# EDUCATIONAL DATA MINING TO SUPPORT PROGRAMMING LEARNING

**[1]Dr. Bhosale R. S, [2]Mr. Deshmukh Suraj, [3]Mr. Malunjkar Avishkar, [4]Ms. Netke Rutuja, [5]Ms. Patki Rutuja**

Department of Information Technology,
Amrutvahini College of Engineering
Sangamner, Maharashtra, India.

***Abstract*- The majority of academic programs in the fields of engineering aim to enhance practical abilities. Due to its numerous uses & global significance, computer programming has today become a necessary talent. We use academic scores and students' programming logs as an experienced dataset. Kaggle.com Learning is where the programming logs are gathered. The method, datasets, evaluation, and output of machine learning approaches were discovered to be necessary. An analyst can find trends by using clustering and visualization in a low-dimensional representation of student data. Early performance data and exploratory characteristics are utilized to classify students who have previously graduated from three performance levels. The results reveal that technology-enhanced commenter-contributes significantly to learning process satisfaction and problem-solving learning outcomes. The proposed educational data mining techniques allow access to student's behaviour in the E-learning system for understanding students' interest in studying the learning materials.**

***Key Words*: Machine Learning, Authentication, Security**

**Published in IJIRMPS (E-ISSN: 2349-7300), Volume 11, Issue 3, May-June 2023**



## INTRODUCTION

E-education is currently experiencing rapid development, and whatever educational activity a student participates in is tracked and recorded in multiple databases, log files, personal profiles, etc. Large amounts of data are continually generated by e-learning to illustrate the constant interplay between training and instruction. To address educational difficulties, data mining for education applies machine-learning data mining methodologies & techniques to many different kinds of educational data sets. As is clear from the name it offers computers that makes similar to people, and educational data mining to promote programming learning being able to learn. Teacher, student, and environment are the three main sorts of sources in an educational institution. These three actors must interact with one another in order to produce some outcome. These three actors interact across the system that has been successfully constructed. Following the interaction between these three actors, the students are divided into groups based on their results using the clustering algorithm k-means, which generates the output and classifies the students according to their academic achievements.

We suggest a framework for unsupervised machine learning techniques used in EDM and data analysis. On a real-world problem-solving dataset, tests are run to verify the effectiveness of the suggested framework. With the goal to aid programming learning, different features, patterns, and rules are retrieved from problem-solving data. On the basis of the extracted features, rules, and data patterns, helpful recommendations are made for students and educators. The knowledge that was extracted (features, rules, and patterns) reveals the shortcomings and the range of potential improvements in programming learning. E-learning platforms can incorporate the proposed EDM architecture.

As a result of a number of factors and demands, such as teacher shortages, unbalanced student-teacher ratios, infrastructure and logistical limitations, the high cost of technical and professional courses, the distribution of education to a large audience, time savings, and easy access to many courses, e-learning platforms have grown in popularity in recent years. A variety of data kinds are constantly created as the use of e-learning systems rises. While some data are unstructured, others are structured. As a result, using conventional statistical algorithms, it is highly challenging to extract usable information from this vast volume of mixed data archives. For the purpose to assess students' application of an e-learning system and determine their interests, an EDM approach like k-means clustering was utilized. The relationship between e-learning system activity and academic success was also noted. Using log data, the programming level of the students was assessed.
Additionally, one of the crucial phases for EDM to get better outcomes is the preparation of data. The value of clustering approaches in the overall framework of EDM was shown by Hooshyar et al. They suggested an evaluation strategy that automatically suggests an appropriate algorithm for EDM after comparing the effectiveness of clustering algorithms on a particular dataset. The suggested framework for EDM, in contrast to previous models, uses data from an OJ system that was acquired using real-world problem-solving techniques to identify data patterns, rules, and characteristics. The identified attributes, guidelines, and data patterns are then used to provide some helpful recommendations for students and teachers.

## LITURATURE SURVEY

• **Recommender System for E-Learning based on Personal Learning Style, Nunung Nurul Qomariyah ,2020** We propose an implementation design of an E-Learning recommender system based on a logic approach, APARELL (Active Pairwise Relation Learner), which has been implemented for used car sales domain. There is an opportunity to apply the same procedure for e-learning system to help the student to choose the best material according to their preferences. We also propose an ontology of material content based on the different learning styles. In this paper, we show that there is a big potential to implement a personalized recommender system in e-learning based on the students learning style [1].

• **A syntactic and semantic multi-agent based question answering system for collaborative e-learning, Abderrazzak Samadi; El Fazazi Hanaa, 2018** The task can take up a considerable amount of the time they spend each day on teaching if the number of students is high and the tutors have a narrow time frame in which to reply to them all. In this paper, we present a Question Answering System (QAS) that helps learners to find the best answers to their questions and helps tutors to answer questions asked by their students in an e-Learning environment. We present a syntactic and semantic multi-agent approach to question answering in e-learning platforms. We try to improve existing approaches by using multiple techniques in a multi-layer system where learners and tutors with the help of automatic agents collaborate to find the best answer to the submitted question[2]. 8 Educational data mining to support programming learning

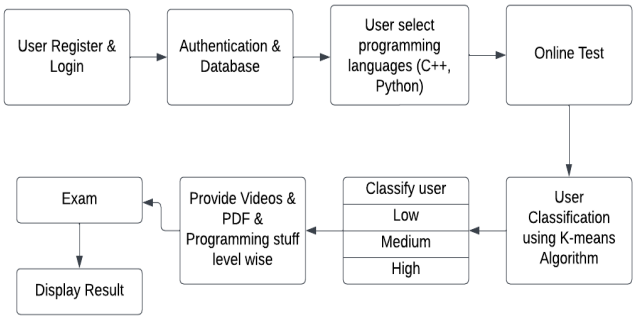
• **A survey on service-oriented architecture for E-learning system, Rani S Jamuna, 2018** E-learning refers to learning that is delivered or enabled via electronic technology. E-learning is classified as synchronous or asynchronous. Service based technology in e-learning provides a common infrastructure to integrate heterogeneous software components, thus enhancing interoperability between different components. Service-oriented architecture for e-learning provides adaptable, interactive, extensible, distributed, collaborative and intelligent e-learning system to effectively realize the learning anytime and anywhere to instructors and learners. This paper presents a detailed analysis of wellknown architectures for service based e-learning system. The architectures that surveyed here gives users the ability to collect, analyze, distribute and use elearning knowledge from multiple knowledge sources[3].

• **E-Learning Supporting System (ELS) in Nahda University in Upper Egypt: Case Study, Samia El-Azab, 2015** Nahda University is the first privet university in Upper Egypt in Banisuif. As the direction for Nahda University is to be paperless, there is the E Learning Support System (ELS) for the undergraduate students of the six faculties of the university; which includes Oral and Dental Medicine, Pharmacy, Business Administration, Engineering, Mass Communication and information Technology. The ELS is considered the communication tool between students and the academic staff members and this is available anywhere and anytime across the internet in different ways to be suitable with variable situations and student abilities. ELS is considered Learning Management System LMS and e learning and it is not a substitute for the traditional learning but it is a blended learning. Generally, ELS is the bridge connecting between students and

university[4]. 9 Amrutvahini COE Sangamner, Department of Information Technology - 2022 Educational data mining to support programming learning

• **Deep E-School-Nurse for Personalized Health-Centered E-Learning Administration, Tannaz Karimi; Babak Majidi, 2019** As a case study, the deep neural networks are used for automatic profiling of students with diabetes. Based on the student diabetes profile a personalized curriculum is designed for the students which includes physical activity and a healthy diet at appropriate intervals during the study. The proposed e-school-nurse can help students to have a healthier e-learning experience[5].

## METHODOLOGY




## AIM & OBJECTIVES

1. To improve the challenges with programming learning.
2. To improve their literacy skills in a fun and engaging way without interfering with their everyday activities.
3. In order to assist for rural India's lack of skilled workers and suitable infrastructure.
4. To evaluate the user's level of programming language study by evaluating their level of expertise.

## MOTIVATION

The development of data and communication in reality has attracted a lot of curiosity about PC programming. One of the primary difficulties is meeting the increasing requirement for extremely knowledgeable software engineers in the ICT industry.

## SYSTEM ARCHITECTURE

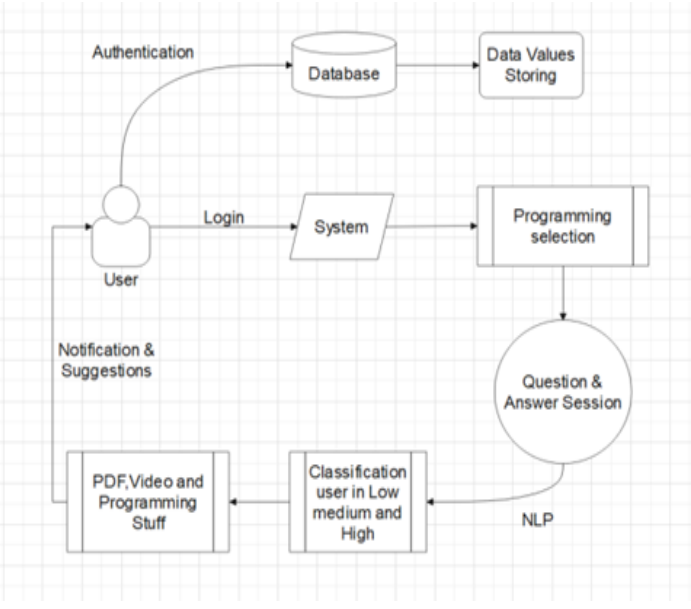



**Fig -1**: System Architecture Diagram

**APPLICATION:**

- School
- Collage
- Classes
- Government sector

**PROPOSED SYSTEM**

This section describes the suggested EDM strategy, which appears in Figure. The proposed method can be broken down into three main phases: first, data preparation and collection; second, data clustering, numerical analysis, and feature extraction; third, pattern and rule mining; and fourth, pattern mining algorithms applied for every cluster. Each phase of the suggested solution uses a distinct algorithm, which isData clustering is one of the crucial stages in our suggested methodology. Data clustering is strongly tied to data analysis, including data patterns, feature extraction, and association rules, and inappropriate clustering might produce disappointing results.

We investigated the applicability of several clustering approaches and algorithms for our clustering aims prior to clustering our multidimensional education data. Finding the ideal clustering method is highly challenging because each approach has a different level of relevance for a certain application. In contrast to partitional clustering approaches, the hierarchical clustering method was examined. This method has high computational and memory requirements. A hierarchical approach may be quite expensive for huge data sets. discussed below. Evaluate time, assumptions, data sets, and clustering goals/results are the primary distinctions between hierarchical and partitional approaches.

By integrating basic mathematical techniques, the unsupervised clustering algorithm is utilised to arrange the vast dataset's most related data into smaller groups. A quick and effective approach for partitioning data is the K-means clustering algorithm. Data science has difficulties as a result of the expanding gathered of structured, unstructured, and multidimensional data across a range of media. Multidimensional data clustering is a complicated and difficult execution, nevertheless.

**System Components:**

The proposed system aims to leverage educational data mining techniques to enhance programming learning experiences. By analyzing student data, such as code submissions, assessment results, and learning behaviors, the system can provide personalized feedback, identify areas of improvement, and optimize the learning process. The system consists of several key components, including data collection, data preprocessing, analysis.

1.Data Collection:

- Learning Management System: This component manages the learning process, tracks student activities, and collects relevant data, including code submissions, quiz results, and user interactions.
- Learning Data Repository: Stores the collected data for further analysis and processing.

2. Data Pre-processing:

- This component prepares the collected data for analysis by cleaning, transforming, and integrating different data sources.
- Data quality checks and filtering may be performed to ensure the accuracy and relevance of the data.

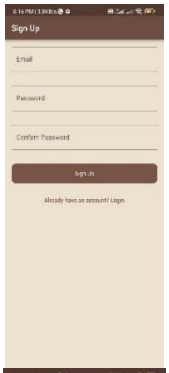
3. Educational Data Mining:

- Utilizes data mining techniques and algorithms to extract valuable insights from the preprocessed data.
- Tasks may include clustering student profiles, identifying learning patterns, and detecting areas of difficulty.

4. Analysis & Modelling:

- Analyses the mined data to generate meaningful models and representations of student performance, learning styles, and programming concepts.
- Statistical analysis, machine learning, and other data analysis methods may be employed to identify trends, correlations, and predictive models.

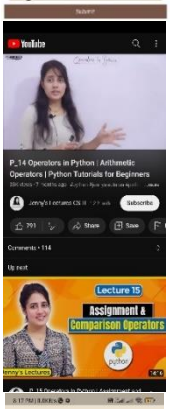
## RESULT


Sign Up
Email
Password
Confirm Password
Already have an account? Login


Login
Email
Password
python
c++
Create an account


1. Who developed Python Programming Language?


2. Which type of Programming does Python support?
object-oriented programming


P_14 Operators in Python | Arithmetic Operators | Python Tutorials for Beginners
Lecture 15
Assignment 4
Comparison Operators


Failed


## CONCLUSION

We have developed an interactive educational application designed for mobile devices, aimed at assisting users in learning multiple programming languages. This application offers users the flexibility to choose the programming language they wish to learn and provides a comprehensive curriculum ranging from basic to advanced concepts. Additionally, our system can be extended to other practical and exercise-based courses, allowing users to explore data patterns, statistical features, and rules. Through our research and development process, we have successfully applied data mining methods to rank student activities based on their impact on final exam performance. This approach enables us to identify key factors that influence students' success and provides valuable insights for both learners and educators. By incorporating this feature into our application, we aim to enhance the learning experience and help students achieve better academic outcomes. The project represents a culmination of extensive studies conducted on various e-learning systems. Our proposed framework stands out as a highly intelligent and comprehensive solution, providing learners with a complete and immersive learning environment.

## REFERENCES

[1] Conference: Sixth International Conference on e-Learning (econf)" E-Learning System Model For University Education Using UML", Nunung Nurul Qomariyah, December 2020

[2] 2018 4th International Conference on Optimization and Applications (ICOA), "A syntactic and semantic multi-agent based question answering system for collaborative e-learning", Abderrazzak Samadi; El Fazazi Hanaa; 2018

[3] 2009 International Conference on Intelligent Agent Multi-Agent Systems, "A survey on service-oriented architecture for E-learning system", Rani S Jamuna; Marie Stanislas Ashok, 2009

[4] Conference: 2015 Fifth International Conference on e-Learning (econf)" ELearning Supporting System (ELS) in Nahda University in Upper Egypt: Case Study", Samia Mostafa Elazab, 2015

[5] 2019 13th Iranian and 7th National Conference on e-Learning and e-Teaching (ICeLeT) Deep E-School-Nurse for Personalized Health-Centered E-Learning Administration:", Tannaz Karimi; Babak Majidi; Mohammad Taghi Manzuri, 2019 [6] J. Qi, P. Yang, L. Newcombe, X. Peng, Y. Yang and Z. Zhao, an overview of data fusion techniques for Internet of Things enab led physical activity recognition and measure", Information Fusion, vol. 55, pp. 269-280, 2020.

[7] P. Yang, J. Liu, J. Qi, Y. Yang, X. Wang and Z.Lv Comparison and Modelling of Country-level Microblog User and Activity in Cyber-physical-social Systems Using Weibo and Twitter Data", ACM Transactions on Intelligent Systems and Technology (TIST), vol. 10, no. 6, pp. 1-24, 2019 40

[8] J. Qi, P. Yang, A. Waraich, Z. Deng, Y. Zhao and Y. Yang," Examining sensor based physical activity recognition and monitoring g for healthcare using Internet of Things: A systematic review", Journal al of biomedical informatics, vol. 87, pp. 138-153, 2018.