International Journal of Advancement in Electronics and Computer Engineering (IJAECE)
Volume 12, Issue 1, Jan. 2023, pp.312-318, ISSN 2278 -1412

Pattern Analysis in Education Data Mining

Rhicha Tiwari[1], Dr. Prabhat Pandey[2]
[1]Research Scholar, APSU, Rewa(M.P.), India
[2] Prof. Computer Science OSD, APSU, Rewa(M.P.), India

Abstract *– Educational organizations play a vital role in the growth and development of any nation, making them an integral part of our society. The mining of information systems within educational institutions holds significant potential for the benefit of both students and the institutions themselves. By leveraging data mining techniques, a proposed system can contribute to the improvement of students' performance in various aspects such as academics, co-curricular and extra-curricular activities, behavior, and overall achievements.*

Through the implementation of this system, educational institutions can derive valuable insights, including the scope of different courses, identification of high-performing students, areas for improvement, and addressing job placement concerns. Data mining techniques such as data cleaning, data integration, and classification and regression analysis are utilized to extract meaningful information. The system also incorporates the use of GoogleVis to present graphical outputs, enhancing the visual representation of the data.

The ultimate goal of this system is to instill confidence in the efficacy of data mining techniques, encouraging their adoption as strategic management tools within the current educational and business systems. By embracing these techniques, educational organizations can effectively leverage their data to make informed decisions and drive improvements in both student performance and institutional outcomes.
Keywords: **Rectfier, Current inject techinique, Diode Rectifer, Power Improvement**

I. Introduction

The field of leveraging educational data through data mining techniques is known as "educational data mining," while the use of patterns derived from educational or instructional data is referred to as "learning analytics." The concept of data mining originated with the development of electronic record-keeping and the utilization of structured records to extract knowledge through pattern analysis. As data availability has increased, new approaches have emerged to transform this data into actionable insights. Learning analytics, similar to management analytics and business intelligence, focuses on utilizing educational data to inform teaching and learning practices.

Empirical research generally considers learning analytics and educational data mining as related endeavors that employ advanced statistical learning methods to analyze learning and educational data. While some view these fields as interchangeable, others argue for a clear distinction between them. The primary argument revolves around the recognition that learning and education are distinct concepts, despite some overlap. Learning analytics primarily examines the processes that influence individual and social learning, while educational data mining focuses on knowledge discovery from all educational data sources generated by individuals and groups within institutional frameworks. In practice, it can be challenging to establish a clear boundary since institutional factors influence individual learning behaviors. Furthermore, there is substantial overlap in research topics and analytical methods within the empirical literature.

Papamitsiou and Economides (2014) summarize the key difference between learning analytics and educational data mining as follows: "Learning analytics adopts a holistic framework, seeking to understand systems in their full complexity. On the other hand, educational data mining adopts a reductionist viewpoint by analyzing individual components, seeking new patterns in data and modifying respective algorithms." However, Aldowah (2019) identifies four dimensions of research that span both disciplines: computer-supported learning analytics, computer-supported predictive analytics, computer-supported behavioral analytics, and computer-supported visualization analytics.

While a theoretically informed study distinguishes between the contributions of learning analytics, educational data mining, and human-computer interaction, the overarching theory that influences all three disciplines is activity theory, as formulated by Leontev (1981) and collaborators. Constructivist psychology has also influenced all three disciplines, particularly in terms of shared interests in group processes and social interaction. Although disciplinary exchange can help bridge differences, it remains uncertain whether these two disciplines are truly distinct or if the differences are more taxonomic than practical. The fruitful exchange between the two disciplines suggests that educational data mining and learning analytics may belong to the same genus, if not the same species.

Romero and Ventura (2020) report that most research in educational data mining and learning analytics is conducted in the same settings, such as virtual learning environments, learning management systems, cognitive tutors, and other computer-based learning environments including mobile devices, massive open online courses (MOOCs), and social learning platforms. Researchers collect various types of data, including login frequency, chat messages, questions submitted to instructors, response times, resource access, grades, profiles, preferences, forum posts, affect observations, and more. Analytical methods employed include classification, clustering, regression, and model discovery. Evaluation criteria for comparing methods encompass precision, accuracy, sensitivity, coherence, similarity weights, fitness measures, and more. The main research areas focus on student behavior modeling and prediction of performance, enhancing students' and teachers' reflection and awareness, and improving feedback and assessment services. Overall, assessing the differences between these two fields would require a systematic analysis of their research production, as research communities are ultimately defined by their output..

II. METHOD

Neural network is a biologically inspired analytical method capable of modeling extremely complex nonlinear functions. Neural networks are part of the popularly employed algorithms in the education domain for student performance prediction. ANN is mostly desired because it can classify patterns without requiring any training. Being inherently parallel and thus able to speed up the computational process makes ANN suitable for prediction activities in the educational data mining domain.

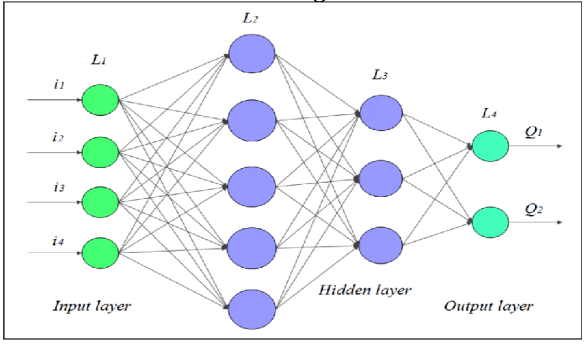


Figure 1 Neural Network

The equation for the neural network is a linear combination of the independent variables and their respective weights and bias (or the intercept) term for each neuron. The neural network equation looks like this:

$Z = Bias + W_1X_1 + W_2X_2 + \ldots + W_nX_n$

where,

- Z is the symbol for denotation of the above graphical representation of ANN.
- Wis, are the weights or the beta coefficients
- Xis, are the independent variables or the inputs, and
- Bias or intercept = W_0

K- nearest-neighbor classifiers are an analogy-based algorithm that learns by comparing themselves with similarly given test training tuples. The algorithms can be employed in numeric predictions to return a real-valued forecast for an unidentified tuple. The algorithm, therefore, restores the reasonable value associated with the k-nearest neighbors of the unidentified tuple. When used for student performance prediction, KNN gave good results [20].

K- nearest neighbor is supervised machine learning algorithm. It is the simplest yet powerful technique that can be used for both classification and regression predictive problems. The basic concept of KNN is to classify the test data in a given dataset by using feature similarity. It calculates the distance (closeness or proximity) between the test data and each training data in the dataset. Then it performs the majority voting and classifies the test data by the majority votes of neighbor classes. The distance can be calculated by using various distance functions like Euclidean, Cosine, Chi-square, Minkowsky, etc. The most intuitive nearest neighbour type classifier is the one nearest neighbour classifier that assigns a point x to the class of its closest neighbour in the feature space, that is

C_n^1nn (x)=Y_((1))

As the size of training data set approaches infinity, the one nearest neighbour classifier guarantees an error rate of no worse than twice the Bayes error rate (the minimum achievable error rate given the distribution of the data).

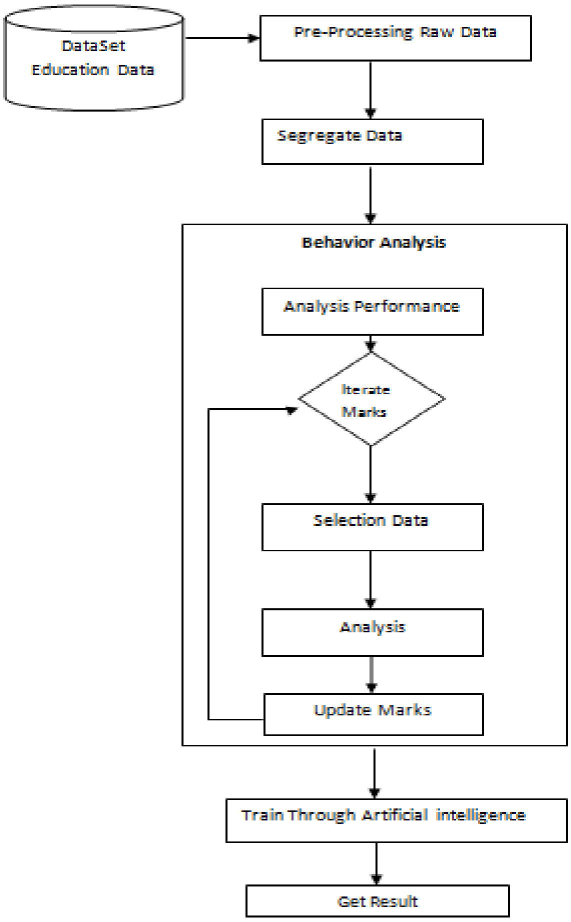
The naive bayes method is a supervised classifier that bases on applying the bayes' theorem with strong naive independence assumptions between the explanatory variables and uses two simplifications where one uses the conditional independence assumption and the other ignores the denominator [4]. This algorithm is also one of the commonly preferred methods by researchers to carry out prediction activities since it learns fast, predicts equally, and does not need ample storage.

International Journal of Advancement in Electronics and Computer Engineering (IJAECE)
Volume 12, Issue 1, Jan. 2023, pp.312-318, ISSN 2278 -1412

Naive Bayes is a classification algorithm that assumes that the predictor variables are independent of each other. The base of the naive Bayes is the Baye's theorem which is derived from the conditional probability. It classifies the test data by computing conditional probability with feature vectors which belong to particular class. Naive Bayes algorithms can be applied in recommendation system spam filtering, sentiment analysis.
The Naive Bayes Classifier is inspired by Bayes Theorem which states the following equation:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

It is a very popular machine learning technique. It can be used to perform both classification and regression. The core idea of SVM is that it tries to find out a hyperplane that separates two classes as widely as possible. In other words, it finds the hyperplane that maximizes the margin. As margin increases the generalization accuracy increases. The points through which the hyperplane passes are called support vectors. The variations to SVM are linear SVM, Polynomial kernel SVM, Radial Basis Function SVM.



III. RESULT

The results of the commonly used student performance prediction methods above from 2010 to 2020 were analyzed and used to plot a graph of how they differ in their overall prediction accuracy. The diagram is shown in figure 4.1.

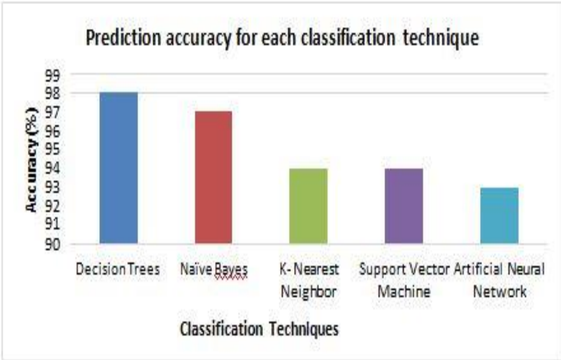
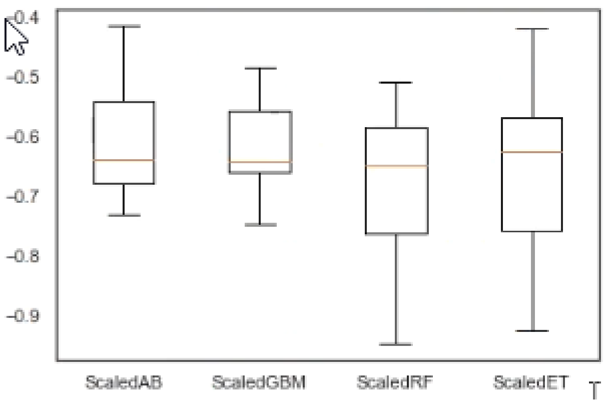


Figure 2: Performance Prediction Accuracy

Figure2 was obtained by studying the commonly used methods for student performance prediction in previous studies and their performances were compared with an aim of finding out the method with the highest accuracy, and the results were ploted on the graph in figure 4.1. The Figure overall shows the prediction accuracy of the conventional prediction methods used to predict student performance that the authors examined in this study, all grouped by their algorithms from 2012 to 2020. As shown in figure 4.1, the decision trees method gave a high prediction accuracy of (98%), compared to all other methods. Naive bayes was the second-highest accurate with (97%). Then followed by the k-nearest neighbor method and support vector machine method, which gave a similar accuracy of (94%). Artificial neural network gave the least student performance prediction accuracy of (93%).



```
Best: -0.620034 using {'n_estimators': 50}
-0.620034 (0.067908) with: {'n_estimators': 50}
-0.620034 (0.067908) with: {'n_estimators': 100}
-0.620034 (0.067908) with: {'n_estimators': 150}
-0.620034 (0.067908) with: {'n_estimators': 200}
-0.620034 (0.067908) with: {'n_estimators': 250}
-0.620034 (0.067908) with: {'n_estimators': 300}
-0.620034 (0.067908) with: {'n_estimators': 350}
-0.620034 (0.067908) with: {'n_estimators': 400}
0.7196023220437311
```

Figure 3 Data processing processes

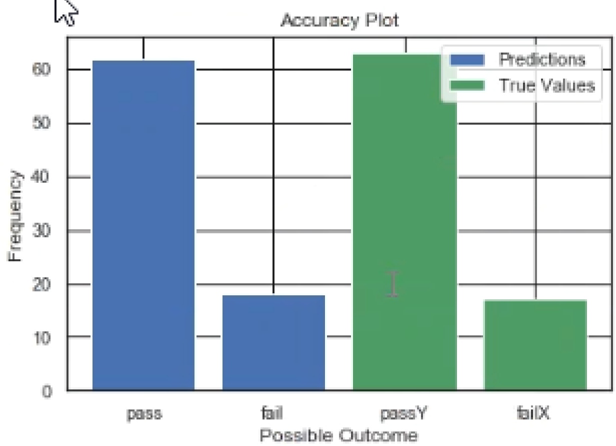


Figure 4 Possibility Outcome using Train and Testing Data

```
Currently on: Linear regression...
Accuracy For Linear regression:0.8237547892720306
Accuracy: 6
Precision: 9
Recall: 36
F-measure: 208
Linear regression took: 0.04388308525085449seconds
```

Figure 4.4 Cross validation of linear regression

```
Accuracy For SVM:0.8275862068965517
Accuracy: 12
Precision: 15
Recall: 25
F-measure: 202
SVM took: 1.5789380073547363seconds
```

Figure 5 Cross validation of SVM

```
Currently on: Random Forest...
Accuracy For Random Forest:0.8084291187739464
Accuracy: 6
Precision: 9
Recall: 31
F-measure: 203
```

Figure 6 Cross validation of random forest

```
Currently on: Decision Tree...
Accuracy For Decision Tree:0.8084291187739464
Accuracy: 6
Precision: 9
Recall: 31
F-measure: 203
Decision Tree took: 0.17970037460327148seconds
```

Figure 7 cross validation of decision tree

```
FutureWarning
Linear regression : 0.914819 (0.052546)
LDA : 0.895665 (0.050663)
KNN : 0.899194 (0.043213)
Random Forest : 0.857762 (0.046514)
DecisionTree : 0.882762 (0.049514)
SVM : 0.892843 (0.065787)
```

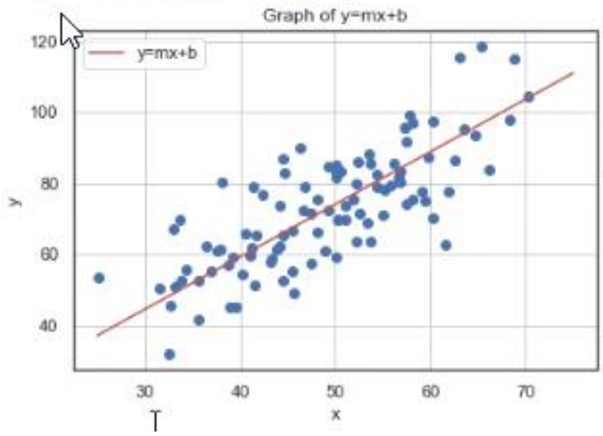
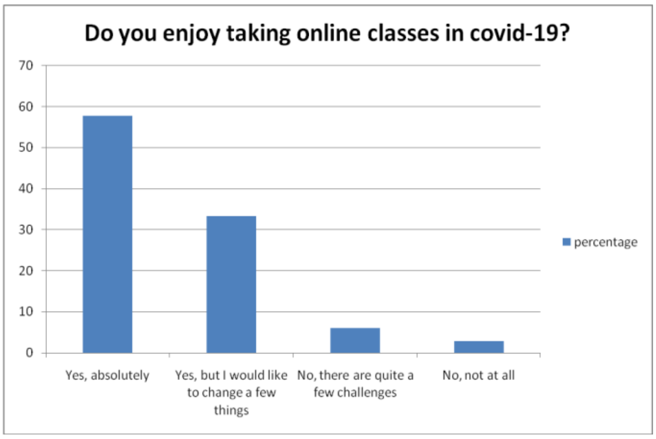


Figure 8 cross validation of Kappa Squared Error for ML Classifier

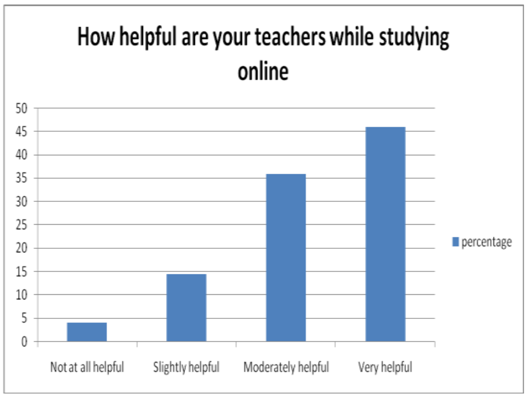
```
Linear regression: 0.6723
SVM: 5421.5080
Random Forest: 0.0889
Decision Tree: 1.4777
```

Figure 4.9 Cross validation of mean error



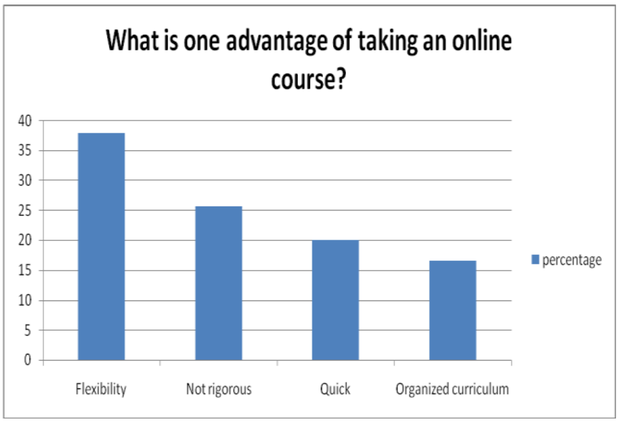
Graph 1 Do you Enjor Taking Online Classes

According to the chart, respondents' enjoyment of taking online classes in the COVID-19 pandemic varied. The majority, accounting for 57.8% of the total, indicated that they enjoyed online classes "Yes, absolutely." Additionally, 33.4% of the respondents expressed that they enjoyed online classes but would like to change a few things to enhance their experience. A smaller proportion of respondents, 6% of the total, mentioned that they did not enjoy online classes due to several challenges they faced. Only 2.8% of the respondents indicated that they did not enjoy online classes at all.



Graph 2 How Helpful your teacher during online classes

The chart provides information about respondents' perceptions of the helpfulness of their teachers while studying online. It shows the number and percentage of respondents who indicated their level of helpfulness.
According to the chart, respondents' perceptions varied regarding the helpfulness of their teachers while studying online. The majority, accounting for 45.8% of the total, considered their teachers to be "Very helpful." Approximately 35.8% of the respondents found their teachers to be "Moderately helpful." A smaller proportion, 14.4% of the respondents, described their teachers as "Slightly helpful." However, a minority of respondents, only 4% of the total, indicated that their teachers were "Not at all helpful" while studying online



Graph 3 Advantage of online classes

According to the chart, respondents identified several advantages of taking an online course. The most commonly mentioned advantage was "Flexibility," with 37.8% of the respondents recognizing this as a benefit. Additionally, 25.6% of the respondents indicated that online courses are "Not rigorous," implying that they perceive online courses to be less demanding or intense compared to traditional in-person classes. Another advantage mentioned by 20% of the respondents was that online courses are "Quick," which could refer to the efficient delivery of course content or the shorter duration of online programs. Lastly, 16.6% of the respondents appreciated the "Organized curriculum" in online courses, suggesting that the structured and well-planned nature of the curriculum is beneficial.

IV. Conclusion

Learning analytics is being employed by higher education institutions to enhance the services they offer and improve tangible and measurable outcomes such as grades and student retention. Even K-12 schools and school districts are beginning to adopt institutional-level analyses to identify areas for improvement, establish policies, and measure results.

With advancements in adaptive learning systems, there is now potential to leverage the power of feedback loops at the individual level for both teachers and students. By measuring and making students' learning and assessment activities visible, they can develop skills in monitoring their own progress and directly witness the positive impact of their efforts on their success. Teachers gain insights into students' performance, enabling them to adapt their teaching methods or implement interventions such as tutoring and tailored assignments. Adaptive learning systems provide educators with rapid feedback on the effectiveness of their adaptations and interventions, facilitating continuous improvement.

The rapid adoption of online and blended learning environments, coupled with the availability of open-source tools for adaptive learning systems and improved understanding of data interpretation, is leading to fundamental shifts in teaching and learning systems. As educational content moves online and mobile devices enable constant interaction with the content, educational data mining and learning analytics enable continuous assessment of learning. This empowers educators at all levels to grasp the potential offered by these developments..

References

[1] Basuki Wibawa, Jenny S. Siregar, D. A. Asrorie, and Halimatus Syakdiyah "Learning Analytic and Educational Data Mining for Learning Science and Technology" 2021.

[2] Myla M. Arcinas, Guna Sekhar Sajja, Shazia Asif , Sanjeev Gour, Ethelbert Okoronkwo, Mohd Naved "Role Of Data Mining In Education For

Improving Students Performance For Social Change” ISSN 2651-4451 2021.

[3] Kousuke Mouri, Fumiya Suzuki, Atsushi Shimada, Noriko Uosaki, Chengjiu Yin, Keiichi Kaneko & Hiroaki Ogata “Educational data mining for discovering hidden browsing patterns using non-negative matrix factorization” ISSN: 1049-4820 2021.

[4] Olga A. Fiofanova “Methodology of Data-Driven Pedagogy and the Development of a Culture of Analysis of Educational Data in Pedagogical Communities” 2020.

[5] Lamya F. Daghestani1 | Lamiaa F. Ibrahim2,3 | Reem S. Al-Towirgi1,4 | Hesham A. Salman “Adapting gamified learning systems using educational data mining techniques” 2020.

[6] Angelos Charitopoulos & Maria Rangoussi1 & Dimitrios Koulouriotis “On the Use of Soft Computing Methods in Educational Data Mining and Learning Analytics Research: a Review” 2020.

[7] Akarshita Tripathi, Mr. Amit Kumar “Analysis of Educational Data Mining Techniques”2019.

[8] Ryan S. Baker “Challenges for the Future of Educational Data Mining: The Baker Learning Analytics Prizes” 2019.

[9] Santosh Ray and Mohammed Saeed “Applications of Educational Data Mining and Learning Analytics Tools in Handling Big Data in Higher Education” 2018.

[10] Andrés Villanueva Manjarres, Luis Gabriel Moreno Sandoval, Martha Janneth Salinas Suárez “Data mining techniques applied in educational environments: Literature review” Digital Education Review - Number 33, June 2018- http://greav.ub.edu/der 2018.

[11] Marcos Wander Rodrigues, Luiz Enrique Zárate, Seiji Isotani “Educational Data Mining: A review of evaluation process in e-learning” TELE 1119 2018.

[12] Ashish Dutt , Maizatul Akmar Ismail , And Tutut Herawan “A Systematic Review on Educational Data Mining” 2017.

[13] Abdulmohsen Algarni “Data Mining in Education” 2016.

[14] N D Lynn and A W R Emanuel “Using Data Mining Techniques to Predict Students' Performance. a Review” 2021.

[15] Wenchao Zhou1,Ting Yang “Application Analysis of Data Mining Technology in Ideological and Political Education Management” 2021.

[16] Xiaohua Guo “Research on the Application of Data Mining in the Analysis of College English Teaching Quality” 2021.

[17]Carlos A. Palacios, José A. Reyes-Suárez, Lorena A. Bearzotti, Víctor Leiva and Carolina Marchant “Knowledge Discovery for Higher Education Student Retention Based on Data Mining: Machine Learning Algorithms and Case Study in Chile” Volume 23 Accepted: 13 April 2021 / Published: 20 April 2021.

[18] Ammar J. M. Karkar, Hayder K. Fatlawi and Ahmed A. Al-Jobouri University of Kufa, Najaf, Iraq “Highlighting E-learning Adoption Challenges using data Analysis Techniques: University of Kufa as a Case Study” 2020.

[19] Aderibigbe IsraelAdekitana OdunayoSalau “The impact of engineering students' performance in the first three years on their graduation result using educational data mining” 2019.

[20] Buenaño-Fernandez, Diego , Villegas-CH, William and Luján-Mora, Sergio “The Use of Tools of Data Mining to Decision Making in Engineering Education - A Systematic Mapping Study” 2019.

[21] Alexandra A. De Souza, Thiago S. Barcelos, Roberto Munoz, (Member, Ieee), Rodolfo Villarroel, And Leandro A. Silva “Data Mining Framework to Analyze the Evolution of Computational Thinking Skills in Game Building Workshops” 2019.

[22] Pankhurhi Mallik , Chandrima Roy , Ekansh Maheshwari , Manjusha Pandey and Siddharth Rautray “Analyzing Student Performance Using Data Mining” 2019.

[23] Dedy Hartama, Agus Perdana Windarto, Anjar Wanto “The Application of Data Mining in Determining Patterns of Interest of High School Graduates” 339 (2019) 012042 IOP Publishing doi:10.1088/1742-6596/1339/1/012042 2018.

[24] Samina Kausar, Xu Huahu, Iftikhar Hussain, Zhu Wenhao, And Misha Zahid “Integration of Data Mining Clustering Approach in the Personalized E-Learning System” Received October 26, 2018, accepted November 13, 2018, date of publication November 20, 2018.

[25] Mukesh Kumar, A.J. Singh2 , Dr. Disha Handa “Literature Survey on Student’s Performance Prediction in Education using Data Mining Techniques”2017.

[26] Hilal Almarabeh “Analysis of Students' Performance by Using Different Data Mining Classifiers” DOI: 10.5815/ijmecs.2017.08.02 2017.

[27] Nastaran Peimani and Hesam Kamalipour “Online Education and the COVID-19 Outbreak: A

Case Study of Online Teaching during Lockdown" Volume 11 Issue 2 10.3390 2021.

[28] Pinaki Chakraborty, Prabhat Mitta, Manu Sheel Gupta ,Savita Yadav, Anshika Arora "Opinion of students on online education during the COVID-19 pandemic" Hum Behav & Emerg Tech. 2021;3:357–365 2021.

[29] Nastaran Peimani and Hesam Kamalipour "Online Education in the Post COVID-19 Era: Students' Perception and Learning Experience" Educ. Sci. 2021, 11, 633. https://doi.org/10.3390/educsci11100633 2021.

[30] Tinggui Chen, Lijuan Peng, Xiaohua Yin, Jingtao Rong , Jianjun Yang and Guodong Cong "Analysis of User Satisfaction with Online Education Platforms in China during the COVID-19 Pandemic "Received: 9 June 2020; Accepted: 2 July 2020; Published: 7 July 2020 .

[31] Olga Korableva, Thomas Durand, Olga Kalimullina "Studying User Satisfaction with the MOOC Platform Interfaces Using the Example of Coursera and Open Education Platforms" DOI: https://doi.org/10.1145/3322134.3322139 2019.

[32] Boqin Liu, Hanrong Chen, He Junmei "Research on the Platform of Online Education Platform Based on Cloud Computing" The 14th International Conference on Computer Science & Education (ICCSE 2019) August 19-21, 2019.

[33] Cindy Sing Bik Ngai , Wing Man Lee, Patrick Pak Kei Ng and Doreen Dongying Wu "Innovating an integrated approach to collaborative eLearning practices in higher education: the case study of a corporate communication e-platform" ISSN: 0307-5079 2019.

[34] Ayten A. Bulat S. İnceismail E "A Study of Generation Z Viewing Habits in Context of Uses and Gratification Theory: The Protector Netflix Series Case" DOI NO: 10.7456/ctc_2019_01 2019.

[35] Mufleh Salem M. Alqahtani, C. Vijaya Bhaskar, Kesavan Vadakalur Elumalai, May Abumelha "WhatsApp: An Online Platform for University-Level English Language Education" Volume 9. Number 4. December 2018.

[36] Ifeanyi Benedict Ohanu & Chiamaka Adaobi Chukwuone "Constraints to the use of online platform for teaching and learning technical education in developing countries" Received: 24 March 2018 /Accepted: 7 June 2018.

[37] Srikanth Kollurua, James T. Varughese "Structured academic discussions through an online education-specific platform to improve Pharm.D. Students learning outcomes" Volume 9, Issue 2, March–April 2017, Pages 230-236 2017.

[38] Colleen Ryan, Louise Young & Margaret McAllister "The impact of an online learning platform about nursing education on Enrolled Nurse preceptor teaching capabilities: A pre-post-test evaluation" ISSN: 1037-6178 (Print) 1839-3535 (Online) Journal homepage: 2017.

[39] Nastaran Peimani and Hesam Kamalipour "Online Education in the Post COVID-19 Era: Students' Perception and Learning Experience" Sci. 2021, 11, 633. 2021.

[40] Rosa M. Rodríguez-Izquierdo "Does service learning affect the development of intercultural sensitivity? A study comparing students' progress in two different methodologies" Volume 82, May 2021, Pages 99-108 2021.

[41] Josep Gustems-Carnicer , Caterina Calderon , Diego Calderon-Garrido, Carolina Martin-Piñol "Academic Progress, Coping Strategies and Psychological Distress among Teacher Education Students" y Vol. 9 No.3 October 2020 pp. 290-312 2020.

[42] Neil Harrison "Patterns of participation in higher education for care-experienced students in England: why has there not been more progress?" ISSN: 0307-5079 (Print) 1470-174X (Online) Journal homepage 2020.