Reduction

Data reductions simplify complex data set while maintenance critical data, creation it significant in PG educational achievement investigates. This alternative uses data decrease to make simpler the record. It is done by select or creates a set of individuality that captures the spirit of student achievement. Principal Component Analysis (PCA) or factoring could

decrease several most probably linked variables. Gathering mechanism into smaller quantity divide variables makes data examination easier.

Data reduction methods let researchers identify key patterns and patterns in the dataset, enabling them to focus on the most important factors affecting academic success. This streamlined approach helps to examine significant correlations between several factors and educational outcomes. It makes it easier to create short programs that capture PG learning results. Data reduction procedures can eliminate redundant or irrelevant features, reducing the risk of over fit and enhancing study generalization.

2.4. Dynamic Particle Swarm Optimized-Discrete Support Vector Machine (DPSO-DSVM)

DPSO-DSVM is a hybrid intelligence technique that combines DPSO and DSVM. It solves complex optimization issues and classifies data sets. This hybrid solution performs when optimization or classification methods fails. Figure 3 illustrates the flow chart of DPSO-DSVM.

2.4.1. Discrete Support Vector Machine (DSVM)

DSVM are supervised learning algorithms that classify data points into different types. It uses a virtual plane to divide the dataset into units to optimize edge width. Parallel lines can be created by using this tool. The edge refers to the greatest gap in the closest measurements with the categories. To decrease the generalization mistake, the greatest gap is used. DSVM functions as follows: Figure 2 depicts flowchart of DSVM.

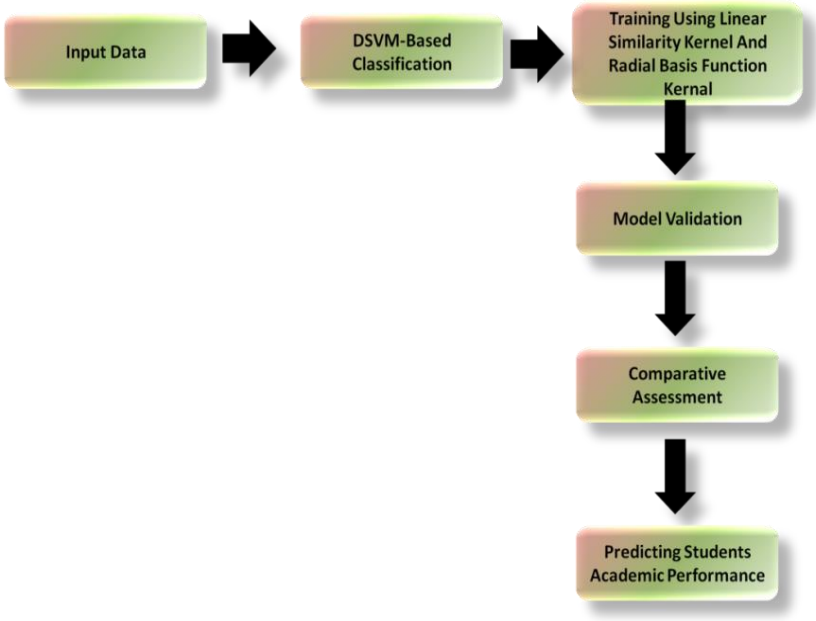

Figure 2. Flowchart of DSVM (Source: Author)

i. Data can be linearly segregated with perfect precision in the separable entity. Many boundaries can be envisaged and the method seeks the hyper plane that maximizes edge separation. The function provided is as follows in Equation (1-3):

$$e(z) = w.z + y, \quad (1)$$

DSVM separates statistics into:

$$e(z) > 0, \text{if } e(z) \in W, \text{and} \quad (2)$$

$$e(z) <= 0 \text{ if } e(z) \in Y \quad (3)$$

The separation from the observer to the hyper plane is calculated as $\frac{|w.z+y|}{||w||}$and the edge is represented as $\frac{2}{||w||}$.

The data points are divided by DSVM:

ii. Inability to distinguish data points. DSVM uses the transformation function (∅) to rearrange data for categorization. Projecting the integer mark result of the points there increases data dimensionality, making the approach successful.

2.4.2. Dynamic Particle Swarm Optimization (DPSO)

The swarming has influenced DPSO. By using swarm behavior principles, swarm performance can fix complex problems. DPSO method gives improved global search results but is imperfect. Evolutionary approaches like DPSO assess optimized survival value. Exploitation involves finding the ideal locations. Exploitation is important to studying intriguing subjects with skill. Fine-tuning factors makes DPSO algorithms work. Wealth and speed quality are balanced by inertial forces in the DPSO algorithm. The DPSO algorithm relies on an archive to enhance outcomes. DPSO solves enormous computer optimization issues owing to quick solutions. DPSO creatures have three primary elements: speed (see Equation 4-5), intellectual and sociable. While the particle updates its location, an additional speed path is computed as follows:

$$vel_j(s+1) = vel_j(s) + D_1 * q_1(s)\left(o_{wbest(j)}(s) - pop_j(s)\right) + D_2 * q_2(s)\left(h_{wbest}(s) - pop_j(s)\right) \quad (4)$$

As seen in Equation (4), the tool's location is dynamically altering.

$$pop_j(s+1) = pop_j(s) + vel_j(s+1) \quad (5)$$

According to the equations (4-5), $pop_j(s)$ symbolize jth substances time-limit location $(s)vel_j(s)$ corresponds to the speed of every single substance x termed inertia weight (IW), $D_1 and D_2$ stand for mental and social aspects. Both q_1 and $q_2$are uniformly random numbers in the range $0-1$ inclusive. In addition, the greatest location of the character in the regional search area is represented by one $o_{wbest(j)}(s)$, where$h_{wbest}(s)$represents the greatest spot in the global search region.

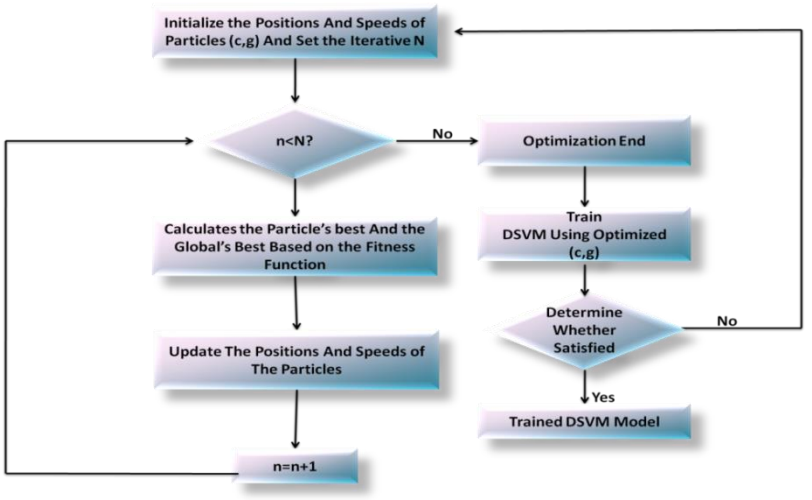

Figure 3. Flowchart of DPSO-DSVM (Source: Author)

3. Results

This section describes a DM approach to improve learning capacity and academic success in high-risk PG distance learning students. To assess the effectiveness of our novel DPSO-DSVM approach, we compared it to known methods such as Support Vector Machine (SVM) [17], Logistic Regression (LR) [17], and Artificial Neural Network (ANN) [17]. The outcomes of the assessment metrics are Recall, Accuracy, Precision and Loss.

The accuracy of DM models and algorithms in predicting whether a student would struggle academically or benefit from an intervention is measured. Its predicting how closely analytical data and remedies fit high-risk students' distance-learning course results and academic achievements. Figure 4 (a) and (b) depicts the accuracy and precision of the proposed and existing methods. Table 1 shows the outcomes of the proposed and existing methods. As evaluated by other methods, the proposed DPSO-DSVM had the highest accuracy of 91.52 %. Accuracy values for existing techniques involving ANN [17], SVM [17] and LR [17] had been 84.48 %, 66.94 % and 81.95 %, respectively.

Precision assesses DM and predictive model accuracy to identify high-risk PG distant learning students who can need further support. Precision measures how many students are at-risk, the DM methods successfully identifies. The proposed DPSO-DSVM had the highest precision of 0.89 %. The precision rates for existing techniques, including ANN [17], SVM [17] and LR [17], were 0.86 %, 0.44 % and 0.80 % accordingly.

Table 1. Outcome of Performance metrics (Source: Author)

Methods	SVM	LR	ANN	DPSO-DSVM [Proposed]
Accuracy (%)	66.94	81.95	84.48	91.52
Precision (%)	0.44	0.80	0.86	0.89
Recall (%)	0.69	0.81	0.617	0.88
Loss (%)	0.340	0.182	0.385	0.151

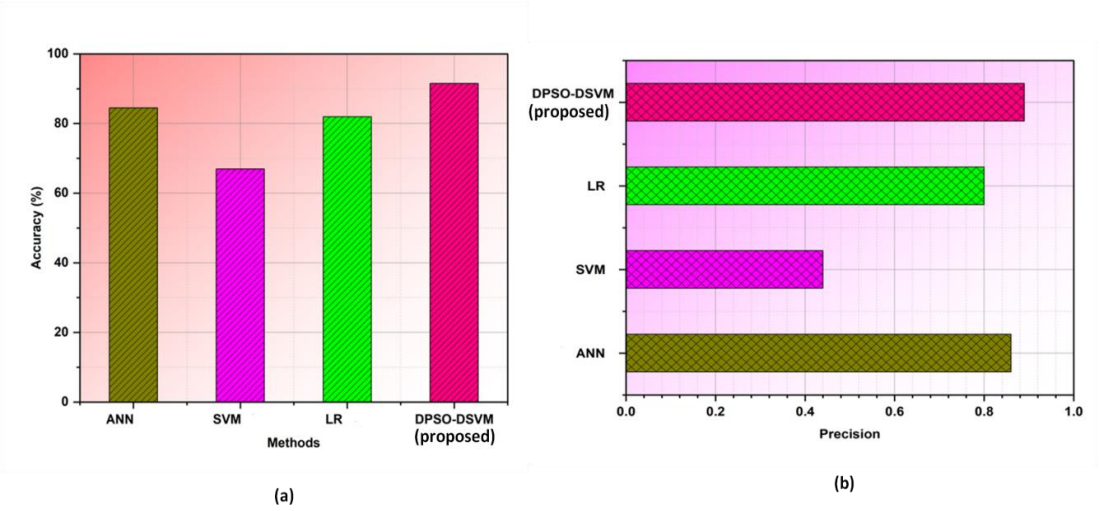

Figure 4. Outcome of (a) Accuracy and (b) Precision (Source: Author)

Loss is crucial to assess DM prediction models and remedies. It conveys the expense of mispredicting or misstating at-risk students and their impact on academic support. The proposed DPSO-DSVM had an outstanding Loss score of 0.151 %. The loss outcomes, recall along with loss of the proposed as well as existing methods is shown in Figures 5 (a) and (b). The loss percentages for current approaches such as ANN [17], SVM [17] and LR [17] were 0.385 %, 0.340 % and 0.182 % respectively.

Recall known as subjected or real-positive level that tests a DM models and algorithms' ability to identify high-risk PG remote learning students who need support and compared to other methods, the proposed DPSO-DSVM had an outstanding recall score of 0.88 %. Existing approaches such as ANN [17], SVM [17] and LR [17] display recall percentages of 0.617 %, 0.69 % and 0.81 % respectively.

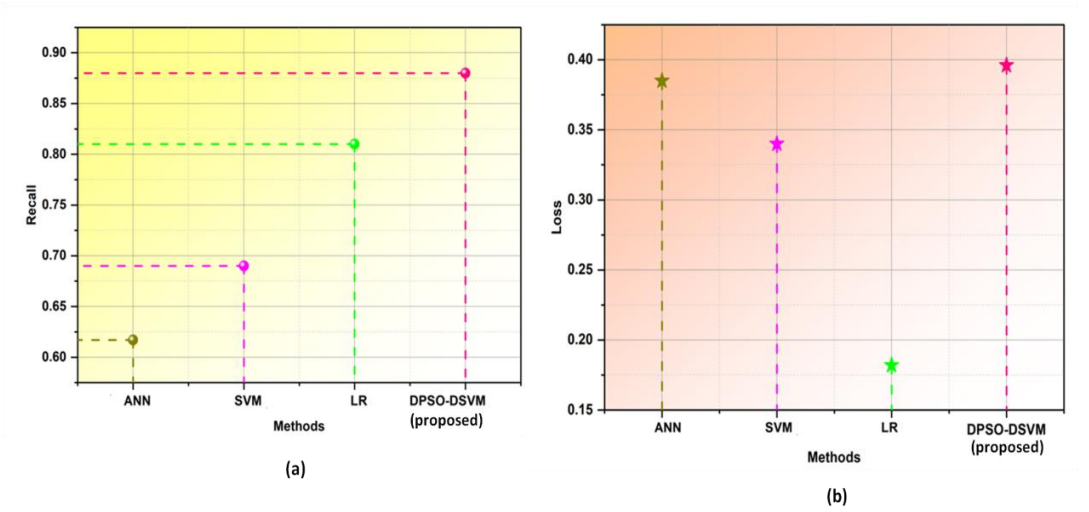

Figure 5. Result of (a) Recall and (b) Loss (Source: Author)

4. Discussion

This study adds to forecast the academic achievement of distance-learning PG students using a DM technique by identifying crucial characteristics influencing a student's academic performance. Massive student data has accelerated educational research, enabling institutions to employ educational DM to evaluate learning outcomes while honoring their achievements. DPSO-DSVM, a DM method to increase high-risk distant learning graduate students' learning capacity and academic outcomes, completed an extensive questionnaire. Cleaning, transformation, reduction and feature selection provide dataset accuracy and reliability for study. This strategy enhances learning path prediction and early detection of at-risk students. To assess whether DPSO-DSVM predicted academic success better than previous approaches, precision, recall, loss and accuracy were examined. High-risk PG distance learning students are improved academically with the significant DM approach. Utilizing the DPSO-DSVM algorithm and academic data, it has developed individualized treatments and support indicates to help at-risk students succeed academically and learn better.

5. Conclusion

Research concludes that DM has transformed student learning and academic achievement. Students' successes get analyzed and recognized using DM methods because to the quantity of student data. Model predictive and early student risk detection is improved by DPSO-DSVM. The research shows that the DPSO-DSVM approach predicts academic achievement better than standard methods using precision (0.89 %), recall (0.88 %), loss (0.151 %) and accuracy (91.52 %). The DPSO-DSVM algorithm and outcome measures offer targeted aid and support for high-risk students, enhancing their academic performance and learning ability. Future studies can help high-risk PG students in distant learning programs to succeed academically that enhance higher education results and student support.

References

1. Nguyen Q, Rienties B, Richardson JT. Learning analytics to uncover inequality in behavioral engagement and academic attainment in a distance learning setting. Assessment & Evaluation in Higher Education. 2020:594-606. https://doi.org/10.1080/02602938.2019.1679088.
2. Olaleye TO, Vincent OR. A predictive model for students' performance and risk level indicators using machine learning. In2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS) 2020;7. https://doi.org/10.1109/ICMCECS47690.2020.240897.
3. Ofori F, Maina E, Gitonga R. Using machine learning algorithms to predict students' performance and improve learning outcome: A literature review. Journal of Information and Technology. 2020; 33-55. https://doi.org/10.105217/068-3-130-060213-6_127.
4. Paramita AS, Tjahjono LM. Implementing machine learning techniques for predicting student performance in an e-learning environment. International Journal of Informatics and Information Systems. 2020; 149-56. https://doi.org/10.47738/ijiis.v4i2.112.
5. Kukkar A, Mohana R, Sharma A, Nayyar A. Prediction of student academic performance based on their emotional wellbeing and interaction on various e-learning platforms.

Education and Information Technologies. 2023; 1-30. https://doi.org/10.1007/s10639-022-11573-9.

6. Singh R, Pal S. Machine learning algorithms and ensemble technique to improve prediction of student's performance. IJATCSE. 2020;9(3). ine at https://doi.org/10.30534/ijatcse/2020/221932020.
7. Verma BK, Singh HK, Srivastava N. Prediction of Students' Performance in e-Learning Environment using Data Mining/Machine Learning Techniques. J. Univ. Shanghai Sci. Technol.. 2021; 569-93. http://doi.org/10.51201/JUSST/21/05179.
8. Saleh MA, Palaniappan S, Abdalla NA. Education is an overview of data mining and the ability to predict the performance of students. Edukasi. 2021; 19-28. https://doi.org/ 10.15294/ edukasi. v15i1.3 0065.
9. Ang, K.L.M., Ge, F.L. and Seng, K.P., 2020. Big educational data & analytics: Survey, architecture and challenges. IEEE access, 8, pp.116392-116414. https://doi.org/10.1109/ACCESS.2020.2994561
10. Hasan R, Palaniappan S, Mahmood S, Abbas A, Sarker KU, Sattar MU. Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. Applied Sciences. 2020; 3894. https://doi.org/10.3390/app10113894.
11. Aydoğdu Ş. Predicting student final performance using artificial neural networks in online learning environments. Education and Information Technologies. 2020; 1913-27. https://doi.org/10.1007/s10639-019-10053-x.
12. Kokoç M, Altun A. Effects of learner interaction with learning dashboards on academic performance in an e-learning environment. Behaviour & Information Technology. 2021; 161-75. https://doi.org/10.10 80/0144929X.2019.1680731.
13. Tsiakmaki M, Kostopoulos G, Kotsiantis S, Ragos O. Implementing AutoML in educational data mining for prediction tasks. Applied Sciences. 2019 Dec 20;10. https://doi.org/10.3390/app10010090.
14. Tomasevic N, Gvozdenovic N, Vranes S. An overview and comparison of supervised data mining techniques for student exam performance prediction. Computers & education. 2020; 103676. https://doi.org/10.1016/j.compedu.2019.103676.
15. Imran M, Latif S, Mehmood D, Shah MS. Student academic performance prediction using supervised learning techniques. International Journal of Emerging Technologies in Learning. 2019; 14(14). https://doi.org/10.3991/ijet.v14i14.10310.
16. Bhatt CJ, Sheth MS. Knowledge, attitude and practice towards evidence based practice in post graduate physiotherapy students. Int J Health Sci Res. 2021; 17-26. https://doi.org/10.52403/ijhsr.20210804.
17. Waheed H, Hassan SU, Aljohani NR, Hardman J, Alelyani S, Nawaz R. Predicting academic performance of students from VLE big data using deep learning models. Computers in Human behavior. 2020; 106189. https://doi.org/10.1016/j.chb.2019.106189.