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

M.Nazir 1† **A.Noraziah 1†** **M.Rahmah 1†**
pcp21002@student.ump.edu.my a.noraziah@yahoo.com drrahmah@gmail.com
Universiti Malaysia Pahang, Pekan 26600, Malaysia

Abstract:
An effective educational program warrants the inclusion of an innovative construction which 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 (EDSS) has currently been a hot topic in educational systems, facilitating the pupil result monitoring and evaluation to be performed during their development. Insufficient information systems encounter trouble and hurdles in making the sufficient advantage from EDSS owing to the deficit of accuracy, incorrect analysis study of the characteristic, and inadequate database. DMTs (Data Mining Techniques) provide helpful tools in finding the models or forms of data and are extremely useful in the decision-making process. Several researchers have participated in the research involving distributed data mining with multi-agent technology. The rapid growth of network technology and IT use has led to the widespread use of distributed databases. This article explains the available data mining technology and the distributed data mining system framework. Distributed Data Mining approach is utilized for this work so that a classifier capable of predicting the success of students in the economic domain can be constructed. This research also discusses the Intelligent Knowledge Base Distributed Data Mining framework to assess the performance of the students through a mid-term exam and final-term exam employing Multi-agent system-based educational mining techniques. Using single and ensemble-based classifiers, this study intends to investigate the factors that influence student performance in higher education and construct a classification model that can predict academic achievement. We also discussed the importance of multi-agent systems and comparative machine learning approaches in EDSS development.

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

1. 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 percent from 2023 to 2026 [1]. 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. 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 [2]. Relevant information is extracted for analyses of academic sources using EDM (Educational Data Mining), a new discipline for discovering important information using technology. [3]. 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. EDM's strengths rise from

Manuscript received October 5, 2023
Manuscript revised October 20, 2023
https://doi.org/10.22937/IJCSNS.2023.23.10.17

linking data from different domains. It is involved with the features extraction to help in the development of academic process from the tremendous amount of data that the institute provides in [4].

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 the model of their student for the prediction of both the features and performance of every student separately [5]. 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 performances. These assessment systems are quite practical in enhancing the student performance and also the entire academic process [6]. Currently, distributed database is of immense helps by considering the strong features used in the variety of its applications. Data is considered as the key feature of any academic institution for having a safe and right control of organizational data. In the last few years, educational data mining (EDM) has gained much focus among the scholars in improving the quality of higher education. Prediction of the academic performance of students is important for the improvement of the value education among the students. Few research works have been carried out whose focus is the prediction and estimation of students' performance during their higher education. However, the research related to performance forecasting at the higher education is very less, while the higher education level intends to be a standard for representing the learning growth of students at much later educational stages also in their graduate studies. The main contribution of this paper is to explore the important factors affecting the performance of students at the higher education level and to develop an effective classification model using the combination of single and ensemble-based classifiers to predict the educational performance. Most of the earlier researcher-related works highlight using classification for forecasting based on registration data, students' performance in a particular course, grade inflation, expected ratio of failing students, and help in the grading system. To be best of information, very few state-of-the-arts are found which utilize ensemble classification schemes for predicting the students' final results based on their scores. The evaluation of all the courses needed in the study plan will ultimately find the list of courses having a major effect on the results. In this research, the students' performance prediction system employing Multi Agent-based distributed Data Mining is introduced for the prediction of the students' performance depending on their data with improved prediction accuracy and yield assistance to the students who score poor marks using optimization rules. A distributed database is specified as databases 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 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 intercommunications.

A. Highlight Overviews

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

1. We explored and modeled 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 multi-agent system which governs mining approach in 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.

The review article's remaining structure is divided into six sections, starting with the literature review and concluding with the Conclusion. The sequential sections are as follows: Section II Literature Review Represents the comparison of existing approaches for students' performance prediction in higher education institutions. Section III introduces Multi-agent system models and their application. Section IV presents the research gap. Section V The discussion section summarizes the work done by comparing the results of this investigation to the existing literature. Section VI discusses the solutions. Finally, Section VII provides the Conclusion.

2. LITERATURE SURVEY

The student success estimation using data mining methods for the advancement of the wide range of research available is shown below: Natek and Zwilling [7] 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. [8] 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. [9] highlighted the identification of dropout students applying DMTs online. Four DM mechanisms, referred, KNNs (K-Nearest Neighbors), DTs, NBs (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 in prior and hardships faced by them in HEIs were predicted [10]-[11]. Using ANNs (Artificial Neural Networks), SVMs (Support Vector Machines), LRs (Logistic Regressions), NB (Na¨ıve Bayes), and DTs. In the study’s experimental results, ANNs and SVMs achieved higher accuracies (57) based only on demographic data [10], while NBs showed sufficient accuracy, despite being a weak framework [11]. In [12], Kavyashree and Laksmi 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 on the failure of the students may aid the management in rendering counselling and improving their retentions and success in passing 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 are helpful in understanding the issue. Classification is performed by one of the prediction sort of classifiers, which categorizes the data on the basis of the training set and utilizes the pattern for the classification of fresh data. The basic objective of the endeavor is to come up with a design an internetworking application, which makes use of data mining approach during the students’ performance prediction in accordance to their activities. This article investigated academic performance using connections between emotions and student’s economic status with NBs Classifications.

Dwivedi and Singh showed in [13] that Students’ background examinations are beneficial for academic planners in institutions in directing them in proper directions. 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. In educational institutions, a considerable role is played by data mining in every single activity of the institution, be it academic, cultural, assessment, training and placement, etc. EDM can help get filtered data from different departments as DMTs extract hidden information from massive databases, which help predict patterns. They can play important roles in predicting students’ placements from performance data. Shahiri and Husain [14] 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 the students’ performance predictions in Malaysian academic organizations. As a result, holistic literature assessments on student performances were predicted using DMTs. The primary goal of this technical paper was to describe DMTs that were used to predict student achievements. This research work also highlights on the way in which the prediction algorithm can be exploited for finding the extremely significant features in student’s data. Students achievement can actually be improved and they can be successful with much efficiency applying the educational DMTs . There are huge advantages and a best influence to students, teachers and educational institutions. Hasan et al. [15] 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. It is revealed by the results of the simulation that the accuracy of the Random Forest Tree algorithm excels the performance achieved with comparative decision tree algorithms. Therefore, a good concordance is attained for the training set given. Hence, the proposed work helps increase the students' grades in the module. It also helps the stakeholders analyze and evaluate the module delivery and results. Prior detection and solution can be found both at the institutional level and module level. On the academic data set of secondary schools acquired from the ministry of education in Gaza Strip for the year 2015, Amra and Maghari [16] 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 in order to improve student learning. Experiments showed that NBs outperformed KNNs, with a maximum accuracy of 93.6 Fernandes et al. [17] 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 yearsend educational results of student performances, the evaluation of demographic characteristics showed that neighborhoods, schools, and ages were strong indicators of student results. Francis and Babu, [18] 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 students' 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. [19] He 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 Daud et al. [20] 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.

Thirtythree 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) [21] 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. However, research on the application of data fragmentation technology in distributed database systems is still limited. In order to increase application efficiency, more data fragmentation methods should be investigated and implemented in distributed database systems for further research. Migueis et al. [22] 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, these segmented study students based their performances which included failures and higher performances before their degrees started to predict their yearend outcomes. 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 in [23] was conducted to anticipate student enrolments in Kenya's HEIs, including Engineering/Mathematics disciplines and Science and Technologies. 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. In [24], PCAs (Principal Component Analyses) were applied on 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) [25] ,This paper focuses on developing a prediction model of student performance based on cloud computing to improve elearning 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) [26], 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 modeling. The limitation of work is that the skills required by problems are labeled. They must acquire the Q matrix automatically and study in the future how to design the personalized intelligent early warning and motivation mechanisms to enhance the students' enthusiasm for learning. The limitation of work is that they used a limited data set and can be extended to multiple datasets for better results. Maraza-Quispe et al.(2022) [27], 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) [11], proposed model 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) [28], 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) [29], valuable results have been obtained from the suggested proposed approach; this work suffers from some limitations that may have affected two results: 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), [30] present the study for predicting undergraduate students based on the reported pre-university and first-year scores. The work has good accuracy for the small dataset but limitations for the generalized case. Liu and 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, 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) [31] also presented a vital baseline for developing machine learning-based models. Liu et al.(2021) [32] 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 [33]. Noraziah et al.(2021), the proposed algorithm "Binary Vote Assignment on Grid Quorum with Association Rule (BVAGQ-AR)" 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 BVAGQAR algorithm is capable of maintaining data consistency in a distributed en-environment.

The limitation of 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 [34]. Kiu and Ching-Chieh (2018), in this paper, researchers used four supervised educational data mining techniques, Na¨ıve Bayesian, Multilayer Perceptron, Random Forest, Decision Tree J48, and Na¨ıve Bayesian, to predict students' especially mathematic performance in secondary school. They used the attributes of student coursework achievement, student background, and student social activities to analyze and identify their impact on student performance. The prediction was performed on the final grade. A data analytics tool, WAKE, is used to analyze the dataset. This study indicated student social activities and background and significantly predicted student performance from the obtained results. using these model, 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 work is that it can't apply to unsupervised education data mining techniques. [35]. Yossy and 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 neighbor, na¨ıve Bayes, extratree, and bernaoulli na¨ıve 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, three algorithms which have the best performance are random forest, adaboost dan K-Nearest Neighboring. In the future, thisresearch can improve to research more than seven algorithms and can research the most influential features of student performance [36].

A. Education Data Mining

Algarni demonstrated in [37] that DMTs help extracts 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 the 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), [38] It has been 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. This study's objective was to assess the efficacy of the KMean 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 [39].

Costa et al. [40] 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. In order to increase the efficiency of these EDM approaches, the research studies the influence of pre-processing the data, and the algorithms are fine-tuned. The authors accomplished this by comparative examinations of four different approaches to EDM, including SVMs, DTs, NBs, and NNs. Two data sources, data for remote education and on-campus programming courses at the entry-level, were gathered from Brazilian institutions. 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.

Table 1: COMPARISON OF EXISTING APPROACHES FOR STUDENTS' PERFORMANCE PREDICTION IN HIGHER EDUCATION INSTITUTIONS

Author Name	***Methods***	Merits	Demerits
Francis and Babu,[2019]	Hybrid algorithm combining clustering and classification	Better in terms of attaining accuracy in prediction	Performance is poor for big dataset
Francis and 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	Improves the true positive results	Generates error with massive volumes of data
Amra, and Maghari[2017]	KNN and Na¨ıve Bayes	Yields the maximum accuracy value of	Speed of Prediction stage is quite
Xu et al[2017]	Novel machine learning	Yields improved performance	in economic
Natek and Zwilling [2014]	Rep Tree, J48 and M5P models	Yields improved performance	Is not validated for student data sets obtained
Erman Yukselturk et. al	KNN, DT, NB and Neural Network	Yields reasonable accuracy	Has not been tested for other applications
Acharya and Sinha [2014]	C4.5, SMO, NB, 1-Nearest Neighbourhood	Achieves 66 % accuracy	the error rate is increased

B. Student success estimation using Machine Learning Methods.

Acharya and Sinha [41] also used MLTs in their study to predict student performances where gender, revenues, boardlevel marks, and attendances 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), [42] provide a brief overview on an elaborate model, which employs MLTs for identifying students with risks of not succeeding in their high school graduations at stipulated time. Data of nearly two lakh students from two American districts were used for the study's objective of rendering proactive tools both to students and schools, and to aid in the identification and prioritization of students with the risk of negative academic results. Even though the work in this study is confined to predicting students with the possibility of to not completing their high school on stipulated time. The study generalized the suggested model in terms of problem formulations, extractions of features, classifications and evaluations so that negative outputs like not applying to colleges, or under matches could also be obtained. It is hoped is that with the school districts witnessing examples of work like this being developed from their peer organizations, they tend to become extremely knowledgeable, inspired and trained to employ data-driven techniques and are capable of utilizing their resources with efficiency in getting the academic results improved for their students.

In this article, Sixhaxa et al. (2022) [43] propose a model that uses different features like academic, behavioral characteristics, and demographic and examines how these characteristics affect student performance and help identify at-risk students. Five machine learning models that combine classification and regression classifiers are used to analyze features and predict student performance. Logistic Regression, Random Forest, KNearest Neighbour, Gaussian Naive Bayes, and Support Vector Machine are five models used in the research study. The result strongly correlates students' behavioral characteristics and academic performance. 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 and Mustafa (2022) [44], 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. This study uses two parameters, the previous achievement grades and a comparison of machine learning algorithms' performance indicators. The proposed model achieved sufficient accuracy, and according to the results, we can say midterm exam grades are a critical factor in predicting the final grades. The proposed model produced a good result, but they only used three parameters. In the future, the input parameter can be extended to get accuracy. Dabhade et al.(2021), [45] researchers applied a few machine learning algorithms to predict the student's performance of 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. Further, it has also revealed that predicting future performance is dependent on a critical

factor called recent past performance. The small data set works perfectly on this model. A large amount of data is to be used to obtain more accuracy in the results, and we can also apply other machine learning algorithms. Tomasevic et al.(2020), [46] in this research work, three Supervised Machine learning techniques indicate a high risk of dropping out from courses and for student exam performance prediction.

The Artificial Neural Network (ANN) technique provided the best result. These techniques are not sufficient for overall analysis and comparison. So, we should include other emerging educational data mining techniques to better overview and comparison. Extensive data set could be a question for the three techniques used.

Vukovic et al. (2021), [47] 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 the engaging activities are associated with the student's performance? Xu et al [48] 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 factorized 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 UCLA revealed that their proposed technique yielded improved performances compared to standard methods.

Table I1: COMPARISON OF EXISTING APPROACHES FOR STUDENTS' PERFORMANCE PREDICTION IN HIGHER EDUCATION

Author Name	***Methods***	***Merits***	***Demerits***
Sixhaxa et al. (2022), [47]	LR,SVM, RF,KNN, GNV	Results improved performance	misclassified students, Not enough large data set
yajc1 and Mustafa (2022), [48]	LR,SVM,nearest neighbour, RF,KNN	Results improved performance	limited input parameter used
Dabhade et al.(2021), [40]	support vector regression linear algorithm	Results enhanced performance	Not effective on large data set
Vukovic et´ al.(2021), [42]	two different machine learning methods	Increase success rate	possibility of supporting other learning styles
Tomasevic et al.(2020), [41]	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 [43], 2017	Machine Learning Techniques	achieves superior performance	elective co urses are excluded
Lakkaraju et al.(2015), [39]	machine learning approaches	Predict not finish high school on time	Small data set
Acharya and Sinha [38],2014	C4.5, SMOs, NBs, 1-Nearest Neighbourhoods, and MLPs	Yields improved performance	limited collection of data

3. 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 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 modeling 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 [49].

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 behavior. 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 out 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 computer 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 [50].

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 modeling complex systems, smart grids, and computer networks. Despite its broad applicability, MAS remains several challenges, including integrant coordination, security, and task assignments. [51].

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 [52]. 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 the 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 [53].

Raza et al. (2019), [54] 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 and 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. The purpose of this research was 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

4. 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 the 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 function or the cost function since both are Inverse proportionality. The proposed study seeks to address the communication cost and the accuracy in a distributed educational data environment.

5. 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 ultimate outcome for students based on their scores.

To assist students in forecasting their performance, we discussed the role and importance of a multi-agent system, which governs the mining technique in a distributed manner. The literature investigates and identifies prospective study domains to enhance EDSS through optimizing students' psychological performance. Besides identifying research needs, we also compared machine learning methodologies for EDSS's advancement.

6. 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, the 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 the administrators and academic scholars in making decisions about how the instructions are delivered. DM approaches have been applied in several fields, which includes retail business, the medical sector, marketing, banking, bioinformatics, counterterrorism, and several others that are also utilizing it to improve the throughput and efficacy.

A maximum prediction accuracy of the students' 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 are helpful in finding the structures or patterns of data, and it aids extremely during the decision-making process. The future work in this field highlights on 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 of it to be reduced to a bunch of sub-binary classification. In addition, a Performance prediction system can be will developed employing distributed database mechanism combined with Multi Agent model for predicting the students' performance in accordance with their data yielding improved prediction accuracy and be of assistance to the poor performing students using optimization rules.

7. CONCLUSION

The performance of the students is considered one of the critical factors in Higher academic organizations. This is due to the reason that a majority of the academic institute depend on the best record of academic performances. This research work explains about the available data mining technology and the framework of distributed data mining model. This research work also presents an overview on their advantages and drawbacks associated with the estimation of student's performance. At the end, a conclusion is reached that the usage of SAMME boosting approach improves AdaBoost to a multiclass classification with no need for it to be reduced to a set of sub-binary classification. Moreover, designing a Performance prediction system utilizing Multi Agent Data Mining for the student performance prediction depending on their data with improved prediction accuracy and render assistance to the poor.

Acknowledgments

This study is supported by the Fundamental Research Grant Scheme (RDU190185) with Reference no: FRGS/1/2018/ICT03/UMP/02/3 that sponsored by Ministry of Higher Education Malaysia (MOHE). Appreciation is also conveyed to University Malaysia Pahang for project financing under UMP Short Term Grant RDU1903122.

References

[1] Ramli, S. F., Firdaus, M., Uzair, H., Khairi, M., & Zharif, A.: Prediction of the unemployment rate in malaysia, International Journal of Modern Trends in Social Sciences. Vol, vol. 1, pp. 38–44, (2018).

[2] Chen, W., Wang, X., Duan, H., Zhang, X., Dong, T., & Nie, S.: Application of deep learning in cancer prognosisprediction model, Sheng wu yi xue Gong Cheng xue za zhi= Journal of Biomedical Engineering= Shengwu Yixue Gongchengxue Zazhi, vol. 37, no. 5, pp. 918–929, (2020).

[3] Bakhshinategh, B., Zaiane, O. R., ElAtia, S., & Ipperciel, D.: Educational data mining applications and tasks: A survey of the last 10 years, Education and Information Technologies, vol. 23, no. 1, pp. 537–553, (2018).

[4] Chen, P., Lu, Y., Zheng, V. W., Chen, X., & Yang, B.: Knowedu: A system to construct knowledge graph for education, Ieee Access, vol. 6, pp. 31553–31563, (2018).

[5] El Aissaoui, O., El Alami El Madani, Y., Oughdir, L., Dakkak, A., & El Allioui, Y, "A multiple linear regression-based approach to predict student performance," in International Conference on Advanced

Intelligent Systems for Sustainable Development. Springer, pp. 9–23, (2019).
[6] Cuevas, R., Ntoumanis, N., Fernandez-Bustos, J. G., & Bartholomew, K. Bartholomew.: Does teacher evaluation based on student performance predict motivation, well-being, and ill-being?, Journal of school psychology, vol. 68, pp. 154–162, (2018).
[7] Natek, S., & Zwilling, M.: Student data mining solution– knowledge management system related to higher education institutions, Expert systems with applications, vol. 41, no. 14, pp. 6400–6407, (2014).
[8] Hamoud, A., Hashim, A. S., & Awadh, W. A.: Predicting student performance in higher education institutions using decision tree analysis, International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, pp. 26–31, (2018).
[9] Yukselturk, E., Ozekes, S., & Turel, Y. K.: Predicting dropout student: An application of data mining methods in an online education program, European Journal of Open, Distance and e-learning, vol. 17, no. 1, pp. 118– 133, (2014).
[10] Hussain, M., Zhu, W., Zhang, W., Abidi, S. M. R., & Ali, S.: Using machine learning to predict student diffi-culties from learning session data, Artificial Intelligence Review, vol. 52, no. 1, pp. 381–407, (2019).
[11] Marbouti, F., Diefes-Dux, H. A., & Madhavan, K.: Models for early prediction of at-risk students in a course using standards-based grading, Computers & Education, vol. 103, pp. 1–15, (2016).
[12] Kavyashree, K. R., & Laksmi, D.: A review on mining students' data for performance prediction, International Journal of Advanced Research in Computer and Communication Engineering, vol. 5, no. 4, (2016).
[13] T. Dwivedi and D. Singh.: Analyzing educational data through edm process: A survey, International Journal of Computer Applications, vol. 136, no. 5, pp. 13–15, (2016).
[14] Dwivedi, T., & Singh, D.: A review on predicting student's performance using data mining techniques, Procedia Computer Science, vol. 72, pp. 414–422, (2015).
[15] Hasan, R., Palaniappan, S., Raziff, A. R. A., Mahmood, S., & Sarker, K. U.: Student academic performance prediction by using decision tree algorithm, in 2018 4th international conference on computer and information sciences (ICCOINS). IEEE, pp. 1–5, (2018).
[16] Amra, I. A. A., & Maghari, A. Y. Maghari.: Students performance prediction using knn and na¨ıve bayesian, in 2017 8th International Conference on Information Technology (ICIT). IEEE, pp. 909–913, (2017).
[17] Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Van Erven, G.: Educational data mining: Predictive analysis of academic performance of public school students in the capital of brazil, Journal of Business Research, vol. 94, pp. 335–343, (2019).
[18] Francis, B. K., & Babu, S. S.: Predicting academic performance of students using a hybrid data mining approach, Journal of medical systems, vol. 43, no. 6, pp. 1–15, (2019).
[19] Al-Shehri, H., Al-Qarni, A., Al-Saati, L., Batoaq, A., Badukhen, H., Alrashed, S., ... & Olatunji, S. O.: Student performance prediction using support vector machine and k-nearest neighbor, in 2017 IEEE 30th canadian conference on electrical and computer engineering (CCECE). IEEE, pp. 1–4, (2017).
[20] Daud, A., Aljohani, N. R., Abbasi, R. A., Lytras, M. D., Abbas, F., & Alowibdi, J. S.: Predicting student performance using advanced learning analytics, in Proceedings of the 26th international conference on world wide web companion, pp. 415–421, (2017).
[21] UBAIDILLAH, S. H. S. A., & AHMAD, N.: Fragmentation techniques for ideal performance in distributed database–a survey, International Journal of Software Engineering and Computer Systems, vol. 6, no. 1, pp. 18–24, (2020).
[22] Miguéis, V. L., Freitas, A., Garcia, P. J., & Silva, A.: Early segmentation of students according to their academic performance: A predictive modelling approach, Decision Support Systems, vol. 115, pp. 36–51, (2018).
[23] Stephen, K. W.: Data mining model for predicting student enrolment in stem courses in higher education institutions, (2016).
[24] Figueira, A.: Predicting grades by principal component analysis: A data mining approach to learning analyics, in 2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT). IEEE, pp. 465– 467, (2016).
[25] Jedidi, Y., Ibriz, A., Benslimane, M., Tmimi, M., & Rahhali, M.: Predicting student's performance based on cloud computing, in WITS 2020. Springer, pp. 113–123, (2022).
[26] Gao, L., Zhao, Z., Li, C., Zhao, J., & Zeng, Q.: Deep cognitive diagnosis model for predicting students' performance, Future Generation Computer Systems, vol. 126, pp. 252–262, (2022).
[27] Maraza-Quispe, B., Valderrama-Chauca, E. D., Cari-Mogrovejo, L. H., Apaza-Huanca, J. M., & Sanchez-Ilabaca, J.: 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, vol. 12, no. 2, pp. 91–99, (2022).
[28] Bravo-Agapito, J., Romero, S. J., & Pamplona, S.: Early prediction of undergraduate student's academic performance in completely online learning: A five-year study, Computers in Human Behavior, vol. 115, p. 106595, (2021).
[29] Injadat, M., Moubayed, A., Nassif, A. B., & Shami, A.: Systematic ensemble model selection approach for educational data mining, Knowledge-Based Systems, vol. 200, p. 105992, (2020).
[30] Asif, R., Merceron, A., Ali, S. A., & Haider, N. G.: Analyzing undergraduate students' performance using educational data mining, Computers & Education, vol. 113, pp. 177–194, (2017).
[31] Kumar, N.: Machine intelligence prospective for large scale video based visual activities analysis, in 2017 Ninth International Conference on Advanced Computing (ICoAC). IEEE, pp. 29–34, (2017).
[32] Liu, X., & Niu, L.: A student performance predication approach based on multi-agent system and deep learning, in 2021 IEEE International Conference on Engineering, Technology & Education (TALE). IEEE, pp. 681–688, (2021).
[33] Al-Obeidat, F., Tubaishat, A., Dillon, A., & Shah, B.: Analyzing students' performance using multi-criteria classification, Cluster Computing, vol. 21, no. 1, pp. 623–632, (2018).
[34] Noraziah, A., Fauzi, A. A. C., Ubaidillah, S. H. S. A., Alkazemi, B., & Odili, J. B.: Bvagq-ar for fragmented database replication management, IEEE Access, vol. 9, pp. 56168–56177, (2021).
[35] Kiu, C. C.: 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). IEEE, pp. 1–5, (2018).
[36] Yossy, E. H., & Heryadi, Y.: Comparison of data mining classification algorithms for student performance, in 2019 IEEE International Conference on Engineering, Technology and Education (TALE). IEEE, pp. 1–4, (2019).
[37] Algarni, A, "Data mining in education," International Journal of Advanced Computer Science and Applications, vol. 7, no. 6, pp. 456–461,(2016).
[38] Czibula, G., Mihai, A., & Crivei, L. M.: S prar: A novel relational association rule mining classification model applied for academic performance prediction, Procedia Computer Science, vol. 159, pp. 20–29, (2019).
[39] M. Rahmah, M. A. Raza, Z. Fauziah, A. Azhar, M. Fahad, and B. Raza, "Analysis of k-mean and x-mean clustering algorithms using ontology-based dataset filtering," pp. 283–287, (2021).
[40] Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., & Rego, J.: Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses, Computers in human behavior, vol. 73, pp. 247–256, (2017).
[41] Acharya, A., & Sinha, D.: Early prediction of students performance using machine learning techniques, International Journal of Computer Applications, vol. 107, no. 1, (2014).
[42] Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., & Addison, K. L.: A machine learning framework to identify students at risk of adverse academic outcomes, in Proceedings of the

21th ACM SIGKDD international conference on knowledge discovery and data mining, pp. 1909–1918, (2015).
[43] Sixhaxa, K., Jadhav, A., & Ajoodha, R., "Predicting students performance in exams using machine learning techniques," in 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, pp. 635–640, (2022).
[44] Yağcı, M.: Educational data mining: prediction of students' academic performance using machine learning algorithms, Smart Learning Environments, vol. 9, no. 1, pp. 1–19, (2022).
[45] Dabhade, P., Agarwal, R., Alameen, K. P., Fathima, A. T., Sridharan, R., & Gopakumar, G.: Educational data mining for predicting students' academic performance using machine learning algorithms, Materials Today: Proceedings, vol. 47, pp. 5260–5267,(2021).
[46] Tomasevic, N., Gvozdenovic, N., & Vranes, S.: An overview and comparison of supervised data mining techniques for student exam performance prediction, Computers & education, vol. 143, p. 103676, (2020).
[47] Vuković, I., Kuk, K., Čisar, P., Banđur, M., Banđur, Đ., Milić, N., & Popović, B.: Multi-agent system observer: Intelligent´ support for engaged e-learning, Electronics, vol. 10, no. 12, p. 1370, (2021).
[48] Xu, J., Moon, K. H., & Van Der Schaar, M, "A machine learning approach for tracking and predicting student performance in degree programs," IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 5, pp. 742–753,(2017).
[49] Falco, M., & Robiolo, G.: A systematic literature review in multi-agent systems: Patterns and trends, in 2019 XLV Latin American Computing Conference (CLEI). IEEE, pp. 1–10,(2019).
[50] Pérez, Y. G., & Kholod, I. I, "Analysis of multiagent sys-´ tem for data analysis," in 2020 XXIII International Conference on Soft Computing and Measurements (SCM). IEEE, pp. 218–221, (2020).
[51] Dorri, A., Kanhere, S. S., & Jurdak, R.: Multi-agent systems: A survey, Ieee Access, vol. 6, pp. 28573– 28593, (2018).
[52] Mahela, O. P., Khosravy, M., Gupta, N., Khan, B., Alhelou, H. H., Mahla, R., ... & Siano, P.: Comprehensive overview of multi-agent systems for controlling smart grids, CSEE Journal of Power and Energy Systems, (2020).
[53] González-Briones, A., De La Prieta, F., Mohamad, M. S., Omatu, S., & Corchado, J. M.: Multi-agent systems applications in energy optimization problems: A state-ofthe-art review, Energies, vol. 11, no. 8, p. 1928, (2018).
[54] Raza, M. A., Mokhtar, R., & Ahmad, N,: A survey of statistical approaches for query expansion, Knowledge and information systems, vol. 61, no. 1, pp. 1–25, (2019).

Muhammad Nazir has pursuing Ph.D. in Computer Science from Universiti Malaysia Pahang, Malaysia. He obtained his degree in Master of Computer Science from International Islamic University Islamabad, Pakistan, in 2010. In 2006, He obtained degree in Bachelor of Science from University of the Punjab, Pakistan. My research interest includes Data Mining, Educational Data Mining, Big Data, Distributed database and machine Learning.

Noraziah Ahmad received the Ph.D. degree in database from University Malaysia Terengganu, Malaysia. She is currently an Associate Professor with the Faculty of Computing, and a Research Fellow with the Centre for Software Development and Integrated Computing, University Malaysia Pahang, Malaysia. She had published 280 scientific research articles. She also supervised more than 20 postgraduate students, and obtained many grants related to her research expertise. Her research interests include distributed database, data grid, data mining, big data, and computational intelligence.

Rahmah Mokhtar received a degree from The National University of Malaysia (UKM). She is currently a Senior Lecturer with the Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Kuantan, Malaysia. She has actively participated in a number of 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.