Predictive Learning Analytics in Higher Education: Factors, Methods and Challenges

[1*]Ghaith Al-Tameemi, [1]James Xue, [1]Suraj Ajit, [1]Triantafyllos Kanakis, [2]Israa Hadi
[1]*Department of Computing and Immersive Technologies*
University of Northampton
Northampton, United Kingdom
[*]ghaith19912@gmail.com, James.Xue@northampton.ac.uk, suraj.ajit@northampton.ac.uk, triantafyllos.kanakis@northampton.ac.uk
[2]*College of Information Technology, Software Department*
The University of Babylon
Babil, Iraq
israa_hadi1968@yahoo.com

***Abstract*—In higher education institutions, a high number of studies show that the use of predictive learning analytics can positively impact student retention and the other aspects which lead to student success. Predictive learning analytics examines the learning data for intervening or improving the process itself that positively reflects on student performance. In our survey, we are considering the most recent research papers focusing on predictive learning analytics and how that affects the final student outcome in educational institutions. The process of predictive learning analytics, such as data collection, data preprocessing, data mining, and others, has been illustrated in detail. We have identified factors that affect student performance. Several machine learning approaches have also been compared to provide a clear view of the most suitable algorithms and tools used for implementing the learning analytics.**

***Index Terms*—Predictive Learning Analytics, Educational Data Mining, Higher education institutions, Data mining, Student performance.**

I. Introduction

The competition for the higher education sectors has increased to ensure their success at all levels by indicating questions about educating and retaining many students [1]. Educational institutions are tending to beneficially use technologies for increasing and improving their students' performance as well as addressing potential issues in higher educations [1]. For that reason, higher education institutions are collecting a vast amount of data from their students as well as the educational process in the universities [1]. Most of these institutions are facing challenges due to the low numbers of students that are enrolled on their courses, and low progression rate [2]. According to the statistics of the Organization for Economic Cooperation and Development, the number of students in the UK in 2014 were over 1 million and only 71% of them completed their four years of study, while the others were leaving their studies before the final year [2]. Therefore, the financial income of higher education institutions is being affected. A vital element that should be considered to solve the dropout rates of student success is the understanding of the student engagement in the meaningful activities available in the Learning Platform (LP) such as Blackboard or Moodle [3].

978-1-7281-6362-8/20/ $31.00

Learning platform grant flexibility for both the learner and the tutor to capture valuable information that are positively reflected on student performance [3] [4]. Every student's interactions and resources accessed inside the learning platform can be extracted and stored. These interactions and other aspects such as student attendance and grade information have always been recognized as a big dataset, due to the vast amount of information that can be retrieved from the learning platform and analyzed using big data analytics techniques [5] [6]. Consequently, many advantages can be gained by using learning platform. Integrating the learning platform with the analytics and notification features will lead to increase in the student retention in the higher education institutions to achieve better engagement rate, and interventions can be made where appropriate [4].

Learning Analytics (LA) is a new field of practice and research that draws on research and techniques from a discipline that measures, collects and analyses the data about the students and their contexts to understand and enhance the learning [7] [8]. The process of the enhancement should consider several steps such as sourcing, cleaning, transforming and analyzing data to extract meaningful information from the big datasets [9]. Learning analytics is defined as "The emerging field that uses the most sophisticated analytics tool in order to enhance the learning quality" [10] [11]. Higher education institutions have typically relied on methods provided by the learning analytics for educational decision making regarding the learners' performance, predict future performance as well as address potential issues [12]. Apart from learning analytics, data mining has also played an essential role in capturing useful information out of large datasets by encompassing a specific method [11]. Despite the effectiveness and importance behind the use of the learning analytics to improve learning and education, however, there are some obstacles that complicate the analysis process such as the data collection and the ethics and privacy issues behind the use of these data [13] [14]. Whenever the data is available, the higher education institution would have a clear view of the current learner situation, and that would allow them to make the most appropriate and careful decision.

A. Survey aim

The main target of this survey is to identify, discuss, compare and contrast the most relevant and recent research papers related to the topic of predictive learning analytics in higher education in order to identify the significance of this approach and its effect on the learning quality and the student performance. Besides, it is about considering the best scenario to obtain the optimal result in the prediction process by discussing all the existing machine learning algorithms.

B. Paper structure

This paper is structured as follows: section 2 provides the background of the research. Section 3 outlines the factors that are affecting student performance. Section 4 gives a detailed overview of the methods that have been employed to achieve the required information. Finally, the last section concludes the paper and outlines research directions.

II. Research background

In higher education institutions, most of the teachers are struggling with how to come up with a method that improves their student engagement and retention rate [6] [15]. The traditional ways that are used to improve student performance are the student evaluations, and questionnaires about their course, their previous course grades, and the intervention made based on these results, as well as their attendance and the graduation rates [15] [16]. Is that strategy positively affecting the learning quality and student outcomes? It may affect, however, it takes a long time to make the right decision for the students based on this strategy and the data generated using this method is limited [15].

A better approach to be considered for addressing the above issues is to take advantage of the learning analytics as a tremendous amount of learner data is available from online sources. For example, the learner interactions, interest, and engagement that are available in the learning platform can be exploited to improve their learning experience [8] [15]. This data can help to generate a pattern that can predict future events and making the appropriate intervention by the education institutions [15] [17]. Additionally, student data on the online platforms can be beneficially exploited to enhance the teaching, learning, and environments as well as modify the learning practices as it illustrates how or when the students are studying.

Learning analytics has a positive impact on higher education institutions because it can save time and money for education institutions. For the retention purposes, undesirable learning

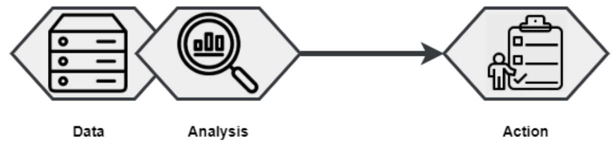

Fig. 1. Learning analytics components

behaviors and emotional conditions can be observed by carefully tracking the learning and persistence of the students, as well as the students at risk can be identified earlier [5] [18]. For instance, if a student does not participate well in the course, his/her interaction in the learning platform has decreased, and his/her attendance rate is also not well, then the student will be less likely to be motivated to stay on the course. Therefore, the utilization of predictive learning analytics will let the students improve their academic performance based on the intervention actions that could be taken by the institutions [2] [5]. In addition, staff can play a crucial role in enhancing a students' retention by providing proper assistance to those who need further support.

Statistical tools and methods are playing a crucial role in analyzing student data and identifying learning strengths and weaknesses. Besides, pattern recognition and predictive techniques can be utilized in the higher education sectors to address students at risk [15]. Moreover, learning analytics, educational data mining, and academic analytics are almost related concepts as all of these concepts aim to explore the relationship between the utilization of the learning platform with a range of results that can undoubtedly reflect on student engagement and retention [19]. Learning analytics has different components as they are shown in Fig. 1. The first component in the learning analytics is raw data, followed by the analysis, which is the process to add intelligence to data using machine learning algorithms. Finally, it is the action which is the critical step towards achieving the purpose of improving the students' performance. Furthermore, learning analytics has five process steps which are capturing, reporting, predicting, acting and refining. These steps are listed in table I.

III. Factors affecting the student performance

There are many factors that could affect student performance. In this section, we will identify the most common factors that can be used for student performance prediction. The factors have been classified into different categories which are student logs in the e-learning or learning platform, historical grades, and performance of the students, student demographics information, and student social information [23].

Student activities on learning platforms have always been considered as useful information that may assist in understanding the student behavior and performance prediction by the

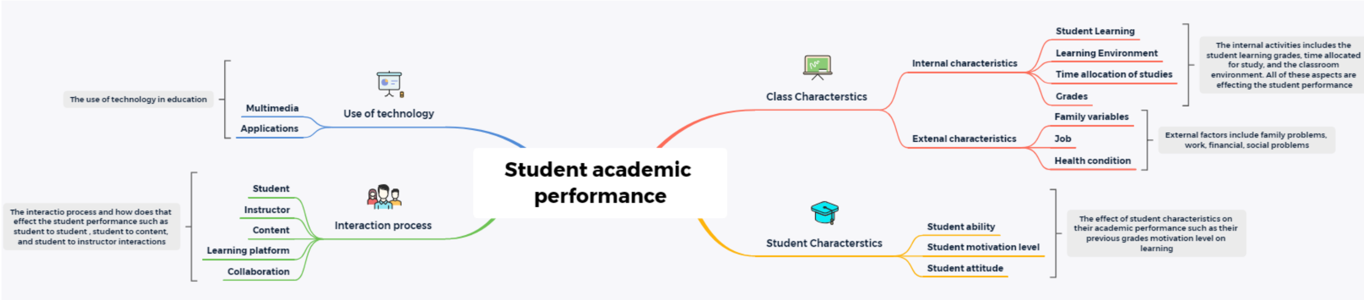

Fig. 2. Factors affecting student academic performance

TABLE I
LEARNING ANALYTIC COMPONENTS

Step	Definition
Capturing	This step involves collecting the data from the learning management system and combining it with the student information [4] [12]
Reporting	After collecting the data, these data will be used to indicate the student's progress and display them in a dashboard for understanding the data in a better way [18] [20].
Predicting	In this step, the data will be used to identify the student outcomes for educational decision making [7] [12] [21].
Acting	After the data is analyzed carefully and the educational decision has been taken, the intervention for the educational institution should be considered for improving the teaching or giving the most appropriate support to the students who are at risk [16] [22].
Refining	This is the last step which can improve the analytical process and ensure the continuity.

utilization of the most appropriate machine learning and data mining algorithms. The activities are student logins, assignments work, quizzes, exploring online, post reading, lectures download, etc [24]. The historical grades and performance of the students have also played a potential role in addressing student performance and outcome. The student environment and demographics can also affect their performance. For instance, the type of school, class, and classroom that a student has taken can be used to predict the student's performance. Another aspect is related to student profile including student age, date of birth, place of birth, and nationality.

Furthermore, this research has shown that instructors can influence student performance, for instance, based on the evaluation result related to an instructor, student performance can be affected in a positive or negative way [25]. The length and difficulty of course or module taken by the student also impacts the student performance. Information about a student's social and personal behavior can be utilized to predict their academic performance and needs. Moreover, course evaluation and student experience are also essential for addressing student performance prediction. Finally, the information related to student satisfaction and the experience related to the course could be beneficially utilized for predicting student performance. Table II summarizes all the factors that affect student performance.

Several aspects have been outlined which can also affect the student performance such as the use of learning platforms, interaction between students, the engagement of the student on the learning platforms, and student ability, motivation as well as the attitude. The interaction between students using features of learning platforms can positively affect student performance and to bridge the gap of the lack of interaction in the traditional learning strategy [26]. Therefore, the use of learning tools on learning platforms will lead to enhancement of the academic performance [27]. In contrast, the isolation of the students in the online environment using a learning platform can be an impediment factor that may affect the student attitude [28].

Student interactions to course contents can also be examined to find out to what extent they contribute to the student academic performance. This aspect is considered as a good predictor of student satisfaction in the course compared to the other aspects [29]. A study has been conducted to determine the impact of multi e-learning systems on the student academic performance, and this study shows that the tools which are provided by the instructor such as e-resources have a positive correlation with the student academic performance [30]. Besides, it also showed that the use of learning platform which contains the essential resources and activities could encourage the students to devote their time to the task [30].

Furthermore, student characteristics have the implicated impact on academic performance. Student skills such as Grade Point Average (GPA) of high school, university GPA, and Scholastic aptitude test can positively affect the student examination, which will affect the student performance [42] [43]. Finally, motivation, attention and the student attitude has also impacted the student academic performance [44] [45].

IV. METHODS

In this section, we will describe the methods that are most frequently used for predicting student performance in higher education institutions. To begin with, we will discuss some of the most challenging parts in predictive learning analytics process including, data collection, data preprocessing which are very essential steps in prediction processs. Then, the most common machine learning algorithms that are used for predictive learning analytics such as neural network, decision tree and naïve bayes are also described. In addition, clustering methods have also been described. Finally, accurate measurement is also illustrated which allows the developer of the

TABLE II
FACTORS AFFECTING STUDENT PERFORMANCE

Factor category	Factor found significant	Data type	Source
Student logs in learning platforms	1- Chatting in a group between the students 2- System logs of a student joining in the virtual room	E-learning activities	[31]
Student logs in learning platforms	1- Assessment activities in the learning platform	E-learning activities	[32]
Student logs in learning platforms	1- Number of virtual classrooms that accessed by the students 2- The different days in the weeks that the student is accessed the virtual classroom 3- Resources that have been visited by the student in the week 4- Number of times that the student visualises the individual resources	E-learning activities	[33]
Student logs in learning platforms	1- The activities that extracted from the online videos in the learning platform 2- The number of questions posted and chat messages 3- Total login times 4- The final grade	E-learning activities	[34]
Student logs in learning platforms	1- Number of home-works that done by the student during the course study 2- Duration of doing the quizzes 3- Number of fail and pass quizzes	E-learning activities	[35], [36], [27]
Student historical grades and class performance	1- Student last education 2- Graduation status	Admission data	[37]
Student historical grades and class performance	1- Activity performance 2- Student attendance 3- Lab experiments 4- Final examination grades 5- Previous semester grades	Admission data	[38]
Student historical grades and class performance	1- Four assignment grades 2- English language score 3-Attendance, performance and study time	Survey and admission data	[39], [40]

TABLE III
LEARNING ANALYTIC COMPONENTS

Data collection technique	Description
Learning platform logs	The data that are collected from the learning platforms (student logs) such as number of clicks, number of downloads (assignments and lectures), time spent in the learning platform, viewing announcements, resources visited, and number of exercised performed), attendance information, and e-resources.
Student information system data	Data that are extracted from the student information system such as demographic data, admission data, and grades.
Surveys	The surveys that taken directly from the students.
Surveys for course evaluation	The typical surveys were taken by the organization to identify the student satisfactory for the course evaluation.
Student social logs	The data extracted based on the students' activities in the university network.

learning analytics process to get a clear view of the accuracy of the techniques that are used.

A. Data collection

Data collection is the most challenging part of the learning analytics process. Data usually collected from different sources such as learning platforms, e-portfolio, e-learning and some other sources for analyzing purposes. There are some techniques the researcher can follow to collect the data such as learning platform activities, student information systems, surveys, course surveys for evaluation purposes, and access logs. Table III summarizes all the mentioned techniques for data collection.

Despite the different techniques of data collection illustrated above, our survey pointed out that most techniques used are the learning platform logs and student information system data. The others such as surveys, course evaluation surveys and the network access logs are less used for analyzing purposes.

B. Data Preprocessing

Data pre-processing involves transforming raw data into an understandable format. The reason to do this is that the real-data almost incomplete, inconsistent and may contains many errors. So, data pre-processing prepares the raw data for further processing. The steps of data pre-processing are data cleaning, integrating, transforming, and reduction [46].

C. Prediction methods

The main aim of prediction is to explore the hidden values such as student performance, score, grades and others [47] [52]. The importance of the predictive methods is to provide proactive measures which can result in improving the student performance [48]. In addition, the utilization of predictive learning analytics can also encourage those students who have low performance in their course study to improve their learning strategy [52]. In this section, we explore the most common

TABLE IV
MOST COMMONLY USED MACHINE LEARNING METHODS FOR LEARNING ANALYTICS

Method	Sources
Naïve bayes	[2, 23, 31, 34, 35, 36, 27, 37, 38, 39, 46, 47, 48, 49, 50, 51].
Decision tree	[23, 31, 34, 35, 36, 27, 37, 38, 39, 41, 52, 48, 49, 53, 50, 51].
Neural network	[12, 21, 22, 31, 34, 36, 27, 39, 46, 52, 48, 54, 55, 49, 56, 53, 57, 50, 51, 58.
Clustering	[12 ,39 ,50 ,59 ,60].

machine learning algorithms that been used to identify those at-risk students, predict student performance, and improve student engagement and retentions.

1) Artificial Neural Network: Artificial Neural Network (ANN) is a machine learning algorithm that is commonly used in the prediction process. It contains a set of inputs and hidden layers that are utilized to achieve the proposed target and the performance affected by the hidden layers and activation function [54]. This approach can be classified as a supervised and unsupervised learning machine learning algorithm. To achieve and solve some of the complex tasks, there are some of the neural network approaches that can be used to achieve whatever targets needed. One of the most common algorithms used for predicting student performance is backpropagation. The network will be trained using the mentioned algorithm, which can have intermediate layers.

The training process of the neural network has different steps. Firstly, the weights of all hidden and output neurons should be initialized. Then, the activation function should be calculated of neurons by computing the signal flow from the input to the output neurons. The direction of the neural network can be represented as follow (input layers, hidden layers, and output layers). The final step is to calculate the neuron weights and compared to the desired output to get the error vectors that used to update the neuron weight in the backward direction. Determining the number of neurons and hidden layers are the essential part of building a neural network model because if there is a small number of neurons [55]. The small number of neurons may cause a reduction in the number of network discrimination power. Alternatively, having many neurons leads to losing the generalization ability in the network [54]. However, neural network needs a considerable amount of input data in order to achieve a satisfactory amount of accurate result [49].

The neural network has played an essential role in predicting student performance and addressing different types of issues in higher education sectors by training and testing historical and time series datasets [52] [56]. For instance, it has been adopted to predict student performance of the Jordanian universities [53]. The prediction of student performance was 96% accurate. Besides, the yearly student performance can also be predicted using neural network approach [57]. Accordingly, we can evidently note that the flexibility of using the neural network approach in prediction process has to provide high accuracy result.

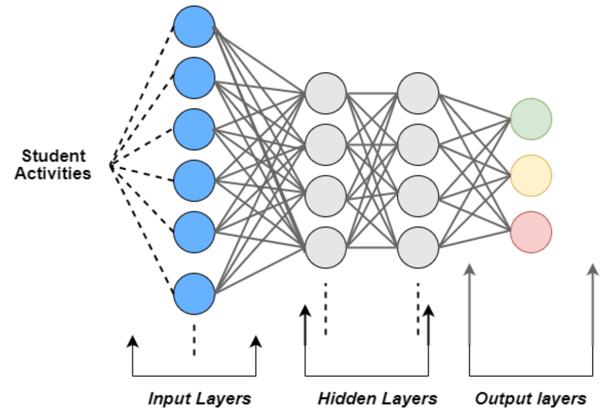

Fig. 3. Neural network architecture for predictive learning analytics

2) Naive Bayes: Naïve Bayes algorithm is a probabilistic classifier which is based on the Bayes theorem and highly performed for data with input dimensions. The Naïve Bayes considered one of the most efficient machine learning algorithms as its works on the isolated features in the datasets as well as the flexibility in the computational process comparing to the other machine learning algorithms. This algorithm is one of the essential machine learning approaches due to the ease of construct; in other words, it does not need complicated iterative parameter estimation schemes. Therefore, it can easily be applied to the vast datasets. In summary, it may not be the best possible machine learning approaches for some application; however, it is commonly applied for learning analytics purpose.

In predictive learning analytics, this technique has widely been used to identify at-risk students, student performance prediction, improve learning strategy, improve student engagement and retention. The naïve bayes has been used to identify at-risk students in early stages based on the score-based grading, and it provides the highest performance accuracy compared to the other techniques [50]. However; this algorithm may not give an accurate prediction result due to the nature of the data [49].

3) Decision Tree: The decision tree is one of the most well-known prediction methods that has multi-level structures based on collected datasets. This technique uses a hierarchy of observations regarding the object to predict the class. The tree comprises of branches and leaves, and each leave represents the class that could belong to while each branch signifies the unification of features which may lead to one of the leaves. This technique is considered as one of the speediest machine learning algorithms in terms of training task, but its speed in the prediction depends on the input values.

A study was conducted in the Open University to predict student performance in a virtual learning environment using Naïve Bayes and decision tree prediction methods [38]. The decision tree has given a higher accuracy percentage than the Naïve Bayes algorithm. Likewise, this technique can be utilized to identify at-risk students on the course so that an appropriate intervention can be taken at an early stage. In

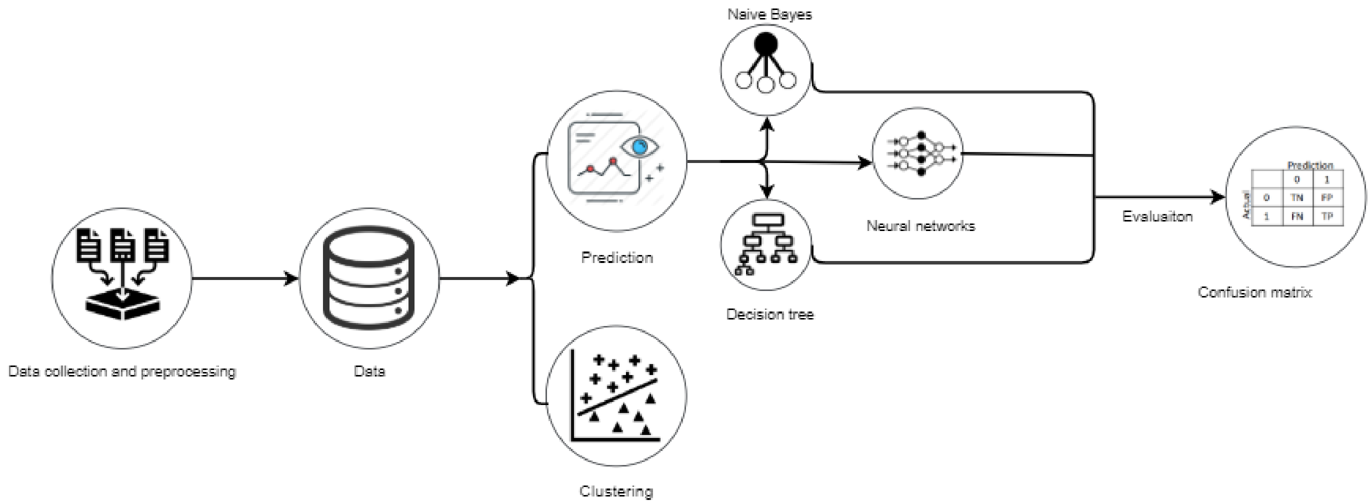

Fig. 4. Process of predictive learning analytics

2019, a model has been built to identify at-risk students who are in the early stage of the semester. This model has been evaluated using eleven different machine learning algorithms, and the decision tree has given the highest accuracy result among the others [51].

D. Clustering

Clustering is an unsupervised method that aims to group similar subsets or objects based on their properties or features to simplify the process of classification. The training or learning does not take place in the clustering methods because they are considered as unsupervised learning methods, and the training sample is unknown previously [61]. Clustering can play an essential role in the classification process because it can be used as a classifier based on the assumption that every cluster corresponding to a class [27].

Firstly, all the class attributes must be removed in order to execute the clustering for training data so that, the mapping between the classes and clusters will be determined to evaluate the obtained clusters as classifiers [27]. Thus, the number of clusters should be matched with the number of classes for obtaining a useful model that associating each cluster to one class. The advantage of using clustering alongside the classification methods is to get a general display of different clusters that generated and that lead to simplify the classification process.

There are many techniques/algorithms for clustering data, such as K-Mean, Mean-shift, and hierarchical clustering. K-mean is parameterized by the number of clusters that the data divide, and then, it assumes the data. While the hierarchical clustering there is no need to specify the number of clusters. In other words, it builds the clusters incrementally by assigning each sample to its own cluster, and then, it merges any two clusters that are similar until all the clusters are merged [61].

The clustering methods are essential in evaluation/ prediction of student performance [59]. One of the effective techniques is the recursive clustering which can be employed to group some students based on a specific students performance to identify those who are prone to fail in their study, and the clustering of the students continues based on the new performance until the end of the course [60]. This technique is beneficial as the students who are in the lower cluster can be intervened to improve their performance before the end of the course [60].

E. Accuracy measurements

Machine learning algorithms evaluation is an essential part of any project to ensure the efficiency of the obtained result. In some cases, some satisfactory results may be achieved using a specific metric, whereas undesirable results using another. In the prediction, classification accuracy measurement is mostly used to evaluate the accuracy of the result; however, this is not enough to judge the model. So, in this section, we will describe all the accuracy measurement approaches [62].

1) Classification accuracy: Classification accuracy is the percentage of the correctly predicted samples to the total number of the input samples. This technique works correctly when the number of samples is equal to each class. For instance, let's assume that we have only 89% samples of the first class and 11% samples of the second class, our proposed model will get 89% of accuracy according to the training dataset. In the same scenario, if we have 70% of class 1 and 30% of class 2, the accuracy measurement will drop down to 70%. In simple words, this technique predicting every training sample related to the first class.

$$Accuracy = \frac{Number of correct prediction}{Number of predicted samples} \tag{1}$$

The real issue of this technique happens when the number of misclassifications of the minor class samples is high.

2) Confusion Matrix: The confusion matrix is a technique used to describe the performance of the classifier on a set of test data based on the known true values [63]. For example, let's assume that we have 265 samples that already predicted using our own classifier and these samples are belong to two classes “Pass and Fail” and we achieved the following result:

- Predicted pass and the actual fail is 15 (FN).
- Actual fail and predicted pass is 3 (FP).
- Actual fail and predicted fail is 87 (TN).
- Actual pass and predicted pass is 160 (TP).

Thus, in our case and based on the accuracy measurement matrix for this technique, we will get the following accuracy measurement result:

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \tag{2}$$

$$Accuracy = \frac{160 + 87}{265} \tag{3}$$

In this scenario, we achieve 93% of accuracy. The form of this matrix considered as the basis for the other types of the accuracy measurement technique.

3) AUC and ROC curve: AUC-ROC is the performance measurement of classification methods. The probability curve represented by the ROC whereas, the AUC represents the measurement of separability. The utilization of this technique is to show the ability of the model for distinguishing between classes. The AUC has two different terms which are true and false positive rates (sensitivities).

V. Conclusion

This survey investigates how predictive learning analytics is affecting the higher educational institutions in terms of student performance, retention, engagement and so many factors. In addition, we have identified the most common factors affecting student performance and the most appropriate predictive learning analysis methods widely used in the higher education sectors. In terms of data collection, our research shows that the most common data collection techniques are the student logs in learning platforms and the student information system (SIS). Our research has shown that the class performance, student activities, as well as the previous grades, can ease the process of the prediction as well as provide an accurate result for the model.

For the machine learning algorithms and specifically for the prediction methods, our survey shows that the Naïve Bayes, Artificial Neural Network, and Decision Tree have widely been used to predict student performance based on the historical datasets, while for the time-series data, the Recurrent Neural Network has widely been used for identifying those who are at-risk situation. Clustering methods have been utilized to group the students based on a specific set of factors, table IV illustrates how the machine learning algorithms frequently used. Finally, we have explained the most eminent techniques to evaluate the accuracy of machine learning algorithms such as classification accuracy, confusion matrix, and AUC-ROC curve.

For future work, the researchers can get benefits from the main outcome of this survey, specifically, the main result that shows the factors that are mostly used to predict student academic performance. Moreover, summarizing imperative machine learning algorithms can also benefit researchers to choose the most appropriate methods for prediction purposes. Based on the information from this survey, one future direction of research is to identify the correlation between student activities on the learning platforms and their final year results using an appropriate machine learning and data mining algorithms.

References

[1] A. Amara , R. Deborah , B. Ayse and M. Mauricio , “Learning Analytics in Higher Education: A Summary of Tools and Approaches,”in 30th ascilite Conference, 1-4 December 2013, Sydney, 2013.

[2] U. Rahila , S. Teo , M. Anuradha and S. Suriadi , “A learning analytics approach: Using online weekly student engagement data to make predictions on student performance,” in International Conference on Computing, Electronic and Electrical Engineering (ICE Cube), Quetta, 2018.

[3] A. Al-Azawei, “What Drives Successful Social Media in Education and E-learning? A Comparative Study on Facebook and Moodle,” Journal of Information Technology Education: Research , vol. 18, pp. 253-274 , 2019.

[4] G. Al-Tameemi and J. Xue, “Towards an Intelligent System to Improve Student Engagement and Retention,” Procedia Computer Science, vol. 151, pp. 1120-1127, 2019.

[5] J. T. Avella, M. Kebritchi, S. G. Nunn and T. Kanai, “Learning Analytics Methods , Benefits , and Challenges in Higher Education : A Systematic Literature Review”,Online Learning, vol. 20, no. 2, 2016.

[6] Engineering Research and Application, vol. 7, no. 6, pp. 25-32, 2017.

[7] T. Yi-Shan and G. Dragan , “Learning Analytics in Higher Education – Challenges and Policies: A Review of Eight Learning Analytics Policies,” in Seventh International Learning Analytics & Knowledge Conference, Vancouver, BC, Canada, 2017.

[8] G. Dragan , “Let’s not forget: Learning analytics are about learning,” TechTrends, pp. 64-71, 2015.

[9] T. R. Rao, P. Mitra, R. Bhatt and A. Goswami, “The big data system, components, tools, and technologies: a survey,” Knowledge and Information Systems, pp. 1-81, 2018.

[10] A. Algarni, “Data Mining in Education,” (IJACSA) International Journal of Advanced Computer Science and Applications, vol. 7, no. 6, pp. 456-461, 2016.

[11] H. Aldowah, A.-S. Hosam and M. F. Wan, “Educational data mining and learning analytics for 21st century,” Telematics and Informatics, vol. 37, p. 13–49, 2019.

[12] M. Alana, M. F. R. A. Joseana and B. C. Evandro, “Monitoring Student Performance Using Data Clustering and Predictive Modelling,” in 2014 IEEE Frontiers in Education Conference (FIE) Proceedings, 2014.

[13] P. Abelardo and S. George , “Ethical and privacy principles for learning analytics,” British Journal of Educational Technology,” vol. 45, no. 3, p. 438–450, 2014.

[14] L. Philipp , E. Markus and E. Martin , “Learning Analytics Challenges to Overcome in Higher Education Institutions,” in Utilizing Learning Analytics to Support Study Success,” Springer, Cham, 2019, pp. 91-104.

[15] P. Nitin , A. Seetharaman , K. Sreekumar and P. Srinivas, “Learning Analytics: Enhancing the Quality of Higher Education, Research Journal of Economics, vol. 2, no. 2, pp. 1-7, 2018.

[16] K. M. L. Jones, “Learning analytics and higher education: a proposed model for establishing informed consent mechanisms to promote student privacy and autonomy,” International Journal of Educational Technology in Higher Education vol. 16, no. 22, pp. 1-22, 2019.

[17] B. T. M. Wong, “Learning analytics in higher education: an analysis of case studies,” Asian Association of Open Universities Journal, vol. 12, no. 1, pp. 21-40, 2017.

[18] J. Ioana, S. Maren , S. Marcus and D. Hendrik , “License to Evaluate: Preparing Learning Analytics Dashboards for Educational Practice,” in 8th International Conference on Learning Analytics and Knowledge, Sydney, New South Wales, Australia, 2018.

[19] E. Wagner and P. Ice, “Data Changes Everything: Delivering on the Promise of Learning Analytics in Higher Education,” EDUCAUSE Review, vol. 47, pp. 32-36, 2012.

[20] P. M. Leah, D. Shane, P. Abelardo and G. Dragan, "Embracing Big Data in Complex Educational Systems: The Learning Analytics Imperative and the Policy Challenge," Research & Practice in Assessment, vol. 9, pp. 17-28, 2014.

[21] L. Wei-Xiang and C. Ching-Hsue , "A hybrid method based on MLFS approach to analyze students' academic achievement," in 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 2016.

[22] H. Jui-Long , E. S. Brett , Y. Juan and D. Xu , "Improving Predictive Modeling for At-Risk Student Identification: A Multistage Approach," IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, vol. 12, no. 2, pp. 1-10, 2019.

[23] H. Mushtaq , Z. Wenhao , W. Zhang and A. Syed Muhammad Raza , "Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores," Computational Intelligence and Neuroscience, vol. 2018, pp. 1-22, 2018.

[24] K. Nikolas and O. Dijana, "Analysis of Student Behavior and Success Based on Logs in Moodle," in MIPRO 2018: 41st International Convention, Opatija, Croatia, 2018.

[25] B. R. A. B. Z. Z. M. H. Christothea Herodotou, "A large scale implementation of predictive learning analytics in higher education: the teachers' role and perspective," Education Tech Research Dev, vol. 67, p. 1273–1306, 2019.

[26] C. S. Nair and R. Patil, "A Study on the Impact of Learning Management Systems on Students of a University College in Sultanate of Oman," IJCSI International Journal of Computer Science , vol. 9, no. 2, pp. 379-385, 2012 .

[27] C. Romero, M.-I. López, J.-M. Luna and S. Ventura, "Predicting students' final performance from participation in on-line discussion forums," Computers & Education, vol. 68, p. 458–472, 2013.

[28] R. Faizi, R. Chiheb and A. El Afia, "Students' Perceptions Towards Using Web 2.0 Technologies in Education," Internation Journal of Emerging Technologies in Learning, vol. 10, no. 6, pp. 32-36, 2015.

[29] S. Chan and D. Bose, "Engage Online Learners: Design Considerations for Promoting Student Interactions," in IGI Global, 2017, pp. 96-118.

[30] R. Burk, P. Lyons, A. Noriega and D. Polovina-Vukovic, "The Impact of Multiple Electronic Resources on Student Academic Performance," Higher Education Quality Council of Ontario, Ontario, 2013.

[31] G. R. P. E. G. S. Xing Wanli, "Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory," Goggins Sean, vol. 47, pp. 168-181, 2015.

[32] M. Imani and S. M. Joel , "Using Learning Analytics to Predict Students' Performance in Moodle Learning Management System: A Case of Mbeya University of Science and Technology," The Electronic Journal of Information Systems in Developing Countries, vol. 79, no. 1, pp. 1-13, 2017.

[33] J. A. Lara, D. Lizcano, M. A. Martínez, J. Pazos and T. Riera, "A system for knowledge discovery in e-learning environments within the European Higher Education Area – Application to student data from Open University of Madrid, UDIMA," Computers & Education, vol. 72, pp. 23-36, 2014.

[34] A. A. Saa, "Educational Data Mining & Students' Performance Prediction," (IJACSA) International Journal of Advanced Computer Science and Applications, vol. 7, no. 5, pp. 212-220, 2016.

[35] A. Zafra and S. Ventura, "Multi-instance genetic programming for predicting student performance in web based educational environments," Applied Soft Computing, vol. 12, p. 2693–2706, 2012.

[36] S. M.-. H. Hasheminejad and M. Sarvmili, "Students' performance prediction based on particle swarm optimization," Journal of AI and Data Mining, vol. 7, no. 1, pp. 77-96, 2019.

[37] S. Helal, J. Li, L. Liu, E. Ebrahimie, S. Dawson, D. J. Murray and Q. Long, "Predicting academic performance by considering student heterogeneity," Knowledge-Based Systems, vol. 161, pp. 134-146, 2018.

[38] E. N. Azizah, U. Pujianto, E. Nugraha and Darusalam, "Comparative performance between C4.5 and Naive Bayes classifiers in predicting student academic performance in a Virtual Learning Environment," in 4th International Conference on Education and Technology (ICET), Indonesia, 2018.

[39] H. Almarabeh, "Analysis of Students' Performance by Using Different Data Mining Classifiers," I.J. Modern Education and Computer Science, vol. 8, pp. 9-15, 2017.

[40] S. B. Kotsiantis, "Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades," Artificial Intelligence Review, vol. 37, no. 4, p. 331–344, 2012.

[41] A. Abazeed and M. Khder, "A Classification and Prediction Model for Student's Performance in University Level," Journal of Computer Science, vol. 13, no. 7, pp. 228-233, 2017.

[42] K. Adair and O. H. Swinton, "Lab Attendance and Academic Performance," International Scholarly Research Network ISRN Education, vol. 2012, pp. 1-5, 2012.

[43] A. A. Aden, Z. Abdiqadir Yahye and A. Mohamed Dahir, "The Effect of Student's Attendance on AcademicPerformance: A Case Study at Simad University Mogadishu," Academic Research International, vol. 4, no. 6, pp. 409-417, 2013.

[44] M. Credé, S. G. Roch and U. M. Kieszczynka, "Class Attendance in College: A Meta-Analytic Review of the Relationship of Class Attendance With Grades and Student Characteristics," Review of Educational Research, vol. 80, no. 2, pp. 272-295, 2010.

[45] A. Williams, E. Birch and P. Hancock, "The impact of online lecture recordings on student performance," Australasian Journal of Educational Technology, vol. 28, no. 2, pp. 199-213, 2012.

[46] G. Akçapınar, A. Altun and P. Aşkar , "Using learning analytics to develop early-warning system for at-risk students," International Journal of Educational Technology in Higher Education, vol. 16, no. 40, pp. 1-20, 2019.

[47] Q. Zhou, Y. Zheng and C. Mou, "Predicting students' performance of an offline course from their online behaviors," in Fifth International Conference on Digital Information and Communication Technology and its Applications (DICTAP), Beirut, 2015.

[48] M. . V. Amazona and A. A. Hernandez, "Modelling Student Performance Using Data Mining Techniques: Inputs for Academic Program Development," in 5th International Conference on Computing and Data Engineering, Shanghai, China, 2019.

[49] V. L. Uskov, . J. P. Bakken , A. Byerly and A. Shah, "Machine Learning-based Predictive Analytics of Student Academic Performance in STEM Education," in IEEE Global Engineering Education Conference (EDUCON), Dubai, United Arab Emirates, 2019.

[50] F. Marbouti, H. A. Diefes-Dux and K. Madhavan, "Models for early prediction of at-risk students in a course using standards-based grading," Computers & Education, vol. 103, pp. 1-15, 2016.

[51] I. Khan, A. Al Sadiri, A. R. Ahmad and N. Jabeur, "Tracking Student Performance in Introductory Programming by Means of Machine Learning," in 4th MEC International Conference on Big Data and Smart City (ICBDSC), Muscat, Oman, 2019.

[52] L. Mutanu and P. Machoka, "Enhancing Computer Students' Academic Performance through Predictive Modelling - A Proactive Approach," in The 14th International Conference on Computer Science & Education (ICCSE 2019), Toronto, Canada, 2019.

[53] Y. S. Alsalman, N. K. Abu Halemah, E. S. AlNagi and W. Salameh, "Using Decision Tree and Artificial Neural Network to Predict Students Academic Performance," in 10th International Conference on Information and Communication Systems (ICICS), Irbid-Jordan, 2019.

[54] O. Fumiya , Y. Takayoshi , S. Atsushi and O. Hiroaki , "A Neural Network Approach for Students' Performance Prediction", in the Seventh International Learning Analytics & Knowledge Conference, Vancouver, British Columbia, Canada, 2017.

[55] A. A. B. D. A. o. d. S. L. B. P. d. N. J. J. d. M. S. J. J. C. d. N. Rosangela Marques de Albuquerque, "Using Neural Networks to Predict the Future Performance of Students," in International Symposium on Computers in Education (SIIE), Setubal, 2015.

[56] Y. Tsung-Yen , G. B. Christopher , J.-W. Carlee and C. Mung , "Behavior-Based Grade Prediction for MOOCs Via Time Series Neural Networks," IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, vol. 11, no. 5, pp. 716-728, 2017.

[57] F. Sikder, J. Uddin and H. Sajal , "Predicting Students Yearly Performance using Neural Network: A Case Study of BSMRSTU," in 5th International Conference on Informatics, Electronics and Vision (ICIEV), Dhaka, Bangladesh, 2016.

[58] W. . W. Fok, Y. He, H. A. Yeung, K. Law, K. Cheung, Y. Ai and P. Ho, "Prediction Model for Students' Future Development by Deep Learning and Tensorflow Artificial Intelligence Engine," in IEEE International Conference on Information Management, Oxford, UK, 2018.

[59] U. Ninrutsirikun, B. Watanapa, C. Arpnikanondt and V. Watananukoon, "A Unified Framework for Student Cluster Grouping with Learning Preference Associative Detection for Enhancing Students' Learning

Outcomes in Computer Programming Courses" in Global Wireless Summit (GWS), Chiang Rai, Thailand, 2018.
[60] A. V. K, A. R. S.K, E. B. George and H. A.S. , "Recursive clustering technique for students' performance evaluation in programming courses," in Majan International Conference, Muscat, Oman, 2018.
[61] L. Rokach, "A survey of Clustering Algorithms," in Data Mining and Knowledge Discovery Handbook, Boston, MA, Springer US, 2010, pp. 269–298.
[62] P. Galdi and R. Tagliaferri, "Data Mining: Accuracy and Error Measures for Classification and Prediction," in Encyclopedia of Bioinformatics and Computational Biology, Elsevier, 2019, pp. 431-436.
[63] J. L. Garcia-Balboa, M. V. Alba-Fernandez, F. J. Ariza-López and J. Rodriguez-Avi, "Homogeneity Test for Confusion Matrices: A Method and an Example," in IGARSS 2018 - IEEE International Geoscience and Remote Sensing Symposium, Valencia, 2018.