STUDENT LEARNING ANALYSIS & RECOMMENDATION SYSTEM

Raghvendra Kumar Jha
M. Tech, Deptt.of Computer Science & Engineering, DPG Institute of Technology & Management
Maharishi Dayanand University
Rohtak, India
raghven11@gmail.com

Neeraj Sharma
M. Tech, Deptt. of Computer Science & Engineering, DPG Institute of Technology & Management
Maharishi Dayanand University
Rohtak, India
ce.neeraj@gmail.com

Richa Nehra
Faculty, Deptt. of Computer Science & Engineering, DPG Institute of Technology & Management
Maharishi Dayanand University
Rohtak, India
richa.dpgitm@gmail.com

Rashmi Verma
HOD, Deptt. of Computer Science & Engineering, DPG Institute of Technology & Management
Maharishi Dayanand University
Rohtak, India
rashmiverma.cse@dpgitm.com

***Abstract*—During global lock down period due COVID-19 pandemic one of the immensely affected area was education sector. On the one hand shutting down of schools & colleges for indefinite period severely affected the traditional education system whereas on the other hand it has given hope to new research areas in education system using machine learning (ML) algorithms. In this research paper we have proposed a new machine learning model for performance evaluation of student vis a vis teacher and making recommendations for the best suitable student teacher pair based on ML algorithms. It introduces the principles behind ML algorithms and illustrates how these algorithms could be applied by the system to evaluating student academic performance based on his academic and behavioral parameters vis a vis teachers performance based on the feedback system. The ML model proposed in this paper aims to adaptively adjust the training for each student based on his/her own pace of learning. The ML model will monitor the student's progress and can make decision about next step training. Several approaches using ML techniques have been proposed to provide a practical method for evaluating student / teacher performance and making final recommendations for student teacher collaboration.**

Keywords—Machine Learning Analysis, Online Education

I. INTRODUCTION

The applications of Machine Learning (ML) and Artificial Intelligence (AI) in marketing and medical science has been wide publicity. However, in the education field the application of the same is still in its nascent stage and under the purview of research only. The applications of Machine learning (ML) model for performance evaluation of Students and Teacher is in general new. However, post COVID-19 pandemic situation it has reached a wide range of application areas in educational systems in addition to evaluation of student academic performance, including the evaluation of curriculum and that of the educators such as lecturers and tutors. Previously, in student academic performance evaluation, the fuzzy logic techniques have been adapted for evaluation based on numerical scores obtained in an assessment of semester examination marks. The Fuzzy Logic techniques were proposed for determining the level of a student's understanding of a certain subject matter in the context of Intelligent Tutoring System, and a fuzzy approach was proposed to assess student performance based on several criteria with a strong suggestion that the method be applied to Computer Assisted Instruction.

Interesting work has been reported along this line of research. This includes evaluation of journal grades, evaluation of vocational education performance, collaborative assessment, and performance appraisal systems of academics in higher education. The focus of attention of this research work is development of a recommendation model for Students for their best suitable teacher based on performance evaluation system to analyze the performance of a student based on his academic records, his learning activities and considering his behavioral traits and that too of a teacher using feedback system. The outcome of this thesis will be a recommendation system which can also be used by institutions and online teaching platforms (commercial and non-for-profit social ventures such as Unacademy, Byju's & Khan Academy) for student teacher collaboration and making recommendations. The similar concept is also being used in the commercial platforms to recommend the best suitable and available products to a buyer using ML based marketing algorithms.

Evaluation of student academic performance usually consists of several components, each involving several judgments often based on imprecise data. This imprecision arises from human (teacher/tutor) interpretation of human (students) performance. Arithmetical and statistical methods have been used for aggregating information from these assessment components. These methods have been accepted by many educational institutions around the world although there are limitations with these traditional approaches.

The use of ML algorithms for the evaluation of teachers and students' performance is newly introduced in academic environment. However, it has reached a wide range of application areas in educational systems in addition to evaluate of students' academic performance, including the evaluation of curriculum and that of the educators.

II. OVERVIEW

A. Use of Big Data Analytics in Online Education

Big Data analysis differs from traditional data analysis primarily due to the volume, velocity and variety characteristics of the data being processed. To address the distinct requirements for performing analysis on Big Data, a step-by-step methodology is needed to organize the activities and tasks involved with acquiring, processing, analysing and repurposing data. The upcoming sections explore a specific data analytics life cycle that organizes and manages the tasks and activities associated with the analysis of Big Data. From a Big Data adoption and planning perspective, it is important that in addition to the lifecycle, consideration be made for issues of training, education, tooling and staffing of a data analytics team.

The Big Data analytics lifecycle can be divided into the following nine stages.

1. Business Case Evaluation
2. Data Identification
3. Data Acquisition & Filtering
4. Data Extraction
5. Data Validation & Cleansing

6. Data Aggregation & Representation

7. Data Analysis

8. Data Visualization

9. Utilization of Analysis Results

The Data Analysis stage as shown above is dedicated to carrying out the actual analysis task, which typically involves one or more types of analytics. This stage can be iterative in nature, especially if the data analysis is exploratory, in which case analysis is repeated until the appropriate pattern or correlation is uncovered. Depending on the type of analytic result required, this stage can be as simple as querying a dataset to compute an aggregation for comparison. Big data analytics is actively being used in many business organizations worldwide in business intelligence and in areas such as marketing and financial decisions.

Analytics refers to the techniques used to analyse and acquire intelligence from big data.

B. Problem Definition

Online education institutions are trying to gain more insight into student experiences to monitor and improve teaching and learning and assist with changes and improvements that will benefit student's success. The use of big data offers some excellent opportunities. The challenges faced by Online Education System are interrelated to each other and may be responsive to the big data capabilities such as:

(a) Real-time feedback and Recommendations

(b) Personalized Learning and Continuous Improvement

(c) Sentiment and Behaviour Analysis

(d) Improved Student Retention

Big data analytics can make instant alerts and provide feedbacks to teachers and students on academic performance by analysing underlying complex data patterns. This approach will help in predicting a dropout student, student who needs additional help or even a student who needs more challenging assignment.

With the help of big data, it is possible for the teacher to formulate formative assessments aptly challenging or demanding according to each student's talent and learning ability. This is possible through various methods such as setting up groups based on ability within the system and allocating a particular student to an appropriate group. Based on the outcome of the formative assessment, students can be advised on their next steps: more advanced learning, or different content or more practice on the same topic.

III. METHODOLOGY

Online academic platforms collect more and more data about their students, and as students' record databases have grown more complex and accessible – we are entering a new era of using data to improve student success, streamline processes, and more effectively utilize resources. Online Education record all academic related data from various activities such as student's data, registration data, assessment data etc. The major objective of learning analytics is concerned with beneficial information for data driven decision making. This can be achieved by the analysis of big data collected from various learning activities performed by students and teachers. However, the greatest challenge is to determine how data is captured, processed, stored, presented, and used for the benefits for tomorrow's outcome.

The recommendation model requires historic data of Students and Teachers learning activities based on which the different classification algorithm will be applied to find the best pair match of Student and Teacher. Therefore, this complete project activity can be carried out in two phases.

In the first phase of project a web-based software tool (LEARNOO) will be developed to record the data of student's day to day learning activities as well as Teacher's feedback. This will include the development and deployment of web application using latest technologies such as Java Spring Boot & Angular. Data collected from students learning activities and feedbacks collected in respect of teachers will be stored in a relational database. This dataset will be further divided in two parts (preferably 80/20 ratio) for training and testing purpose. The ML model trained on this dataset will be used for so called recommendation model.

In the second part we will integrate this ML model in our web application to make it a complete recommendation system which will not only guide the students in choosing their best learning paths but also recommend the best available teacher as per their learning goals.

A. Architecture of Learning Management System (LMS)

In Fig. 1, the architecture considered for the Learnoo project is presented. As observed, this architecture considers different technologies for the Back End and Front-End parts of the web-service. At the Back-End side, web application is programmed in Java Spring Boot with integration of Spring Security for consideration of API Security. At the Front-End part, Angular 10 has been used which is further integration of HTML5, TypeScript and CSS3 along with different NPM packages for specific requirements of the application. The request/response functionalities are implemented by means of the HTTP protocol with SSL encryption (HTTPS). As RDBMS, MySQL has been used as primary database.

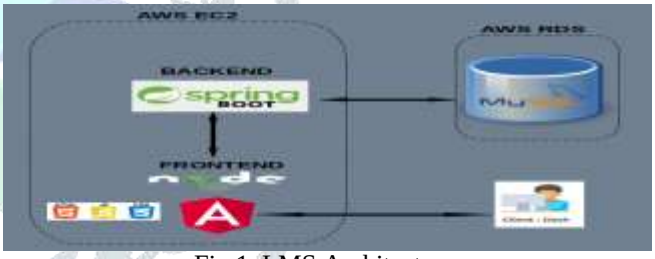

Fig 1. LMS Architecture

B. Architecture of Learning Recommendation System

As depicted below in Fig. 2 the dataset obtained from the Phase-I i.e., Learning Management System (LMS) must pass through ETL (Extract, Transform and Loading) process prior to be used for data analysis and final recommendation.

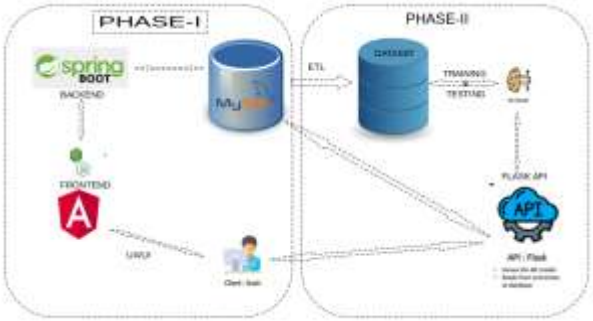

Fig 2. LRS Architecture

The students interact with the LMS to watch lectures, complete assignments, and read materials. Finally, student interactions are recorded and stored in database. The database contains student behavioural data, such as their interactions with the Learning system. An instructor can utilize these data to understand student behaviour and improve the students' performance based on the pattern

identified for the students' performance. This analysis will help the instructor analyse those activities to understand student behaviour.

The various components of recommendation model are as follows:-

i. Data Pre-processing Model

The data provided by the LMS cannot be directly used as inputs in the ML model. Initially, the input features of the current study sourced from different tables needed to be integrated into a single table. We performed various pre-processing steps on the data using python libraries as mentioned below to format the raw data into a form acceptable for ML algorithms.

- Pandas
- NumPy
- Matplotlib
- Plotly
- Seaborn
- sklearn
- dabl

ii. ML Model (Inference Model)

The ML techniques tested as predictive models in the current project are described as below:

(a) Logistic Regression - is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.). In other words, the logistic regression model predicts P(Y=1) as a function of X.

(b) Random Forest - Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of over fitting to their training set.

iii. Web API

Flask API - Flask is a micro web framework written in Python. It can create a REST API that allows us to send data, and receive a prediction as a response.

IV. RESULT S AND ANALYSIS

In the current study, we used the dummy dataset ('***StudentsPerformance.csv***') of quite good size to demonstrate the student's performance analysis and train/test our inference model. The dataset contains various parameters such as Gender, Age, Ethnicity, Parent's level of Education, Test preparation course, Math Score, Reading Score, Writing Score etc.

	gender	racelethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	72	72	74
1	female	group C	some college	standard	completed	69	90	88
2	female	group B	master's degree	standard	none	90	95	93
3	male	group A	associate's degree	free/reduced	none	47	57	44
4	male	group C	some college	standard	none	76	78	75

Using the above-mentioned dataset, we performed a data analysis using Python's rich library of Data science such as Pandas, NumPy, Seaborn and Matplotlib. The report of complete analysis performed over the chosen dataset is attached as **Annexure** to this document. In this analysis we used various Machine Learning algorithm to analyse the data and Train & Test our model before final deployment which are mentioned next.

A. Logistic Regression

```
from sklearn.linear_model import  LogisticRegression

# creating a model
model = LogisticRegression()

# feeding the training data to the model
model.fit(x_train, y_train)

# predicting the test set results
y_pred = model.predict(x_test)

# calculating the classification accuracies
print("Training Accuracy :", model.score(x_train, y_train))
print("Testing Accuracy :", model.score(x_test, y_test))
```

```
Training Accuracy : 0.3586666666666667
Testing Accuracy : 0.304
```

B. Random Forest

```
from sklearn.ensemble import RandomForestClassifier

# creating a model
model = RandomForestClassifier()

# feeding the training data to the model
model.fit(x_train, y_train)

# predicting the x-test results
y_pred = model.predict(x_test)

# calculating the accuracies
print("Training Accuracy :", model.score(x_train, y_train))
print("Testing Accuracy :", model.score(x_test, y_test))
```

```
Training Accuracy : 1.0
Testing Accuracy : 0.796
```

As mentioned above this complete project was divided in two phases. Both the phases together complete the project goal, and both the phases have equal importance. In the process of development of predictive model, we experienced many technological hindrances and gained deeper insights into implementation of model. At many instances we had to change our decisions in selecting the resources due to their limitations and compatibility issues. The complete trial and testing of product was carried out after completion of Phase-II.

Machine learning is having a tremendous impact on the teaching industry. Teaching industry is adopting new technologies to predict the future of education system. It is Machine learning which predict the future nature of education environment by adapting new advanced intelligent technologies. This work explores the application of Machine Learning in teaching and learning for further improvement in the learning environment in online education. We explored the application of machine learning in customized teaching and learning environment and explore further directions for research. Customized teaching and learning consider student background, individual student aptitude, learning speed and response of each student. This customized teaching and learning approach provide feedback to teacher after real time processing of the data. This way a teacher can easily recognize student attention and take corrective measures. This will improve student participation and hence the overall results. Individual student concepts and goals can easily be track with the help of Machine learning by taking real time feedback. Based on that feedback, curriculum, topics, and methodology can be improved further. In simple terms, machine learning makes the process automatic for decision making process and analysed the individual student data. Overall, the assessment process is made more streamlined, accurate and unbiased with the help of machine learning. In the near future, machine learning will be more efficient and produce even better results.

V. CONCLUSION

It is found that data and analytics can help Institutions to better understand themselves, up-to- date and accurate knowledge of their own institutions. It is essential to shape the successful university of the future. There is potential for learning analytics to be used to help institutions to support their students through their educational journey. All online institutions should consider introducing an appropriate learning analytics system to improve student support and performance. Future prediction is difficult, since anything can happen led by leading companies with heavy financial backup they are investing heavily on Big Data, so it is clear that "Big data" is the extremely near future. Big Data Analytics is a trend that will increase substantially in the coming period and will have a large impact on any education institution due to the many advantages. Traditional face-to-face instruction can support traditional data- driven decision-making processes, Institutions should identify appropriate big data analytics tool to harvest the benefits from the huge amount of data and how Big Data Analytics tools can be used in a way to take decisions and drive the institution towards benefitting from the data.

VI. ACKNOWLEDGEMENT

The authors wish to acknowledge the unanimous reviewers for their kind suggestions and comments for improving this paper.

VII. REFERENCES

[1] Data Mining: Practical Machine Learning Tools and Techniques, Second Edition Ian H. Witten and Eibe Frank 5.

[2] Minimol Anil Job, An efficient way of Applying big data Analytics in Higher Education sector for Performance Evaluation, International Journal of Computer Applications, Vol 180-No 23,February 2018

[3] Amal S.Alblawi & Ahmad A. Alhamed , Big Data and Learning Analytics in Higher Education, 2017 IEEE Conference on Big Data and Analytics (ICBDA).

[4] Vatsala, Rutuja Jadhav, Sathyaraj R,A Review of Big Data Analytics in sector of Higher Education, International Journal of Engineering Research and Application, Vol 7(Issue 6)June 2017.

[5] Julius Murumba & Elyjoy Micheni, Big Data Analytics in Higher Education: A Review, The International Journal of Engineering and Science (IJES) || Volume || 6 || Issue || 6 || Pages || PP 14-21 || 2017 ||

[6] Yuqian Li,Peng Li,Feng Zhu,Ruchuan Wang: Design of Higher Education Quality Monitoring and Evaluation Platform Based on Big Data. 12th International Conference on Computer Science & Education (ICCSE 2017) August 22-25, 2017.

[7] Fezile Matsebula, Ernest Mnkandla: A big data Architecture for Learning Analytics in Higher Education, IEEE Africon 2017 Proceedings.

[8] Venkatesh Naganathan, Comparative Analysis of Big Data ,Big Data Analytics : Challenges and Trends, International Research Journal of Engineering and Technology(IRJET),May 2018.

[9] M.M.M.A Riffai, David Edgar, Peter Duncan, Ahmed Hassan Al-Bulushi: The Potential for Big Data to Enhance the Higher education Sector in Oman, 2016 3rd MEC International Conference on Big Data and Smart City.

[10] Thomas Erl, Wajid Khattak, Paul Behler: Big Data Fundamentals ,Concept Drivers and Techniques-Text Book.

[11] Matsebula, Fezile. A Big Data Architecture for learning analytics in higher education. 978-1-5386-2775-4/17/$31.00 ©2017 IEEE.

[12] Data Mining: Practical Machine Learning Tools and Techniques, Second Edition Ian H. Witten and Eibe Frank International Journal of Computational Intelligence and Informatics, Vol. 6: No. 4, March 2017.

[13] Learning Spring Boot 2.0 -: Simplify the development of lightning-fast applications based on microservices and reactive programming by Greg L. Turnquist (Author)

[14] U. Spagnolini, L. Fontana, A. Paganoni, A. Torrebruno,, M.A. Prada , M. Domínguez, A. Morán, R. Vilanova, J. Lopez Vicario, M.J. Varanda, P. Alves, M. Podpora and M. Barbu. Io5 - data mining tool for academic data exploitation. Technical report, ERASMUS + KA2 / KA203 SPEET Project, March 2019.