Students' Performance Prediction in Higher Education Using Multi-Agent Framework-Based Distributed Data Mining Approach:
A Review

M. Nazir, Universiti Malaysia Pahang Al Sultan Abdullah, Pahang, Malaysia*

A. Noraziah, Universiti Malaysia Pahang Al Sultan Abdullah, Pahang, Malaysia

https://orcid.org/0000-0002-0419-7213

M. Rahmah, Universiti Malaysia Pahang Al Sultan Abdullah, Pahang, Malaysia

ABSTRACT

An effective educational program warrants the inclusion of an innovative construction that enhances the higher education efficacy in such a way that accelerates the achievement of desired results and reduces the risk of failures. Educational decision support system has currently been a hot topic in educational systems, facilitating the pupil result monitoring and evaluation to be performed during their development. In this literature survey, the authors have discussed the importance of multi-agent systems and comparative machine learning approaches in EDSS development. They explored the relationship between machine learning and multiagent intelligent systems in literature to conclude their effectiveness in student performance prediction paradigm. They used the PRISMA model for the literature review process. They finalized 18 articles published between 2014-2022 for the survey that match the research objectives.

KEYWORDS

Classification, Computational Intelligence, Distributed Data Mining, Educational Data Mining, Higher Education, Multi-Agent System, Prediction

INTRODUCTION

Predicting students' performance with a reasonable degree of accuracy is beneficial in finding the students who perform poorly when the learning process begins. The main objective of any educational institution is to render the best chances for education and skills for the students. To reach this goal, it is essential to recognize that students need extra help and meaningful steps to improve their results.

Malaysia is experiencing a moderate increase in the unemployment rate each year. The projected unemployment rate for 2017 is 3.42221%. Future unemployment will rise slightly over the next ten years. Then the remaining four years began to grow at a rate of at least 3.8 per cent from 2023

to 2026 (Ramli et al., 2018). This is a highly alarming figure, given that the findings indicate that unemployed graduates from public universities have the highest unemployment rate. However, the education sector has also revealed that unemployment has increased. The first objective is to analyze the critical variables affecting Malaysia's unemployment rate. According to the results, general factors such as inflation and population growth in Malaysia significantly impact the unemployment rate. In short, graduate students need to understand the situation and prepare for the uncertainty associated with this unemployment. Governments must also be accountable for taking appropriate action to tackle unemployment and not affect other social and economic conditions.

Today, due to the tremendous importance of this topic for the advancement of nations globally, the prediction of student performance is becoming increasingly significant. The educational process is entirely dependent on producing generations that are capable of leading this country and its steps toward progress in every area (scientific, economic, social and military, etc.). Consequently, one of the key criteria that motivate governments to ensure that academic institutions represent vast and scrupulous efforts to move the academic process towards continuous and improving advancement is advancing the academic process. Prediction can help in getting future knowledge. The more the volume of data is, like in massive databases, the more the forecasting is generated; this process is referred to as data mining which helps find the concealed information by examining various data sources associated with diverse domains, including social enterprises, healthcare, and academics (Chen et al., 2020; Miguéis et al., 2018). Relevant information is extracted for analyses of academic sources using EDM (Educational Data Mining), a new discipline for discovering important information using technology (Bakhshinategh et al., 2018).

The efficacy of learning environments improves as a result of statistical analysis and deep learning analysis. There has been a rapid increase in the significance of EDM currently due to the rise in the data gathered, based on the academic data obtained from various e-learning systems, along with the progress made in conventional academic systems. It is a kind of education in which the emphasis is on the instructor rather than the students, and it has drawbacks such as large class sizes, a lack of individual attention, and the use of traditional teaching tools such as black and whiteboards, desks, and pens. EDM's strengths rise from linking data from different domains. It is involved with feature extraction to help in the development of the academic process from the tremendous amount of data that the institute provides. Educational data mining refers to the processes of analyzing, investigating, forecasting, clustering, and classifying data found in educational institutions. (Chen et al., 2018).

In contrast to the traditional database search, which may answer questions like," That is the ward who failed the exam?" EDM can answer complex questions, such as predicting whether or not a student will pass an exam. Educational organizations attempt to develop a model of their student for the prediction of both the features and performance of every student separately (El Aissaoui et al., 2019). Hence, the scholars working with the EDM domain make use of diverse approaches of data mining for evaluating the lecturers, to lead their educational institutes. As due significance is not given to forecasting the performance of students in the present academic systems, these systems are bogged down due to a deficit of efficiency. The procedure of estimating the lessons, which the student might find interesting and having knowledge of his activity in academic organizations helps increase the educational efficacy. Many academic institutions are using MLTs (machine learning techniques) and EDM to assess their student's performances. These assessment systems are quite practical in enhancing student performance and also the entire academic process (Cuevas et al., 2018).

Currently, a distributed database is of immense help by considering the strong features used in the variety of its applications. Data is considered the key feature of any academic institution for having safe and right control of organizational data. In the last few years, educational data mining (EDM) has gained much focus among scholars in improving the quality of higher education.

This paper presents a study on the factors that influence the academic performance of higher education students and develops a classification model that utilizes both single and ensemble-based classifiers to predict student performance. The ensemble model combines different techniques to

achieve more accurate results and is particularly suited for creating a robust and reliable predictive model. By using this approach, it is possible to identify at-risk students early on and propose measures to improve their academic performance. Previous research in this area has focused on using classification for predicting outcomes based on registration data, course performance, and grading systems. However, there have been few studies that use ensemble classification schemes to predict students' final results based on their scores. In this study, a Multi-Agent based Distributed Data Mining approach is employed to predict student performance and optimization rules are used to assist students who score poorly. Ultimately, the evaluation of all courses in the study plan will determine which courses have the most significant impact on student performance.

A distributed database is specified as a database deployed at various machines at the same or multiple places, but to the end user, it seems to be one centralized database. This distribution of databases allows the handling of loads without loading one machine. These distributed databases are synchronized to work together, thus allowing the execution of multiple processes simultaneously, ensuring faster delivery of data/results. The systems, including workstations, microcomputers, desktops, and servers, are connected using wireless or network technologies for their inter-communications.

Highlight Overviews

The major components of the outcomes of this research can be summarized as follows:

1. We explored and modelled the essential factors affecting students' performance at the higher education level. Next, we created a performance check paradigm for developing an effective classification model by utilizing the combination of single and ensemble learning.
2. For supporting students' performance prediction, we described the role and importance of a multi-agent system which governs the mining approach in a distributed fashion.
3. The detailed literature explores and highlights the possible research fields to improve EDSS by considering the optimized psychological pave of students' performance.
4. Finally, apart from finding the research gaps, we also highlighted comparative machine learning techniques with their merits and demerits in developing EDSS.

This paper is structured as follows: Section II provides a review of the relevant literature on predicting students' performance in higher education institutions and a comparison with existing approaches. Multi-agent system models and their application are introduced in section III. Section IV presents a discussion section that summarizes the research findings and compares them to the existing literature. Section VI proposes potential solutions. Finally, the paper concludes in Section VII with a summary of the main findings and their implications.

LITERATURE SURVEY

The student success estimation using data mining methods for the advancement of the wide range of research available is shown below.

Natek & Zwilling, (2014) analyzed DMTs with less sized datasets on students by comparing two diverse DMTs. Their conclusions were assertive and showed that combining the tools based on DMTs has an essential role in higher education information management systems. Data collected between the years 2010–2011 (42 students), 2011–2012 (32 students), and 2012–2013 (42 students) (32 students) included several attributes of students, including historical school records, family histories, and demographics. Their academic achievements were determined using three classifiers: Rep Tree, J48, and M5Pwhere experimental trials demonstrated that J48 had much less accuracy than Rep Tree despite higher sensitivity values. Hamoud et al., (2018) utilized Weka for evaluating university student performances and factors that impacted their successes or failures. The study used Google

Forms for their questionnaire with 161 questions, where the open-source tool "Lime Survey" was used on University of Basrah's College of Computer Science and Information Technology students. The study used J48, Random Tree, and Rep Tree for categorizations where J48's categorizations were better in terms of accuracy when compared to the other two approaches. The study's DTs (Decision Trees) results were remarkable and accurate. Though MLTs could be used in several other domains to predict accurate outcomes, this study's experimentations indicated student statuses as 'passed or failed' but did not exhibit valid scores.

Yukselturk et al., (2014) highlighted the identification of dropout students applying DMTs online. Four DM mechanisms, referred, to as KNNs (K-Nearest Neighbors), DTs, NB (Naive Bayes) and NNs (Neural Networks) were used. The performance of KNNs was the best among all the classifiers, yielding an accuracy of 87. But, the model just considered the four algorithms for the prediction of the dropouts and not the real scores got by the students. The past studies evaluated the performance achieved with five well-known MLTs, which help in the classification of students who are at risk prior and hardships faced by them in HEIs were predicted (Hussain et al., 2019; Marbouti et al., 2016). Using ANNs (Artificial Neural Networks), SVMs (Support Vector Machines), LRs (Logistic Regressions), NB (Naïve Bayes), and DTs. In the study's experimental results, ANNs and SVMs achieved higher accuracies (57) based only on demographic data (Hussain et al., 2019), while NBs showed sufficient accuracy, despite being a weak framework (Marbouti et al., 2016). Kavyashree & Laksmi, (2016) have mentioned that a country's progress is closely linked to the quality of its academic system. The educational sector has seen a tremendous shift concerning its operation. Recently, it has been identified as an industry, and several problems are faced here in this process. The declining student success rate and the non-completion of courses are the main concerns of higher education. A timely prediction of the failure of the students may aid the management in rendering counselling and improving their retention and success in passing 5 using special coaching pupil retentivity. Data mining has found extensive application in the academic field to discover novel hidden patterns using the student's information that helps understand the issue. Classification is performed by one of the prediction sorts of classifiers, which categorizes the data based on the training set and utilizes the pattern for the classification of fresh data. The basic objective of the endeavour is to come up with a design for an internetworking application, which makes use of a data mining approach during the students' performance prediction under their activities. This article investigated academic performance using connections between emotions and students' economic status with NBs Classifications.

Hastie et al. (2009) introduces a novel algorithm that expands upon the original AdaBoost algorithm for multi-class classification, eliminating the need to transform it into multiple two-class problems. The proposed method combines weak classifiers and sets a requirement for each weak classifier to outperform random guessing. FSAMME is an innovative AdaBoost algorithm designed for multi-class classification. It incorporates a novel multi-class exponential loss function and employs forward stage-wise additive modeling. Unlike traditional approaches, FSAMME directly tackles the multi-class problem without transforming it into multiple two-class problems. It ensures that each weak classifier performs better than random guessing. Notably, FSAMME achieves higher classification accuracy compared to AdaBoost (Hu et al., 2008).

Dwivedi & Singh, (2016) showed that Students' background examinations are beneficial for academic planners in institutions in directing them in the proper direction. In case the class of students is predicted during the mid-year of the institution during the final year. The academic planner can easily plan for a few essential workshops to improve student performance, which, in turn, can help in their placement at the academic year-end.

Fedushko et al. (2022) proposes a solution for improving academic specialty selection among Ukrainian students in higher educational institutions. By utilizing a decision-making architecture and a modeled database, the system provides detailed descriptions of university specialties. The study highlights that many students choose popular but potentially less in-demand specialties instead of considering tuition-free education. To address this issue, incorporating intelligent algorithms,

user behavior analytics, and consultations with academic and career orientation experts to enhance decision-making in specialty selection.

Shahiri & Husain, (2015) have reported that the Prediction of the students' performance has become a vast challenge owing to the enormous amount of data in educational databases. Recently in Malaysia, attention has not been paid to the deficit of available systems to assess and monitor the improvement and performance of students. There are two major reasons behind this. At first, analyzing the available prediction techniques is still inadequate in identifying the most desirable techniques designed for student performance predictions in Malaysian academic organizations. As a result, holistic literature assessments on student performances were predicted using DMTs.

Hasan et al., (2018) investigated student academic performance with the help of a decision tree algorithm in terms of parameters such as the Academic Information and Activity of Students. The records of 22 students are collected from the Spring 2017 semester, registered for their undergraduate degree from Oman's private Higher Education Institution. The proposed research work uses the Electronic Commerce Technologies module as it is a primary module that has been proposed in all the computing specializations. Also, the WEKA data mining tool is utilized for assessing the decision tree algorithm to measure the student's performance based on Moodle access time (Figueira, 2016). On the academic data set of secondary schools acquired from the Ministry of Education in the Gaza Strip for the year 2015, Amra & Maghari, (2017) presented a framework for student performance prediction using KNNs and NBs classifiers. Because of the timely forecasting of student performance, the primary purpose of this categorization may be valuable to the educational sector for performance improvement. Educators may also accurately assess to improve student learning. Experiments showed that NBs outperformed KNNs, with a maximum accuracy of 93.6. Fernandes et al., (2019) evaluated educational achievement predictions for public school students in the Federal District of Brazil for the years 2015 and 2016. Statistical analyses produced clear perspectives based on facts. Two datasets were then extracted where the first dataset encompassed the school's academic year starting information, whereas the other dataset contained semester information of two months since the beginning. The classifications were based on GBMs (Gradient Boosting Machines), which estimated students' academic grades for the latter portions of the study year. Even though the features' grades' and 'absences' had the highest relevance for predictions of years-end educational results of student performances, the evaluation of demographic characteristics showed that neighbourhoods, schools, and ages were strong indicators of student results.

Francis & Babu, (2019) Presented a novel prediction algorithm to evaluate the student's academic performance depending on classification and clustering approaches and has been validated in real-time using a student dataset of different academic fields present in higher educational institutions in Kerala, India. It is proven from the result that the ensemble algorithm, which combines clustering and classification approaches, provides results that are better in terms of yielding accuracy in predicting the student's educational performance. While the proposed model has yielded promising results, it may be further expanded in the future to support a wide variety of student datasets.

Al-Shehri, et al., (2017) compared supervised learning classifiers, such as SVMs and KNNs, on the University of Minho's data with thirty-three characteristics which were converted into numeric forms from their nominal forms. The data gathered through questionnaires and reports from two Portuguese schools encompassed nominal (4), binary (13), and numeric (16) for a total of 395 occurrences. Experimentations included several data set partitions using Weka, and the study discovered that SVMs yielded the best accuracy while using 10-Fold cross-validations and partitioning.

Sverdlyka et al. (2022) investigates the integration of video content within student learning, focusing on its dual role of entertainment and education. It particularly emphasizes the utilization of YouTube as a platform for educational activities, addressing organizational and technical aspects, as well as its connection with social networks and related services. This integration of video content into e-learning is regarded as a noteworthy achievement in the field, signifying a significant advancement in enhancing the learning journey for students.

Daud et al., (2017) looked at how contemporary learning analytics might be used to predict student success with data on students studying in Pakistan on scholarships. The discriminative models of CART, SVMs, and C4.5, as well as generative models of Bayes Networks and NBs were investigated in this study for their comparatives values of precisions, recalls, and F-scores for predictions. The data of 3,000 students was collected between 2004 to 2011, however after pre-processing and removing duplication, the number of students got reduced to 776. 690 of the 776 students had completed their degrees successfully, while 86 did not. Thirty-three variables, grouped into four categories, contained data on Family expenditures, incomes, and students' personal information. Their research revealed that expenditures on natural gas, electricity, self-employment, and locations were significant factors in predicting student academic achievement. SVMs with F1 scores of 0.867 outperformed other compared approaches.

Hafizah et al., (2020) conducted a study in which they briefly introduced different fragmentation methods for distributed database systems. All fragmentation methods described in this research can improve the efficiency of distributed database systems, reducing transmission/communication costs and access/response times. Based on the researchers' work, it can be concluded that fragmentation strategies are an essential technique that can be used to improve distributed database systems and data health. It is vital to maintain an appropriate data exchange structure for the full utilization of resources, so choosing a reliable and efficient data exchange structure is essential to improve the efficiency of the distributed database system (Ubaidillah and Ahmad, 2020) . However, research on the application of data fragmentation technology in distributed database systems is still limited. To increase application efficiency, more data fragmentation methods should be investigated and implemented in distributed database systems for further research.

Migueis et al., (2018) proposed a two-stage framework using DMTs and worked on students' first-year career completion data for predicting their educational achievements. Academic performance is portrayed by both the average rank reached and the time spent to complete the program, in contrast to other literature studies on academic data mining. Furthermore, this segmented study students based their performances which included failures and higher performances before their degrees started to predict their yearend outcomes (Yağcı, 2022). The proposed framework was evaluated on 12 years of student data belonging to the European Engineering School of a public research university. Their empirical results showed that their suggested model could accurately forecast the level of students' performance in the early stages of their educational journey, with an accuracy of more than 95. RFs (Random forests) have been found to perform significantly better than other classification methods like DTs, SVMs, NBs, and bagged or boosted trees. Along with the prediction framework, the proposed segmentation model was found to be an effective tool for determining the best mechanisms to achieve higher performance levels and reduce academic failures, resulting in an overall improvement in the quality of educational experience in HEIs.

The test was conducted to anticipate student enrolment in Kenya's HEIs, including Engineering/ Mathematics disciplines and Science and Technologies (Wanjau, 2016). Nearly 18 traits were discovered based on a questionnaire. Their classifications in terms of Chi-Square and IG feature selection algorithms, CART DTs, showed enhanced prediction accuracies.

Figueri, (2016) PCAs (Principal Component Analyses) were applied to a dataset of students registering for bachelor's degrees in computer science to predict student rankings. The work used PCA to generate DTs based on information extracted from Moodle Logs to predict student results.

Jedidi et al., (2022) This paper focuses on developing a prediction model of student performance based on cloud computing to improve e-learning in the educational environment by incorporating new information technologies and communication. They demonstrate the overall architecture for predicting students' performance using google cloud service, which can be divided into three layers: infrastructure layer, cloud service layer, and user layer. The limitation of this work is the proposed approach that could be implemented in the future would be used to support students with low grades. The data is small, and it can be escalated.

Gao et al., (2022) proposed a deep cognitive diagnostic framework to predict students' performance. The experimental results show that the proposed framework is better than the competing strategies of cognitive modelling. The limitation of work is that the skills required by problems are labelled. The limitation of the work is that they used a limited data set and can be extended to multiple datasets for better results. Maraza-Quispe et al., (2022) the researchers aim to predict the model using a simple regression tree algorithm to predict academic performance, particularly to identify at-risk students at the beginning of the course. The data set was obtained from an LMS, and they utilized data from one semester and three courses, although a larger data set may be used for greater accuracy.

Marbouti et al., (2016) proposed a model that is helpful for both instructor and student prediction by using the Support Vector Machine (SVM) technique to anticipate the student's early risk based on semester data. Students' retention rates can be increased and used to modify the course. The SVM is valuable to some extent, but still, it needs some improvement. Also, it may not apply to general situations. The early prediction may predict accuracy and vice versa for late prediction, which uses a limited dataset. Bravo-Agapito et al., (2021) recommended a group of models to overcome the Complete Online (CO) university that can predict student academic performance after studies. Research results indicate that online education is mainly related to intervention approaches. The results of the present studies are not generalizable due to the limited data set and features.

Injadat et al., (2020) valuable results have been obtained from the suggested proposed approach; this work suffers from some limitations that may have affected two results: a limited data set, limited features, and unbalanced data. Furthermore, they used some methods to expand the complexity of the models, resulting in model overfitting.

Asif et al., (2017) present a study for predicting undergraduate students based on the reported pre-university and first-year scores. The work has good accuracy for the 2 | P a g e small dataset but limitations for the generalized case Liu & Niu, (2021) propose a new approach to predict student learning performance using the Agent-Based Modeling Feature Selection (ABMFS) model. ABMFs select the targeted features and then use the selected features as input to a Convolutional Neural Network (CNN)-based. CNN model is the Deep learning technique for training to obtain prediction results. Kumar et al., (2017) also presented a vital baseline for developing machine learning-based models.

Liu et al., (2021) used standard evaluation metrics to evaluate the experiments' proposed ABMFS and CNN-based models. They compared the prediction performance of the current mainstream classifiers using ABMFS model selected features with using all features. The obtained results signify that using the ABMFS model selected features improves the prediction results on Portuguese and Mathematic data. However, a large data set and more algorithms can enhance the prediction accuracy in students' performance, and better results can be achieved.

Al-Obeidat et al., (2017) The data analysis techniques are presented in real case studies to predict students' performance using their past academic experience. They suggested a new hybrid classification technique that utilizes a decision tree and fuzzy multi-criteria classification. This approach uses several criteria such as age, school, address, family size, evaluation in previous grades, and activities to indicate students' performance. The current work is compared with other prominent classifiers to check the model's accuracy, and the acquired result showed that this is a promising classification tool (Al-Obeidat et al., 2018; Pérez and Kholod, 2020).

Noraziah et al., (2021) the proposed algorithm "Binary Vote Assignment on Grid Quorum with Association Rule (BVAGQAR)" to classify the data and manage fragmented database synchronous replication. Classify and fragment techniques improved the performance of the distributed database system, increased data accessibility, and reduced the transfer cost and access time. The BVAGQ-AR algorithm can split the database into separate disjoint fragments. The result of the experiment indicates that handling fragmented database synchronous replication by the proposed BVAGQ-AR algorithm is capable of maintaining data consistency in a distributed en-environment. The limitation of the work is that BVAGQ-AR does not support handling fragmented database replication transaction management

by considering failure cases, and BVAGQAR has to deal with fragmented database failure cases and fault tolerance (Noraziah et al., 2021)

Kiu & Ching-Chieh, (2018) in this paper, researchers used four supervised educational data mining techniques, Naïve Bayesian, Multilayer Perceptron, Random Forest, Decision Tree J48, and Naïve Bayesian, to predict students' especially mathematic performance in secondary school. This study indicated student social activities and background and significantly predicted student performance from the obtained results. using these models, early prediction of student performance in a particular subject can be achieved. Therefore, these models are helpful for teachers and students for early prediction. The limitation of the work is that it can't apply to unsupervised education data mining techniques (Kiu, (2018).

Yossy & Heryadi, (2019) This study aims to determine the accuracy of the most accurate classification algorithm to measure student performance predictions. This study uses seven methods: random forest, classification and regression trees, AdaBoost, K-nearest neighbour, naïve Bayes, extra tree, and bernaoulli naïve Bayes. Student math data is used for prediction. The technology used to compare the seven methods uses Python programming. Cross-validation is used in the testing of the performance of different approaches. Based on these results we know that the best classification method is the random forest. The result of the study shows the best classification algorithm for student performance with the student math data is the random forest. In addition, the three algorithms which have the best performance are random forest, AdaBoost dan KNearest Neighboring. In the future, this research can improve to research more than seven algorithms and can research the most influential features of student performance (Yossy & Heryadi, 2019)

Education Data Mining

Algarni, (2016) demonstrated that DMTs help extract valuable information from unprocessed data and can impact decisions significantly. EDM was used to extract useful information as the usage of technology in educational systems has resulted in massive volumes of student data requiring storage, rendering it significant to make use of EDM to improve the processes of teaching and learning. EDM is useful in various domains, such as identifying at-risk students and recognizing the learning required to be given priority for multiple categories of students. It is also helpful to improve graduation rates, efficiently evaluate the institute's performance, improve campus resources, and optimize the subject curriculum redevelopment. This paper analyses the various projects in the EDM area and includes the data and methodologies used in those projects.

Czibula et al., (2019) have learned from the obtained results; that the used classifier is better than supervised classifiers already applied in EDM literature for measuring Students' performance prediction and can only be applied to classification problems. The Student Performance prediction using Relational Association Rules (SPRAR) may not solve the regression problems. is study's objective was to assess the efficacy of the K-Mean and X-Mean clustering algorithms by applying them to two distinct datasets consisting of students enrolled in higher education. These datasets were taken from the Kaggle repository and were used in the study. According to the results, X-Mean is more suitable for usage with large datasets in terms of detecting clusters and the accuracy of such discoveries. This is the case because X-Mean is ideal for large datasets. It was also discovered when comparing the two algorithms that the K-Mean technique performed well on the short dataset compared to the X-Mean algorithm, showing that the X-Mean methodology performs better over the enormous datasets. This was discovered while comparing the two methods (Kesorn & Poslad, 2011).

Costa et al., (2017) Studied the EDM approaches to measure its efficiency in the performance prediction of students. The research was contributed, and it is different from other relevant works. By utilizing this efficient EDM approach, the students who fail during the early times of courses are identified. Later, it will help decide to limit the failure rate. To increase the efficiency of these EDM approaches, the research studies the influence of pre-processing the data, and the algorithms are fine-

tuned. Their research showed that EDM approaches successfully predicted students' academic failures on time, making it easier for instructors or academics to make performance improvement decisions.

The existing educational systems face various challenges, leading to the need for Educational Decision Support Systems (EDSS). These challenges include the lack of individualized learning, difficulties in managing and analyzing educational data, high teacher workloads, inefficient resource allocation, the need for early identification of at-risk students, limited professional development opportunities, and the necessity for improved parent and stakeholder engagement. EDSS can address these challenges by providing personalized learning, automating data management and analysis, reducing teacher workload, optimizing resource allocation, identifying at-risk students, facilitating professional development, and enhancing parent and stakeholder communication.

Student Success Estimation Using Machine Learning Methods

Acharya & Sinha, (2014) also used MLTs in their study to predict student performances where gender, revenues, board-level marks, and attendance were input parameters for analyses. The study used C4.5, SMOs, NBs, 1-Nearest Neighbourhoods, and MLPs (Multi-layer Perceptrons) for classifications where SMOs were found to be very effective for improving model performances concerning students studying courses, yielding maximum mean test accuracy of 66% when compared to the rest of the techniques. Lakkaraju et al., (2015) provide a brief overview of an elaborate model, which employs MLTs for identifying students with risks of not succeeding in their high school graduations at the stipulated time.

Table 1. Comparison of existing approaches for students' performance prediction in higher education institutions

Author Name	Methods	Merits	Demerits
Francis & Babu, (2019)	Hybrid algorithm combining clustering and classification	Better in terms of attaining accuracy in prediction	Performance is poor for big dataset
Francis & Babu, (2019)	Gradient Boosting Machine	Effective and scalable	High sensitivity to outliers
Fernandes, et. al., (2019)	Random forests	Yields improved results	Training speed is low
Alaa Khalaf Hamoud, et. al., (2018)	J48, Random Tree and Rep Tree	Yields improved accuracy	Time consumption is high
Al-Shehri, et. al., (2017)	Support Vector Machine algorithm and K-Nearest Neighbour	Improves the truly positive results	Generates errors with massive volumes of data
Amra, & Maghari, (2017)	KNN and Naïve Bayes	Yields the maximum accuracy value of 93.6%	The speed of the Prediction stage is quite low
Xu et .al., (2017)	Novel machine learning	Yields improved performance compared to standard techniques	in economic
Natek & Zwilling, (2014)	Rep Tree, J48 and M5P models	Yields improved performance	Is not validated for student data sets obtained from multiple nations
Erman Yukselturk et. al., (2014)	KNN, DT, NB and Neural Network	Yields reasonable accuracy	Has not been tested for other applications
Acharya & Sinha, (2014)	C4.5, SMO, NB, 1-Nearest Neighbourhood and Multi-layer Perceptron	Achieves 66% accuracy	the error rate is increased

In this article, Sixhaxa et al., (2022) propose a model that uses different features like academic, and behavioural characteristics, and demographic and examines how these characteristics affect student performance and help identify at-risk students. The limitation of this study is the sample size or small data set, but a large dataset can be used to produce accurate predictions. Researchers Yagcı & Mustafa, (2022) propose a new model based on machine learning algorithms to predict the final exam grades of undergraduate students, taking their midterm exam grades as the source data. Machine learning algorithms RF, NN, SVM, LR, NB, and kNN algorithms are used to predict students' performance in final exam grades.

Dabhade et al., (2021) researchers applied a few machine learning algorithms to predict the student's performance in final year undergraduate in an Institute, and it may be used to enhance institution ranking. The support vector regression linear algorithm has obtained good results compared to other models. Tomasevic et al., (2020) in this research work, three Supervised Machine learning techniques indicate a high risk of dropping out from courses and for student exam performance prediction.

Vukovic et al., (2021) Recently, the multi-agent system ´ has been reported to identify the various parameters like students' engagement with the assessment activity using the two different machine learning approaches. However, the work cannot answer how much engaging activities are associated with the student's performance.

Xu et. al., (2017) conceived a novel approach based on MLTs for predicting degree student performances. The proposed technique offers two essential features. The recommended solution uses dual layers with numerous base predictors and sequential ensemble predictors. Their data-driven approach was based on the model's latent elements and probabilistic factorised matrices for determining the relevance of courses required for accurate base predictions. The study's extensive simulations on undergraduate student datasets with three years of information from the University of California, Los Angeles (UCLA) revealed that their proposed technique yielded improved performances compared to standard methods.

The novel aspect of this research is that it combines machine learning and mathematical modelling techniques with Operational intelligence and SRE to create a highly accurate model for transaction tracing in a real-time distributed environment, with the end goal of improving technology and, in particular, the success rates of data-based projects, as well as gaining additional insights into a specific tier/layer of a data/IT operation process. Due to a lack of operational insights, cutting-edge infrastructure, monitoring of data pipelines, and mathematical modelling, technological ventures frequently fail. It is necessary to evaluate prediction models in real time and to assess accuracy using thresholds. Operational intelligence, machine learning, and big data analytics are therefore becoming increasingly popular in this era of cloud and quantum computing and have the potential to enable the implementation of high-quality data projects (Najafabadi, et al., 2015).

This paper describes a machine-learning strategy for analyzing academic performance and related data. This study can be used to categorize and regress data sets, identify and recognize current educational difficulties, and assess student samples' individual and group characteristics. We describe the data analysis, algorithm employed, and results. Determine the dataset's most essential traits, facts, and conclusions. The approach delivers accurate assessments of academic achievement. If students and children are divided into courses based on their projected qualities, they will be better prepared to communicate, lead, and manage themselves. According to the findings, performance metrics assessment appears to play an essential role in developing contemporary education and improving individual students' knowledge (Fedushko, et al., 2019).

MULTI-AGENT SYSTEM

Intelligent Agents are software and hardware organizations carrying out specific activities for users with a certain degree of independence. To aid someone, an agent has to contain a certain degree of

Table 2. Comparison of existing approaches for students' performance prediction in higher education institutions

Author Name	Methods	Merits	Demerits
Sixhaxa et al., (2022)	LR, SVM, RF, KNN, GNV	Results improved performance	Misclassified students, Not enough large data set
yajc1 & Mustafa, (2022)	LR, SVM, nearest neighbour, RF, KNN	Results improved performance	limited input parameter used
Dabhade et al., (2021)	support vector regression linear algorithm	Results enhanced performance	Not effective on a large data set
Vukovic et. al., (2021)	two different machine learning methods	Increase success rate	possibility of supporting other learning styles
Tomasevic et al., (2020)	DT, SVM, K-NN, NB, RLR, ANN	Improved performance	Not provided sufficient amount of data, more techniques can be used for good results
Xu et al., (2017)	Machine Learning Techniques	achieves superior performance	Elective courses are excluded
Lakkaraju et al., (2015)	machine learning approaches	Predict not finishing high school on time	Small data set
Acharya & Sinha, (2014)	C4.5, SMOs, NBs, 1-Nearest Neighbourhoods, and MLPs	Yields improved performance	limited collection of data

intelligence capable of choosing between several policy options, planning, communicating, adapting to environmental change, and learning from experience. An intelligent agent may generally be defined as an electrical signal, a recognizer or classifier that identifies which event took place, a set of logic ranging from a hard-coded programmer to rule-based inferences, and a method to take action. Multi-agent systems have become effective for modelling and solving problems in complex and dynamic environments. Agents and multi-agent systems (MAS) have opened up new modes for analyzing, designing, simulating, and implementing complex software systems for over 20 years. In a setting characterized by a single agent, there is only one agent, and that agent bears complete responsibility for the activities it takes. However, the environment consists of numerous active agents, making it a multi-agent environment (Falco & Roiolo, 2019; González-Briones et al., 2018).

Movement and learning are also critical features of the agent paradigm. It is mobile if an agent can travel a network and execute remote machine tasks. In response to the environment, a supervised learning algorithm will adapt to the needs of its user and will automatically modify its behaviour. An event-condition-action paradigm may be defined for learning or intelligent agents. An event is defined as something that changes the surroundings or something that the actor should be conscious of in the framework of intelligent agents. For instance, the delivery of a new letter or a modification to a web page may represent an event. In case of an occurrence, the agent must recognize and evaluate the significance of the event and react to it. This second stage might be simple or highly complex depending on the circumstances and determine what the situation or status of the planet is. When mail comes, the event automatically describes itself, then the agent may query the mail program to identify who sent the message and what the topic is, or even scan the mail content to identify keywords. All these elements are part of the cycle identification. The agent can wake up to the first occurrence, but the agency has to understand how important the event is for its tasks. The communication can be categorized as urgent if the mail comes from the employer of the user. This makes clever actors the most valuable component. The idea of autonomy is the principal problem in the usage of intelligent agents. This" intelligent" program can be done by the user for various time-consuming computer procedures. This allows users to go on to other activities and even unplug from their computers during the operation of the software agent. Moreover, the functioning of the computer is not to be learned by

the user. Indeed, clever agents are capable of acting as a software layer to give usability to computer experts that many novice people seek (Perez & Kholod, 2020; Vuković et al., 2021).

A multi-agent system (MAS) consists of multiple intelligent collaborators working together to solve a problem beyond a single agent's capacity. MAS systems have attracted attention for their use in many real-world applications that individual agents cannot handle. Multi-agent systems (MAS) have received much attention from scientists of all disciplines, including computer science and civil engineering, to solve complex problems by breaking them down into smaller tasks. MAS has found many applications, including modelling complex systems, smart grids, and computer networks. Despite its broad applicability, MAS remains several challenges, including inter-agent coordination, security, and task assignments (Dorri, et al., 2018).

In recent years, MAS has attracted the attention of researchers due to its potential applications in various fields such as biology, physics, systems engineering, control systems, smart grids, etc. A large amount of literature can be found on the use of the MAS in different areas. MAS is suitable and can work effectively for achieving individual goals instead of the general goal. MAS systems have received attention for their use in many practical applications. Agents can act flexibly and independently, making informed decisions based on their intelligence and experience. A multi-agent system (MAS) consists of multiple intelligent collaborators working together to solve problems that a single agent cannot solve (Mahela, et al., 2018).

MAS is a widely used method in many fields and is widely used due to its ability to communicate, coordinate and cooperate between agents and its ability to assign agents for different tasks. MAS systems can be used for a variety of purposes and provide dynamic solutions to complex problems that can be solved by MAS systems. A Multi-Agent System (MAS) is designed to achieve multiple goals based on a set of rules and regulations. MAS is a system that integrates a set of tools that communicate, interact and coordinate with each other to achieve a specific goal. New features and capabilities allow MAS to be used in a variety of fields and environments (Gonzalez-Briones, et al., 2018).

Raza et al., (2019) have reported statistical-based QE methods such as browse log analyses, web knowledge analyses, and search and document analyses. They have also discussed the merits and demerits of each technique. The current research will help understand some of the essential SQE strategies and select the best approach considering research, group inquiries, and computational proficiency requirements. The results indicate that choosing the optimal approach depends on the type of nature and availability of data sources, search query, and performance efficiency requirements. In addition, a hybrid combination of these techniques should be considered to improve the average rate of retrieval performance in the future. This research aimed to investigate the efficiency of the K-Mean and X-Mean clustering methods by using two datasets of student enrollment in higher education that are collected from the Kaggle repository.

DISCUSSION

The primary objective of this study is to investigate the significant factors that have a bearing on the academic performance of students who are enrolled in higher education and to design an efficient classification model to forecast educational performance by utilizing a combination of single and ensemble-based classifiers. The vast majority of the older research-related works emphasize the use of categorization for predicting based on registration data, students' performance in a specific course, grade inflation, the projected percentage of failing students, and assistance in the grading system. According to the most up-to-date information available, only a few state-of-the-art methods have been discovered that use an ensemble classification scheme to forecast the outcome for students based on their scores.

Multi-agent systems and comparative machine learning approaches are essential components of Educational Decision Support Systems (EDSS), whereas they contribute by enabling collaboration, personalization, scalability, performance evaluation, decision-making support, and adaptive learning

within EDSS. Multi-agent systems facilitate communication, personalization, and scalability, while comparative machine learning allows for performance evaluation, decision support, and adaptive learning. The advantages and applications of these approaches in EDSS include individualized learning, intelligent tutoring systems, collaborative learning, decision support for teachers, and informing educational policy design. The technologies used in the approaches enable collaborative decision-making, personalized learning paths, intelligent course recommendations, early intervention, and continuous improvement. By integrating multiple agents with specialized knowledge, systems can provide comprehensive recommendations and adapt learning paths to individual needs of students. While, comparative machine learning helps to identify patterns and correlations in student data, while multi-agent systems facilitate timely interventions and support. Additionally, these approaches enable ongoing assessment and refinement of educational strategies. Overall, these technologies have the potential to transform educational decision-making and support, empowering students and enhancing their academic success.

RECOMMENDED SOLUTIONS

EDM is a developing research field, which is currently exploited for examining the data for various educational objectives. EDM is primarily applied in predicting the academic performance of students. In data mining, the evaluation and elucidation of the academic performance of students are considered to be the apt analysis, evaluation, and assessment means. In the current times of a knowledge economy, students constitute the primary faces for the socio-economic development of any nation, therefore it is important to keep their performance on track. Data mining (DM) techniques are used for learning the hidden knowledge and patterns that help administrators and academic scholars in making decisions about how the instructions are delivered. DM approaches have been applied in several fields, which include retail business, the medical sector, marketing, banking, bioinformatics, counterterrorism, and several others that are also utilizing it to improve the throughput and efficacy. Educational data mining is a rapidly growing field employed in educational contexts to enhance students' understanding and learning process. Its primary objective revolves around the identification, extraction, and analysis of data associated with educational procedures and student performance. Educational data mining encompasses the exploration and application of cutting-edge techniques to unearth valuable insights from educational domains. Although, Distributed Data Mining (DMT) provides several benefits for predicting students' performance. Its scalability allows for processing large educational datasets, improving the accuracy of prediction models. DMT's parallel processing capabilities enable faster analysis and model building. It also ensures privacy preservation by keeping data distributed. DMT optimizes resource utilization and adapts to changing educational environments. Collaboration and knowledge sharing among stakeholders enhance prediction outcomes. In summary, DMT provides scalability, accuracy, speed, privacy, efficiency, adaptability, and collaboration for effective performance prediction in education.

Maximum prediction accuracy of the student's performance is of immense help in identifying the students who are poorly performing during the start of the learning process. Data mining is useful in achieving this goal. DMTs help find the structures or patterns of data, and it aids extremely during the decision-making process. Future work in this field highlights the concept of making use of (stage-wise additive multi-Modeling using Multiclass Exponential loss function) SAMME boosting approach improves AdaBoost to a multiclass classification with no need for it to be reduced to a bunch of sub-binary classification. In addition, a Performance prediction system can be well developed by employing distributed database mechanism combined with a Multi-Agent model for predicting the students' performance under their data yielding improved prediction accuracy and being of assistance to the poor-performing students using optimization rules.

CONCLUSION

In higher academic organizations, the performance of students plays a crucial role as it directly impacts the reputation and success of academic institutions. To address this, researchers have turned their attention to data mining technology and a distributed data mining framework to effectively estimate student performance. The purpose of this research is to explore how these techniques can be leveraged for accurate prediction. This research work provides an overview of the benefits and limitations associated with predicting student performance. EDSS enhances college effectiveness and mitigates failure risks by leveraging multi-agent systems and comparative machine learning. Distributed Data Mining Techniques (DMTs) helps decision-making by identifying models and data. The Intelligent Knowledge Base Distributed Data Mining framework employs multi-agent system-based educational mining to assess student performance in mid-term and final exams, while investigating academic achievement. The development of EDSS heavily relies on multi-agent systems and comparative machine learning approaches. By analyzing the available data mining technology, the SAMME boosting approach is utilized. Unlike traditional AdaBoost, this method enhances multiclass classification without the need to convert it into sub-binary classifications. This advancement allows for more efficient and effective estimation of student performance. Furthermore, the research work suggests the development of a Performance Prediction System that incorporates Multi-Agent Data Mining. This system utilizes student data to improve the accuracy of performance prediction. Its primary objective is to assist underperforming students by providing personalized support based on their individual data. By leveraging the power of Multi-Agent Data Mining, the system can extract valuable insights and patterns from the data to enhance prediction accuracy.

LIMITATIONS AND FUTURE DIRECTIONS

Prediction of Students' performance accurately involves a complex process that requires more intelligent approaches to consider the evolving facts and circumstances. These facts and circumstances vary for different student communities based on personal attributes. There are research gaps that exist in the present data mining models, which are:

1. There are only a few hybrid methods that combine the benefits of both supervised and unsupervised learning for automating the prediction and enhancing prediction accuracy in students' performance.
2. The present models are inflexible for analyzing the major academic and personal features that greatly influence the students' performance.
3. Although few hybrid approaches are available, these cannot dynamically adjust their potential by predicting performance based on personal and external features.
4. Many of the existing models use a single data set, and hence there is a question of performance when applied in distributive multi-data.

Only a few studies have dealt with the challenge of integrating heterogeneous data and knowledge in a combined hybrid framework. Furthermore, the current literature addresses either the accuracy or cost functions since both are Inverse proportionality. The proposed study seeks to address communication cost and accuracy in a distributed educational data environment.

ACKNOWLEDGMENT

REFERENCES

Acharya, A., & Sinha, D. (2014). Early prediction of students' performance using machine learning techniques. *International Journal of Computers and Applications*, *107*(1).

Al-Obeidat, F., Tubaishat, A., Dillon, A., & Shah, B. (2018). Analyzing students' performance using multi-criteria classification. *Cluster Computing*, *21*(1), 623–632. doi:10.1007/s10586-017-0967-4

Al-Shehri, H., Al-Qarni, A., Al-Saati, L., Batoaq, A., Badukhen, H., Alrashed, S., . . . Olatunji, S. O. (2017, April). Student performance prediction using support vector machine and k-nearest neighbour. In *2017 IEEE 30th Canadian Conference on Electrical and computer engineering (CCECE)* (pp. 1-4). IEEE.

Algarni, A. (2016). Data mining in education. *International Journal of Advanced Computer Science and Applications*, *7*(6). Advance online publication. doi:10.14569/IJACSA.2016.070659

Amra, I. A. A., & Maghari, A. Y. (2017, May). Student's performance prediction using KNN and Naïve Bayesian. In *2017 8th International Conference on Information Technology (ICIT)* (pp. 909-913). IEEE.

Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, *113*, 177–194. doi:10.1016/j.compedu.2017.05.007

Bakhshinategh, B., Zaiane, O. R., ElAtia, S., & Ipperciel, D. (2018). Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies*, *23*(1), 537–553. doi:10.1007/s10639-017-9616-z

Bravo-Agapito, J., Romero, S. J., & Pamplona, S. (2021). Early prediction of undergraduate Student's academic performance in completely online learning: A five-year study. *Computers in Human Behavior*, *115*, 106595. doi:10.1016/j.chb.2020.106595

Chen, P., Lu, Y., Zheng, V. W., Chen, X., & Yang, B. (2018). Knowedu: A system to construct a knowledge graph for education. *IEEE Access : Practical Innovations, Open Solutions*, *6*, 31553–31563. doi:10.1109/ACCESS.2018.2839607

Chen, W., Wang, X., Duan, H., Zhang, X., Dong, T., & Nie, S. (2020). Application of deep learning in cancer prognosis prediction model. *Sheng wu yi xue Gong Cheng xue za zhi= Journal of Biomedical Engineering= Shengwu Yixue Gongchengxue Zazhi, 37*(5), 918-929.

Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., & Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behaviour, 73*, 247-256.

Cuevas, R., Ntoumanis, N., Fernandez-Bustos, J. G., & Bartholomew, K. (2018). Does teacher evaluation based on student performance predict motivation, well-being, and ill-being? *Journal of School Psychology*, *68*, 154–162. doi:10.1016/j.jsp.2018.03.005 PMID:29861025

Czibula, G., Mihai, A., & Crivei, L. M. (2019). S PRAR: A novel relational association rule mining classification model applied for academic performance prediction. *Procedia Computer Science*, *159*, 20–29. doi:10.1016/j.procs.2019.09.156

Dabhade, P., Agarwal, R., Alameen, K. P., Fathima, A. T., Sridharan, R., & Gopakumar, G. (2021). Educational data mining for predicting students' academic performance using machine learning algorithms. *Materials Today: Proceedings*, *47*, 5260–5267. doi:10.1016/j.matpr.2021.05.646

Daud, A., Aljohani, N. R., Abbasi, R. A., Lytras, M. D., Abbas, F., & Alowibdi, J. S. (2017, April). Predicting student performance using advanced learning analytics. In *Proceedings of the 26th international conference on world wide web companion* (pp. 415-421). doi:10.1145/3041021.3054164

Dorri, A., Kanhere, S. S., & Jurdak, R. (2018). Multi-agent systems: A survey. *IEEE Access : Practical Innovations, Open Solutions*, *6*, 28573–28593. doi:10.1109/ACCESS.2018.2831228

Dwivedi, T., & Singh, D. (2016). Analyzing educational data through EDM process: A survey. *International Journal of Computers and Applications*, *136*(5), 13–15. doi:10.5120/ijca2016908392

El Aissaoui, O., El Alami El Madani, Y., Oughdir, L., Dakkak, A., & El Allioui, Y. (2019, July). A multiple linear regression-based approach to predict student performance. In *International Conference on Advanced Intelligent Systems for Sustainable Development* (pp. 9-23). Springer.

Falco, M., & Robiolo, G. (2019, September). A systematic literature review in multi-agent systems: Patterns and trends. In *2019 XLV Latin American Computing Conference (CLEI)* (pp. 1-10). IEEE. doi:10.1109/CLEI47609.2019.235098

Fedushko, S., & Ustyianovych, T. (2020). Predicting pupil's successfulness factors using machine learning algorithms and mathematical modelling methods. In *Advances in Computer Science for Engineering and Education II* (pp. 625–636). Springer International Publishing. doi:10.1007/978-3-030-16621-2_58

Fedushko, S., Ustyianovych, T., & Syerov, Y. (2022). Intelligent Academic Specialties Selection in Higher Education for Ukrainian Entrants: A Recommendation System. *Journal of Intelligence*, *10*(2), 32. doi:10.3390/jintelligence10020032 PMID:35736004

Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Van Erven, G. (2019). Educational data mining: Predictive analysis of the academic performance of public school students in the capital of Brazil. *Journal of Business Research*, *94*, 335–343. doi:10.1016/j.jbusres.2018.02.012

Figueira, A. (2016, July). Predicting grades by principal component analysis: A data mining approach to learning analytics. In *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)* (pp. 465-467). IEEE. doi:10.1109/ICALT.2016.103

Francis, B. K., & Babu, S. S. (2019). Predicting the academic performance of students using a hybrid data mining approach. *Journal of Medical Systems*, *43*(6), 1–15. doi:10.1007/s10916-019-1295-4 PMID:31037484

Gao, L., Zhao, Z., Li, C., Zhao, J., & Zeng, Q. (2022). Deep cognitive diagnosis model for predicting students' performance. *Future Generation Computer Systems*, *126*, 252–262. doi:10.1016/j.future.2021.08.019

González-Briones, A., De La Prieta, F., Mohamad, M. S., Omatu, S., & Corchado, J. M. (2018). Multi-agent systems applications in energy optimization problems: A state-of-the-art review. *Energies*, *11*(8), 1928. doi:10.3390/en11081928

Hamoud, A., Hashim, A. S., & Awadh, W. A. (2018). Predicting student performance in higher education institutions using decision tree analysis. *International Journal of Interactive Multimedia and Artificial Intelligence*, *5*(2), 26–31. doi:10.9781/ijimai.2018.02.004

Hasan, R., Palaniappan, S., Raziff, A. R. A., Mahmood, S., & Sarker, K. U. (2018, August). Student academic performance prediction by using a decision tree algorithm. In *2018 4th international conference on Computer and information sciences (ICCOINS)* (pp. 1-5). IEEE. doi:10.1109/ICCOINS.2018.8510600

Hastie, T., Rosset, S., Zhu, J., & Zou, H. (2009). Multi-class adaboost. *Statistics and Its Interface*, *2*(3), 349–360. doi:10.4310/SII.2009.v2.n3.a8

Hu, J., Luo, G., Li, Y., Cheng, W., & Wei, X. (2008). An AdaBoost algorithm for multi-class classification based on exponential loss function and its application. *Hangkong Xuebao*, *29*(4), 811–816.

Hussain, M., Zhu, W., Zhang, W., Abidi, S. M. R., & Ali, S. (2019). Using machine learning to predict student difficulties from learning session data. *Artificial Intelligence Review*, *52*(1), 381–407. doi:10.1007/s10462-018-9620-8

Injadat, M., Moubayed, A., Nassif, A. B., & Shami, A. (2020). Systematic ensemble model selection approach for educational data mining. *Knowledge-Based Systems*, *200*, 105992. doi:10.1016/j.knosys.2020.105992

Jedidi, Y., Ibriz, A., Benslimane, M., Tmimi, M., & Rahhali, M. (2022). Predicting student performance based on cloud computing. In *WITS 2020* (pp. 113–123). Springer. doi:10.1007/978-981-33-6893-4_11

Kavyashree, K. R., & Laksmi, D. (2016). A review on mining students' data for performance prediction. *International Journal of Advanced Research in Computer and Communication Engineering*, *5*(4), 1104–1106.

Kesorn, K., & Poslad, S. (2011). An enhanced bag-of-visual word vector space model to represent visual content in athletics images. *IEEE Transactions on Multimedia*, *14*(1), 211–222. doi:10.1109/TMM.2011.2170665

Kiu, C. C. (2018, October). Data mining analysis on student's academic performance through exploration of student's background and social activities. In *2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA)* (pp. 1-5). IEEE. doi:10.1109/ICACCAF.2018.8776809

Kumar, N. (2017, December). Machine Intelligence Prospective for Large Scale Video Based Visual Activities Analysis. In *2017 Ninth International Conference on Advanced Computing (ICoAC)* (pp. 29-34). IEEE. doi:10.1109/ICoAC.2017.8441320

Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., & Addison, K. L. (2015, August). A machine learning framework to identify students at risk of adverse academic outcomes. In *Proceedings of the 21st ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1909-1918). doi:10.1145/2783258.2788620

Liu, X., & Niu, L. (2021, December). A Student Performance Predication Approach Based on Multi-Agent System and Deep Learning. In *2021 IEEE International Conference on Engineering, Technology & Education (TALE)* (pp. 681-688). IEEE. doi:10.1109/TALE52509.2021.9678811

Mahela, O. P., Khosravy, M., Gupta, N., Khan, B., Alhelou, H. H., Mahla, R., Patel, N., & Siano, P. (2020). Comprehensive overview of multi-agent systems for controlling smart grids. *CSEE Journal of Power and Energy Systems*, *8*(1), 115–131.

Maraza-Quispe, B., Valderrama-Chauca, E. D., Cari-Mogrovejo, L. H., Apaza-Huanca, J. M., & Sanchez-Ilabaca, J. (2022). A predictive model implemented in knime based on learning analytics for timely decision-making in virtual learning environments. *International Journal of Information and Education Technology (IJIET)*, *12*(2), 91–99. doi:10.18178/ijiet.2022.12.2.1591

Marbouti, F., Diefes-Dux, H. A., & Madhavan, K. (2016). Models for early prediction of at-risk students in a course using standards-based grading. *Computers & Education*, *103*, 1–15. doi:10.1016/j.compedu.2016.09.005

Miguéis, V. L., Freitas, A., Garcia, P. J., & Silva, A. (2018). Early segmentation of students according to their academic performance: A predictive modelling approach. *Decision Support Systems*, *115*, 36–51. doi:10.1016/j.dss.2018.09.001

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, *2*(1), 1–21. doi:10.1186/s40537-014-0007-7

Natek, S., & Zwilling, M. (2014). Student data mining solution–knowledge management system related to higher education institutions. *Expert Systems with Applications*, *41*(14), 6400–6407. doi:10.1016/j.eswa.2014.04.024

Noraziah, A., Fauzi, A. A. C., Ubaidillah, S. H. S. A., Alkazemi, B., & Odili, J. B. (2021). BVAGQ-AR for Fragmented Database Replication Management. *IEEE Access : Practical Innovations, Open Solutions*, *9*, 56168–56177. doi:10.1109/ACCESS.2021.3065944

Pérez, Y. G., & Kholod, I. I. (2020, May). Analysis of Multiagent System for Data Analysis. In *2020 XXIII International Conference on Soft Computing and Measurements (SCM)* (pp. 218-221). IEEE. doi:10.1109/SCM50615.2020.9198765

Ramli, S. F., Firdaus, M., Uzair, H., Khairi, M., & Zharif, A. (2018). Prediction of the unemployment rate in Malaysia. *Int. J. Mod. Trends Soc. Sci*, *1*(4), 38–44.

Raza, M. A., Mokhtar, R., & Ahmad, N. (2019). A survey of statistical approaches for query expansion. *Knowledge and Information Systems*, *61*(1), 1–25. doi:10.1007/s10115-018-1269-8

Shahiri, A. M., Husain, W., & Rashid, N. A. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, *72*, 414–422. doi:10.1016/j.procs.2015.12.157

Sixhaxa, K., Jadhav, A., & Ajoodha, R. (2022, January). Predicting Student's Performance in Exams Using Machine Learning Techniques. In *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 635-640). IEEE.

Sverdlyka, Z., Klynina, T., Fedushko, S., & Bratus, I. (2022). Youtube Web-Projects: Path from Entertainment Web Content to Online Educational Tools. In N. Kryvinska & M. Greguš (Eds.), *Developments in Information & Knowledge Management for Business Applications. Studies in Systems, Decision and Control* (Vol. 421). Springer. doi:10.1007/978-3-030-97008-6_22

Sy Ahmad Ubaidillah, S. H., Ahmad, N., & Beneoluchi Odili, J. (2020). Fragmentation Techniques for Ideal Performance in Distributed Database–A Survey. *International Journal of Software Engineering and Computer Systems*, *6*(1), 18–24. doi:10.15282/ijsecs.6.1.2020.3.0066

Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & Education*, *143*, 103676. doi:10.1016/j.compedu.2019.103676

Vuković, I., Kuk, K., Čisar, P., Banđur, M., Banđur, Đ., Milić, N., & Popović, B. (2021). Multi-agent system observer: Intelligent support for engaged e-learning. *Electronics (Basel)*, *10*(12), 1370. doi:10.3390/electronics10121370

Wanjau, S. K. (2016). *Data mining model for predicting student enrolment in stem courses in higher education institutions*. Academic Press.

Xu, J., Moon, K. H., & Van Der Schaar, M. (2017). A machine learning approach for tracking and predicting student performance in degree programs. *IEEE Journal of Selected Topics in Signal Processing*, *11*(5), 742–753. doi:10.1109/JSTSP.2017.2692560

Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1), 1–19. doi:10.1186/s40561-022-00192-z

Yossy, E. H., & Heryadi, Y. (2019, December). Comparison of Data Mining Classification Algorithms for Student Performance. In *2019 IEEE International Conference on Engineering, Technology and Education (TALE)* (pp. 1-4). IEEE. doi:10.1109/TALE48000.2019.9225887

Yukselturk, E., Ozekes, S., & Turel, Y. K. (2014). Predicting dropout student: An application of data mining methods in an online education program. *European Journal of Open, Distance and E-learning*, *17*(1), 118–133. doi:10.2478/eurodl-2014-0008

A. Noraziah received the Ph.D. degree in database from University Malaysia Terengganu, Malaysia. She is currently an Associate Professor in Faculty of Computing, and a Research Fellow in Centre for Software Development and Integrated Computing, University Malaysia Pahang, Malaysia. She had published 300 scientific research articles. She also supervised more than 20 postgraduate students, and obtained many grants related to her research expertise. She received many international and national awards. She has served as an International Program Committees, and reviewers for many numerous international journals and conferences; and also judges for exhibitions/competitions too. Her research interests include Distributed Database, Data Grid, Data Mining, Big Data, and Computational Intelligence.

M. Rahmah received a PhD from The National University of Malaysia (UKM). She is currently a Senior Lecturer with the Faculty of Computing, Universiti Malaysia Pahang, Pekan, Pahang, Malaysia. She has actively participated in several research and commercialize projects in the domain of computer science. Her current research interests include information retrieval, human–computer interaction, and data processing and manipulation, such as big data and data mining related to artificial intelligence.