**EMITTER International Journal of Engineering Technology**
Vol. 12, No. 2, December 2024, pp. 213~236
DOI : 10.24003/emitter.v12i2.855

# A Deep Dive into a Groundbreaking Approach to Machine Learning-Powered E-Learning

**Subhabrata Sengupta[1], Rupayan Das[1], Satyajit Chakrabarti[2]**

[1]Information Technology, Institute of Engineering & Management, University of Engineering & Management, Kolkata, West Bengal, India.
[2]Computer Science & Engineering, Institute of Engineering & Management, University of Engineering & Management, Kolkata, West Bengal, India.
Corresponding Author: subhabrata.sengupta@iem.edu.in

*Received April 18, 2024; Revised May 29, 2024; Accepted December 1, 2024*

**Abstract**

Information retrieval aims to find the most important data for specific queries. The challenge is retrieving relevant data efficiently due to the large search area. Existing solutions lead to unnecessary processing costs. Additionally, identifying the main focus of the query is crucial for targeted retrieval. Current methods struggle to address these issues effectively. To overcome these challenges, we have proposed a goal-question-indicator (GQI) approach for personalized learning inquiry (PLA). This approach allows for efficient retrieval of variable-sized data with reduced processing requirements. We have also presented the open learning analytics platform's (Open-LAP) pointer motor segment, which helps end users specify goals, generates discussion topics, and provides self-characterizing pointers.

**Keywords**: ICT, Information Retrieval, Knowledge Base, Learner Model, Learning Analytics, Open Learning Analytics, Open-LAP, TF-IDF

## 1. INTRODUCTION

Up until now, all pertinent research has been done in the fields of predicted content and textual similarity. We will make use of both of these work developments while constructing our framework, which will respond to a query delivered as an information message. After comparing each question's semantic value to a related label, the closest question will be assigned the highest weight. Labels will be considered first. This innovative initiative will be focused on e-learning computerization. Learning analytics, or LA, is a fast-growing field of research that manipulates data to regulate processes and find ways to manipulate education to facilitate learning. Recently, a new area of study known as OLA has emerged. OLA focuses on learning scenarios that are more self-organized, self-sorted, and open. Information laborers, such as physicians or investigators dealing with money, sometimes have to acquire explicit inquiry languages (e.g., SQL and SPARQL1) in order to get information from an information base (KB). These days, the main focus of informative data mining [1-6] is creating new instruments for drawing ideas out of data. However, users may find it difficult to keep up with the most recent advancements in these query languages that allow them to access the data they require for their jobs because different Knowledge Bases (KBs) (such as Relational Databases, Triple Stores, No-SQL stores, etc.) are adopting new query languages so quickly. This makes it impossible for someone without a lot of PC experience to apply the knowledge in the KB effectively. Non-specialized clients will have easier access to the knowledge base's contents if specific language interfaces are made simple to use. The technology we have provided in this research is a Natural Language Interface that lets users ask natural language questions about the underlying knowledge bases. In this study, a Natural Language Interface system has been proposed that enables users to ask questions about the underlying knowledge bases in natural language. Rather than expecting the users to complete the whole query on their own, our framework will offer ideas to assist the users in completing their questions. Ultimately, our framework produces dynamic analysis for the result sets to aid clients in comprehending the subject and the particular concerns. Our framework will be parsed to its First Order Logic (FOL) depiction given a comprehensive inquiry employing a phrase structure derived from connected datasets.

Previous studies [7-11] have focused on personalized learning analytics, but there is a need for a more flexible and dynamic approach to address diverse stakeholders' concerns and goals within the learning analytics practice. While deep learning techniques [12-13] have been explored for knowledge extraction from classroom-generated data, there is a gap in effectively applying these techniques to create comprehensive learning assessment frameworks. There is a distinction between learning analytics and educational data mining, with the former focusing on larger scales such as courses and institutions. The gap lies in integrating tools and procedures to anticipate and act upon educational outcomes. Current learning analytics [14] implementations rely on predefined questions and indicators, which may not be suitable for open learning analytics due to its unpredictable nature. There is a need for new approaches that can adapt to the variable nature of learning analytics.

The key contributions of the work are fourfold:

- The work introduces a GQI approach for PLA, which efficiently retrieves variable size data with less CPU usage.
- Discusses the application, structure, usage, and assessment nuances of the Open-LAP, which aids end-users in defining goals and providing self-characterizing pointers.

- Proposes a NLP-based Interface system that allows users to ask questions about underlying knowledge bases in natural language, assisting users in completing their queries.
- Introduce a framework that provides dynamic analysis for result sets to help clients understand the subject and specific concerns, enhancing comprehension of the learning material.

Reminder part of the paper organized as follows. Section 2 presents related works. The originality of the paper is addressed in Section 3. Section 4 deals with problem statement and system model. Experiment and analysis are discussed in Section 5. Finally, the paper is concluded in Section 6.

## 2. RELATED WORKS

Customized LA testing was conducted, and the reasonable, structural, and usage nuances of a standard-based marker definition apparatus were presented in order to aid in the flexible definition and dynamic age of pointers to address the concerns of diverse partners with varying goals and questions in the LA practice [7]. The advantages conditions, potential uses, and challenges of applying deep learning techniques to extract knowledge from data generated by IoE devices in the classroom have been explored to create such learning assessment frameworks. Finally, a subtle but insightful connection is shown between the suggested Learning Analytics (LA) based methodologies and traditional instructing learning technique in terms of execution criteria such as student achievement, maintenance, awareness, and consideration [8]. These days, the focus of informative data mining is creating new instruments for drawing ideas out of data. These examples often have to do with the smaller-scale ideas that are learned, such copying a single digit or making a logical inference, among other things. The focus of learning examination, as it now varies from information mining, is on using tools and procedures at bigger sizes, including in courses, schools, and post-secondary institutions. However, the two controls work on the basis of expectations and examples: If we can follow the example provided in the material and comprehend what's occurring, we may anticipate what should happen immediately and take the necessary action. In many fields of study, including those where students work together with learning frameworks that leave helpful traces, artificial intelligence (AI) is concerned with a wide range of calculations that improve their display with experience [9]. Instructional data mining is the study of obtaining meaningful information from large-scale informational indexes or databases that contain understudies' communications during their learning process—for instance, in a virtual environment. In summary, learning analytics comprises a set of strategies aimed at comprehending and enhancing both the learning process and its surroundings. There are many processes involved in its construction, the first of which is identified carefully utilizing Educational Data Mining to extract information using specific AI algorithms. A useful overview of the taking in forms from a range of perspectives and goals is intended to be provided by the research on the convergences and connections between these three study topics in this work. Many models are presented and discussed. In view of representative relapses, an algorithm that utilizes non-exploratory data on prior test outcomes acquired by the institution as information has been tried [10]. It may forecast a 60–70% variance in understudy exam results. Clickers were found to be a more persuasive method of instruction than traditional written assignments done by hand, according to the computation used to assess the impact of teaching strategies in a traditional classroom with differential conditions. However, clickers were viewed as significantly less feasible than online assignments with timely feedback. The main goal of e-LAT's advancement is to process massive informational collections in microseconds regarding unique information examination premiums of educators and information protection issues [11]. This will force them to reflect on their innovative, improved teaching and learning environments and identify opportunities for improvements and mediations. Innovative methods that can produce distinct data on the outcomes of students participating in venture-based, community-oriented learning activities can be found in Multimodal Learning Investigation (MMLA) [12]. Using the multimodal learning research stage, we gathered several data surges and controlled and deleted multimodal linkages to address the linked question: which MMLA highlights are suitable indicators of cooperative critical thinking in venture-based learning open-ended tasks? Handwritten CPS ratings were relapsed using AI algorithms. The response to the query provides potential ways in which venture-based learners may be able to recognize specific parts of a team effort with intuition. The majority of current LA implementations rely on a predetermined set of questions and cues, which is unreasonable for OLA since the markers are unpredictable [13]. In this paper, we introduce the goal-question-pointer (GQI) approach for PLA and provide the theoretical underpinnings, practical application, and evaluation nuances of the Open-LAP's marker motor segment. This segment engages end-clients in the pointer age process by helping them define their goals, offering conversation starters, and providing self-characterizing pointers. It is possible to arrange multimodal learning examination data using both new and promising approaches—such as neural systems—and increasingly traditional relapse approaches [14]. The pros and cons of each approach vary depending on the settings and research questions. The work presented here is a crucial commitment to developing strategies that will subsequently identify the crucial components of students' success in project-based learning scenarios and ultimately help teachers provide appropriate and timely assistance to students in these primary areas. Previous frameworks for typical language queries over intricately linked datasets need the user to submit a comprehensive and highly structured query and provide the relevant answers as rudimentary element arrangements [15]. With the use of a component-based language that has a complete conventional semantics, we have developed a framework that can support rich autosuggest and provide gradually generated research for every result it gives. Research subjects by [16] generate digital traces that may be captured and analyzed when they engage with the many learning enhancements within their curriculum. These algorithms offer notable suggestions that can support understudy learning by adhering to the structure of the new type of information that is most frequently used in learning studies. This study looks at how such an analysis may be used to overcome the problems that keep teachers from providing students with individualized feedback on a large scale. The contextual inquiry included in the paper illustrated the relationship between the approach and understudies' favorable opinions of input quality and academic achievement. First-year students creating understudies tried their hand at PC frameworks as the research leaders [17]. The work by [18] introduces a comprehensive exploration of Interactive Information Retrieval (IIR), focusing on modeling and optimizing the interactive search process. It covers formal models like cooperative game theory, decision-theoretic approaches, feedback techniques, and result diversification, concluding with insights on evaluation strategies and future research avenues in IIR. The article [19] outlines the significance of archives and the importance of efficient information retrieval. It highlights the utilization of methods like Term Frequency-Inverse Document Frequency and Vector Space Model, coupled with stemming algorithms like Nazief-Adriani, to enhance retrieval performance by assessing document relevance through similarity calculations and subsequent sorting. The article [20] introduces a novel approach to adaptive e-learning systems, integrating learner motivation levels into the customization of learning paths and content selection. By combining Content-Based Filtering with machine learning algorithms, it aims to optimize learning experiences, with experimental results showcasing enhanced learning outcomes when considering learners' motivation levels. The paper [21] introduces a Reinforcement Learning (RL) based e-learning framework utilizing Markov Decision Process (MDP) to personalize learning paths, overcoming the challenge of selecting appropriate courses in the era of Massive Open Online Courses (MOOCs). By employing RL techniques like Q-learning for Sequential Path Recommendation (SPR), it adapts recommendations based on learners' feedback, enhancing learning experiences and achieving significant performance improvements, as evidenced by experimental results. The work by [22] presents a proof-of-concept blockchain based e-learning platform aimed at enhancing security and transparency in assessments while enabling personalized curriculum in higher education. By automating assessments and credential issuance, the platform increases trust in online education providers and improves transparency in education history and credentials, as indicated by the evaluation results. The study by [23] introduces Gamification into a Learning Management System (LMS) to enhance student motivation and cognitive abilities, aiming to create a Smart Learning Environment. Through the ADDIE method, gamified elements are integrated to increase engagement and cognitive improvement, as evidenced by higher N-Gain scores compared to a non-LMS group.

## 3. ORIGINALITY

A particular usage of IR is web search [18][24], which focuses on locating pertinent data from a vast collection of text documents—web pages in this case—that are available online. Web search is infused with traditional information retrieval (IR) concepts and techniques to help users quickly and effectively locate the information they need. Finding the web pages from the immense quantity of information on the World Wide Web that are most relevant to the user's query is the primary goal of online search. Web search engines create an index of web pages via the use of crawling and indexing techniques. This enables the speedy retrieval and ranking of pertinent sites in response to user queries. The mechanism of an online search engine evaluates a user's query and ranks the web sites according to how relevant they are to the query after the user inputs the query. Web pages are ranked and their relevance assessed using a variety of information retrieval (IR) approaches, including relevance ranking algorithms like PageRank, keyword matching [25], and term frequency-inverse document frequency (TF-IDF) weighting [15][24]. Beyond standard IR, web search has developed to include new features and methods [25] [26], such as entity recognition, natural language processing, tailored search results, and semantic analysis [27]. By improving the relevance and accuracy of search results, these advancements are intended to provide users with an improved search experience. All things considered, web search expands upon the ideas and methods of information retrieval, using web pages as the documents that are part of the World Wide Web. By obtaining and displaying pertinent web pages based on user queries, the aim is to empower users to efficiently browse and access the immense amount of information available on the internet. It makes sense to believe that Web search is the most important use of IR. For IR, web searches were also quite beneficial. Indeed, IR has gained significant attention due to search engines' immense popularity. But search is more than simply a simple application of traditional IR models. A plethora of information is available on the modern World Wide Web in a manner intended for human vision and understanding. Links may be followed, search engines can be used to get information, and domain names can be used to attempt to access websites. The content of the web pages that are acquired is simply understood by the program as a collection of random characters. Software developers are not able to understand the contents of a file, webpage, or document just by opening it. Even while the software may make assumptions based on HTML (Hyper Text Markup Language) or XML (Extensible Markup Language) tags, a programmer would still need to become involved and determine the semantics, or meaning, of each tag.

## 4. SYSTEM DESIGN

This section is divided into two parts, such as problem statement and workflow Design of the System:

### *4.1 Problem Statement*

Technology could seem simple. Indeed, it is when it can be executed flawlessly. However, there are some obstacles connected to the technological dissemination of information. The internet serves as the foundation for this kind of research project, but there are a number of issues that might come up. Online learning activities including group collaboration using tools like social networking sites, blogs, wikis, online games, online video sharing, and immersive virtual worlds. Collaboration refers to any active activity in which two or more partners work together to attain goals that they are unable to accomplish separately. Since ICT was introduced, collaboration has been described as utilizing its capabilities to support cross-organizational cooperation. When combined with peer and self-evaluation, collaborative e-learning may enhance comprehension, foster cooperation abilities, and highlight the procedures necessary for productive group projects. Conversely, "collaboration learning" is an approach where students at all performance levels cooperate in small groups to achieve a shared objective. Participants assist and collaborate with one another while also sharing with others what they have learned about the subject. Higher emotional and cognitive skill levels are used in linked and focused processes by them. Another way to think about collaborative learning is as an ongoing process where team members build new knowledge together. Together with it, the experience of working together produced a collective output. The following crucial elements will be included in an intelligent e-learning process:

- Core Content Knowledge
- Digital Literacy
- Cross Disciplinary Knowledge
- Ethical or Emotional Awareness
- Cultural Competence
- Creativity & Innovation
- Problem Solving & Critical Thinking
- Communication and Collaboration
- Actual Problem Statement
- Hardness of the Problem Statement
- Duration of the Problem Statement
- Speed with Accuracy factor of a learner
- Prerequisite Knowledge Concepts
- Visualizations of the problem statement with various simulative environment
- Adaptability

### 4.2 Workflow Design of the System:

***Central Domain Knowledge Base Generation:***

A Central Domain Knowledge Base (CDKB) is an organized collection of knowledge unique to the course or area being taught in an intelligent e-learning system. The foundation of individualized learning, content adaptation, and evaluation is this knowledge base.

**System Model:**

1. **Knowledge Acquisition:**
    - **Subject Matter Experts (SMEs):** Experts in the field define and structure the core concepts, relationships, and rules of the domain.
    - **Curriculum Materials:** Textbooks, lecture notes, presentations, and other teaching materials are used as sources of knowledge.
    - **External Resources:** Relevant research papers, articles, websites, and databases are integrated to enrich the knowledge base.

    **Knowledge Representation:**

    - **Ontologies:** Formal representations (like OWL or RDF) of the domain's concepts, properties, and relationships are created. This provides a structured way to organize and access the knowledge.
    - **Semantic Networks:** Graphical representations of knowledge are developed, linking concepts and showing their relationships.
    - **Rule-Based Systems:** If-then rules capture domain-specific knowledge for reasoning and decision-making within the e-learning system.

    **Knowledge Base Population:**

    - **Manual Duration:** SMEs meticulously review and validate the extracted knowledge, add missing information, and resolve inconsistencies.
    - **Automated Extraction:** Natural Language Processing (NLP) techniques are used to automatically extract knowledge from unstructured texts (like textbooks) and add it to the knowledge base.

    **Storage and Management:**

    - **Databases:** Relational or graph databases are used to store the structured knowledge, ensuring easy access and querying.
    - **Triple Stores:** Specialized databases (like RDF stores) are used to manage knowledge graphs efficiently.
    - **Version Control:** Changes to the knowledge base are tracked to maintain a history of updates and revisions.

**Knowledge Utilization:**

- **Personalized Learning Paths:** The CDKB tailors the learning experience based on individual learner profiles, knowledge gaps, and preferences.
- **Adaptive Content:** Content is dynamically selected and presented based on the learner's current knowledge level and learning style.
- **Intelligent Tutoring:** The system provides personalized feedback, hints, and explanations based on the learner's interactions with the knowledge base.
- **Automated Assessment:** The CDKB powers automated quizzes and tests that adapt to the learner's progress.

The process of Central Domain Knowledge Base Generation plays a pivotal role in establishing the foundation of the Learning Object (LO) repository and the individual student knowledge bases. The knowledge base is a comprehensive collection of information that serves as the backbone for both the repository and the personalized learning experience of each student. The student knowledge base is particularly significant as it contributes to tailoring the learning style and incorporating background knowledge specific to each student's needs. In Figure 1, the intricate process of knowledge base construction is visually represented. It illustrates how the LO Extractor is actively involved in generating metadata and building relationships, providing a detailed insight into the steps involved in shaping and enriching the overall knowledge base. This phase is crucial for ensuring a robust and personalized learning environment by capturing and organizing essential information for both the repository and individual learners.

***Learner Model:***
The Learner Model is a pivotal component where the learning style identifier and background knowledge identifier work in tandem. This model is proposed by maintaining the following segments:

**Learner Model Components:**

- The Learner Model is the core element.
- It incorporates two main identifiers:
  - **Learning Style Identifier:** Determines an individual's preferred way of learning.
  - **Background Knowledge Identifier:** Assesses the learner's existing knowledge base.

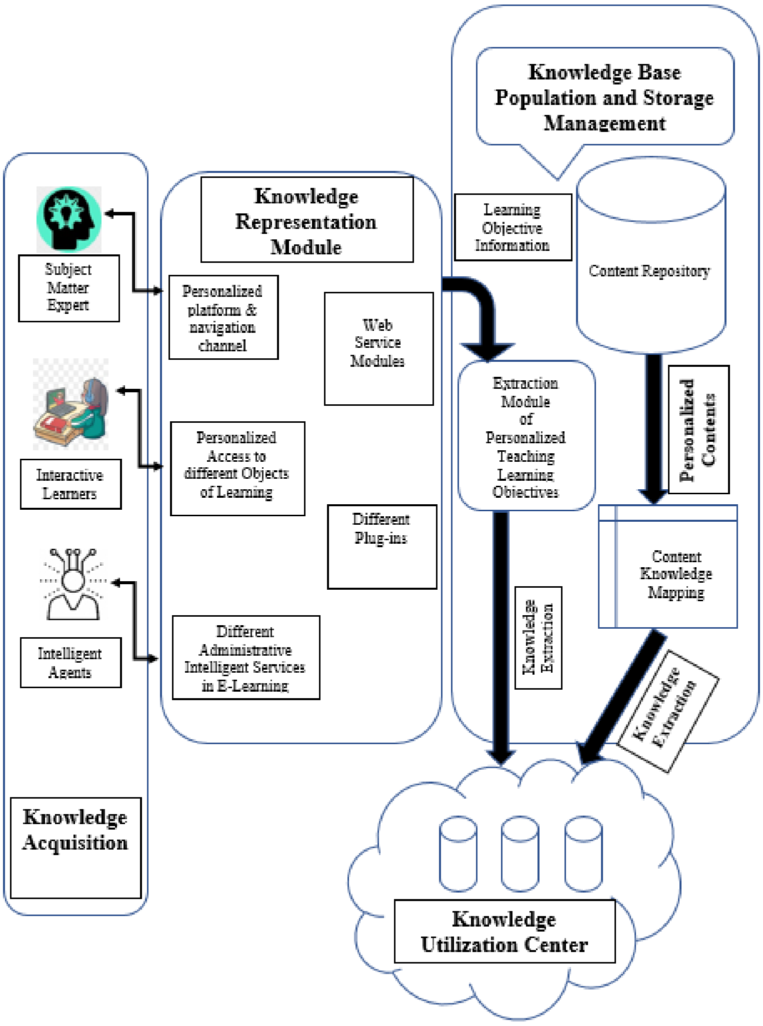


**Figure 1.** Proposed System Model for Central Domain Knowledge Base Generation

## Purpose of the Learner Model:

- Create a comprehensive profile for each learner.
- This profile provides a holistic understanding of the individual's:
  - Unique learning preferences
  - Prior knowledge

## Benefits of Integrating Identifiers:

- Enables the implementation of one-to-one delivery systems.
- Transforms learning into a more collaborative, interactive, thorough, and intelligent experience.

**Learner Model as a Roadmap:**

- Functions as a personalized guide for content delivery.
- Ensures educational materials are aligned with the learner's:
  - Preferred learning style
  - Existing knowledge level

**Visual Representation:**

As depicted in Figure 2, the visual representation is included in

- Offers a graphical depiction of the proposed Learner Model.
- Demonstrates how the two identifiers merge to create a tailored learning profile.

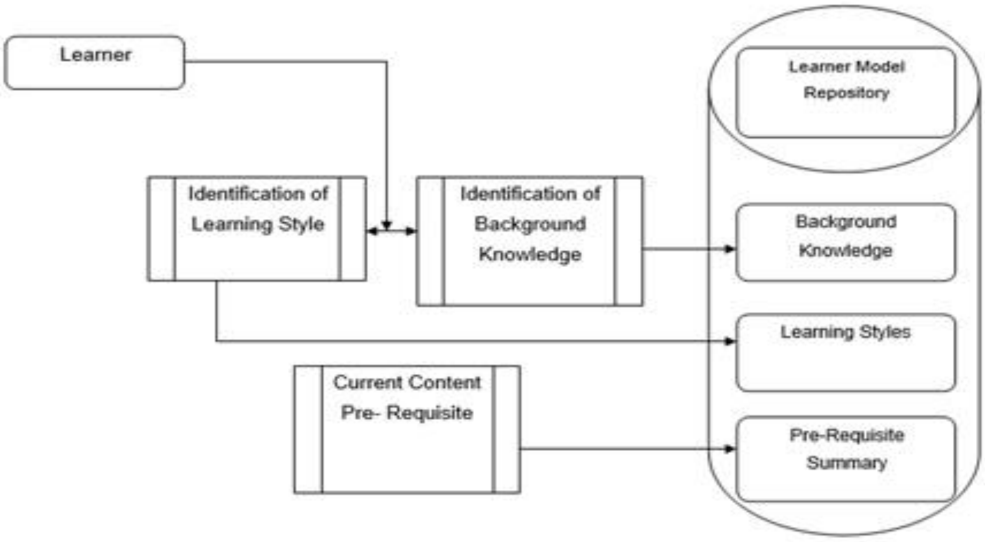


**Figure 2.** Proposed Learner Model for Central Domain Knowledge Base Generation

## 5. EXPERIMENT AND ANALYSIS

Different domains of E-Learning are analyzed to produce a robust and meaningful analysis for a Smart E-Learning System. The different analysis has been carried out as per the following aspects:

- Data Collection & Preprocessing
- Feature Engineering & Selection
- Machine Learning Model Development
- Model Evaluation & Optimization
- Deployment & Continuous Improvement

At the very first outset, the impact of Core Content Knowledge (CCK) is analyzed here as per Figure 3. CCK acts as the foundation, guiding the selection of relevant data sources, identifying potential biases, and determining the scope of the analysis. CCK is essential for understanding the underlying domain and transforming raw data into meaningful features that machine learning models can leverage. CCK informs the choice of appropriate machine learning algorithms based on the nature of the data and the specific learning objectives. CCK enables a deeper understanding of model performance, identification of areas for improvement, and interpretation of results within the context of the e-learning domain. CCK is vital for iteratively refining the e-learning framework, incorporating new knowledge and adapting to evolving learning needs.

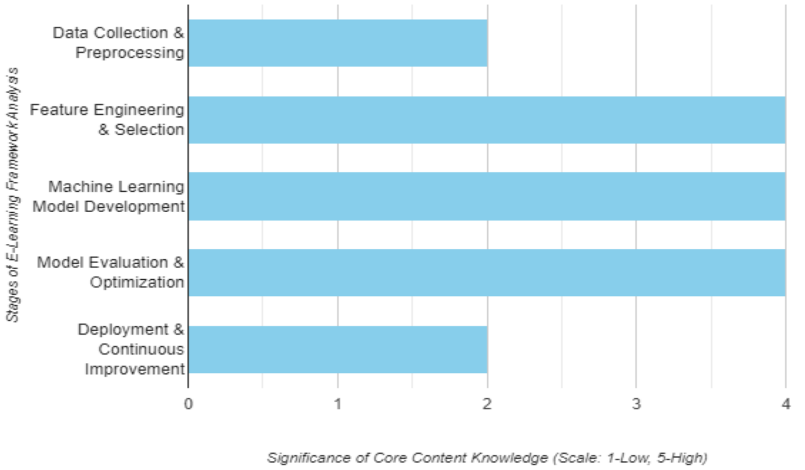


**Figure 3.** Impact of Core Content Knowledge on Intelligent E-Learning Analysis

The process of E-Learning Analysis is also based on Digital Literacy, Cross-Disciplinary Knowledge and Ethical/Emotional Awareness of the different knowledge domains. Comparative analysis of these three aspects is properly analyzed in Figure 4. Digital Literacy starts moderate, peaks during feature engineering (handling complex data), dips slightly during model development, then rises to become highly significant during deployment and improvement. This reflects the growing importance of communicating findings, understanding implications, and adapting to new technologies. Cross-Disciplinary Knowledge is consistently high throughout the process, underscoring its critical role in understanding the learning domain, formulating relevant features, choosing suitable models, interpreting results in context, and making informed decisions for improvement. A slight dip during deployment indicates a shift towards more technical and ethical considerations. Ethical/Emotional Awareness starts low, gradually increases to peak importance during model development and evaluation, signifying the critical need to address bias, fairness, and potential emotional impact on learners. This high significance continues throughout deployment and improvement, highlighting the importance of responsible AI and user-centric design.

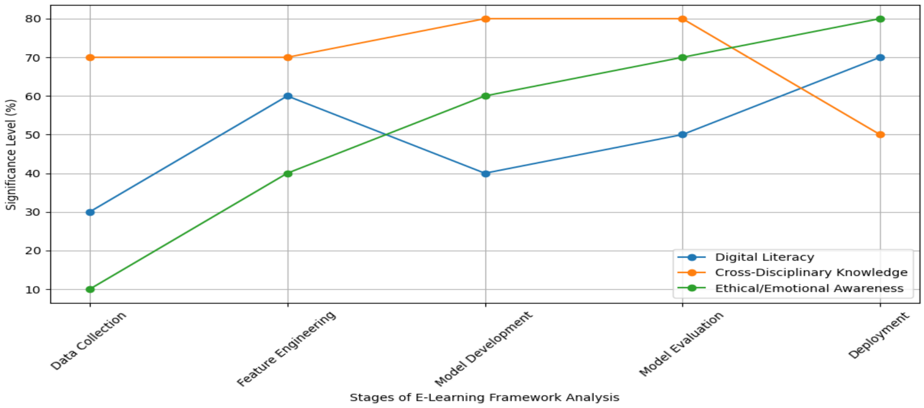


**Figure 4.** Significance of Competencies (Digital Literacy, Cross-Disciplinary Knowledge and Ethical/Emotional Awareness) in E-Learning Framework Analysis

The E-Learning Analysis Process is also dependent on Cultural Competence of the learner, Creativity & Innovation skill of the learner, Problem Solving & Critical Thinking mindset of the learner and Communication and Collaboration power of the learner. Comparative analysis of these four aspects is properly analyzed in Figure 5.

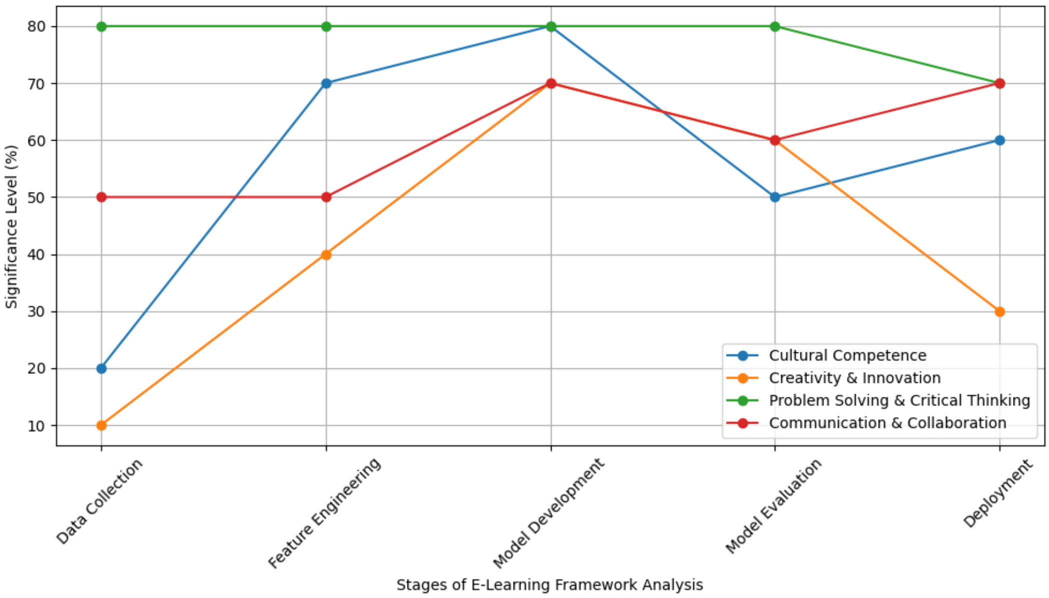


**Figure 5.** Significance of Competencies (Cultural Competence of the learner, Creativity & Innovation skill of the learner, Problem Solving & Critical Thinking mindset of the learner and Communication and Collaboration power of the learner) in E-Learning Framework Analysis

Cultural Competence starts moderate as it's crucial for understanding diverse user needs during data collection. Peaks during feature engineering and model development, ensuring that features and algorithms consider cultural nuances. It dips slightly during evaluation but remains high during deployment to ensure culturally sensitive implementation and adaptations. Creativity & Innovation is initially low, but becomes essential during feature engineering and model development, where innovative solutions are needed to address unique challenges. Remains high during evaluation to identify areas for improvement and then moderately important for creative adaptation during deployment. Problem Solving & Critical Thinking is consistently high throughout the process, underlining its importance in defining the problem, analyzing data, making informed decisions about features and models, evaluating results critically, and troubleshooting issues during deployment. Communication & Collaboration begins moderately important for data collection and then maintains high significance throughout. This highlights its crucial role in collaborating with stakeholders, explaining complex technical concepts, discussing ethical implications, and working effectively as a team to deploy and improve the system.

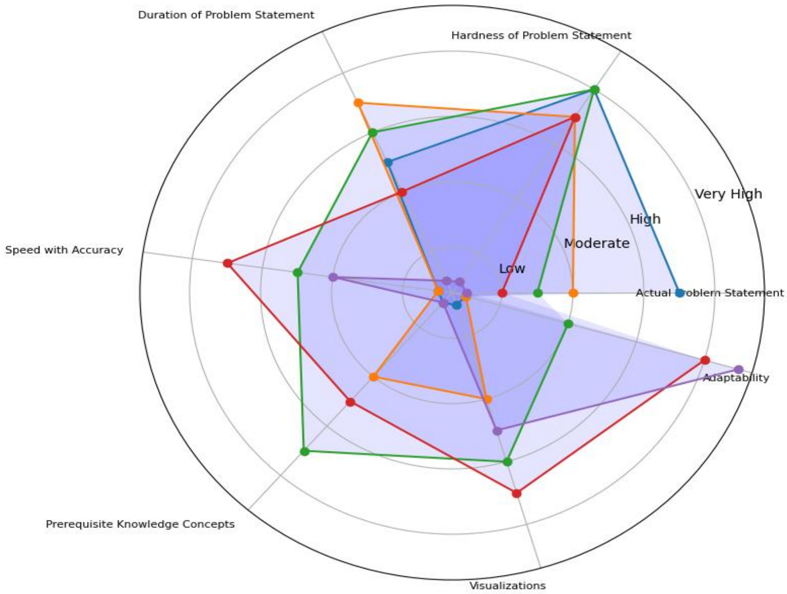


**Figure 6.** Significance of Competencies (Actual Problem Statement, Hardness of the Problem Statement, Duration of the Problem Statement, Speed with Accuracy factor of a learner, Prerequisite Knowledge Concepts, Visualizations of the problem statement and Adaptability) in E-Learning Framework Analysis

The impact of Actual Problem Statement, Hardness of the Problem Statement, Duration of the Problem Statement, and Speed with Accuracy factor of a learner, Prerequisite Knowledge Concepts, Visualizations of the problem statement and Adaptability are huge in the proper analysis of a robust Smart E-Learning system. Comparative analysis of these five aspects is properly analyzed in Figure 6. The following plots are based on the dataset of different universities. The visualizations are the representation in the form of counts versus information with respect to different criteria. Figure 7 depicts the Gender Count ratio by which a proper analysis of student count is found.

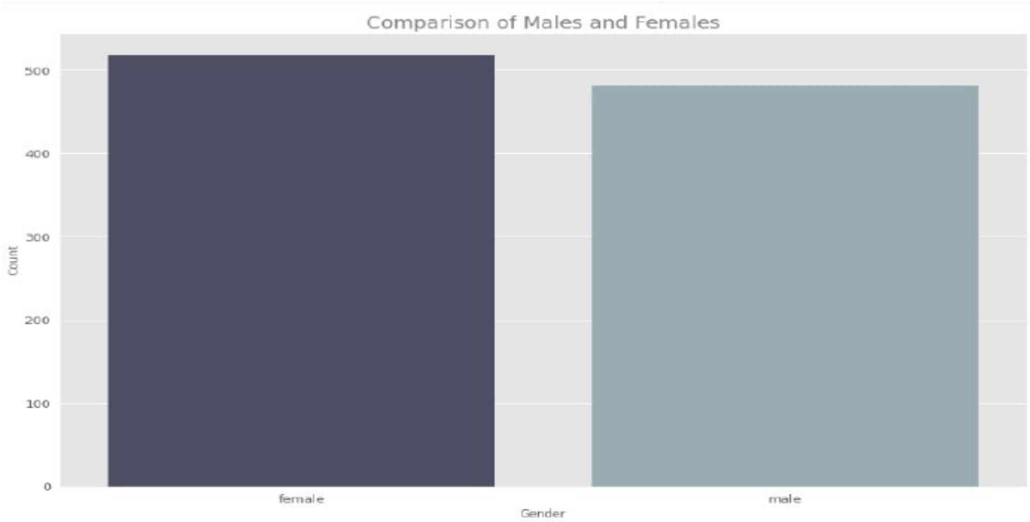


**Figure 7.** Gender Count Comparisons

Analysis of different ethnic groups amongst the student community is very essential. Figure 8 indicates that the total student community is segregated in Group A, Group B, Group C, Group D and Group E. Group A has the lowest number of students and Group C is having the highest number of students.

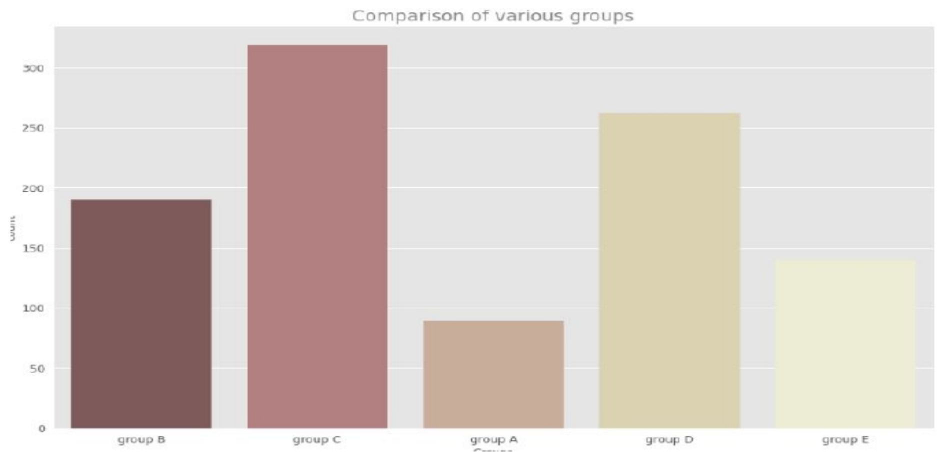


**Figure 8.** Comparison of various ethnic groups

Parental educational background plays a pivotal role for the students' interest in further studies in most cases. Figure 9 gives us an idea about the minimal qualification that most parents have. It is seen that most parents have some professional education.

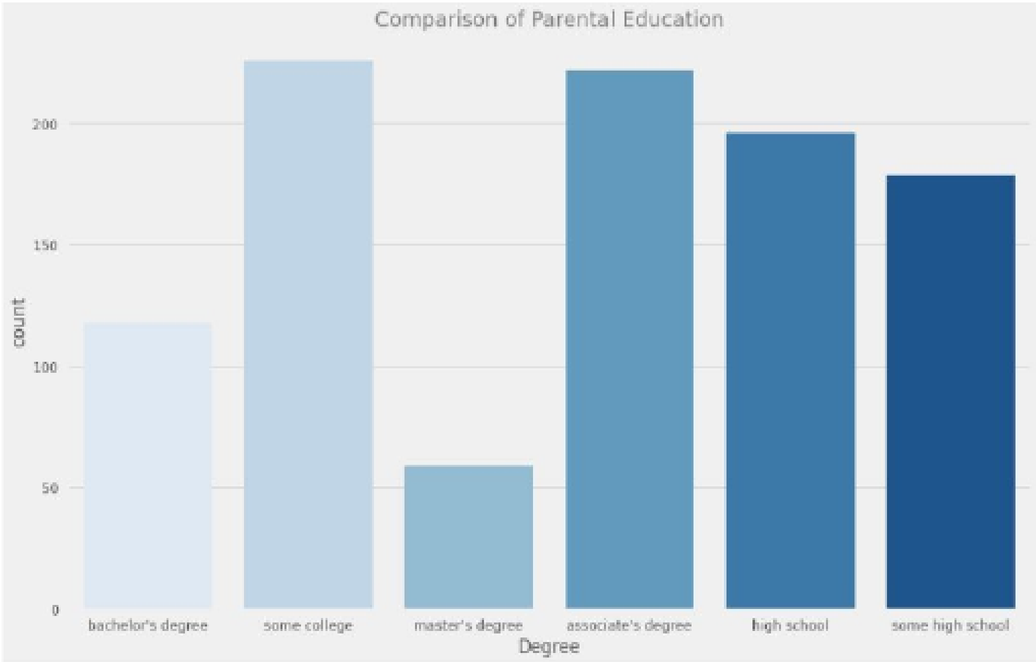


**Figure 9.** Comparison of Parental Education

In this research, synchronization factor of Parental Education with Test Preparation for the students are also needed. Figure 10 depicts that synchronization factor.

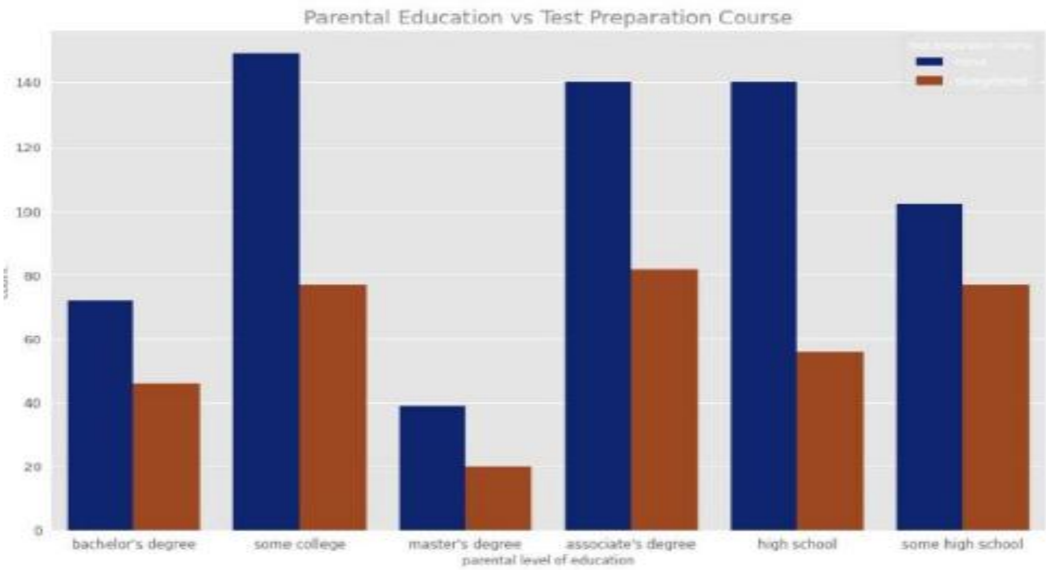


**Figure 10.** Plot to show Parental Education synchronized with Test Preparation

In this research, the methodology of comparative analysis of total scores obtained by the students has been taken care of. This gives a clear idea about the overall performance of the students. Figure 11 demonstrates the comparative analysis of the range of total scores obtained by the students.

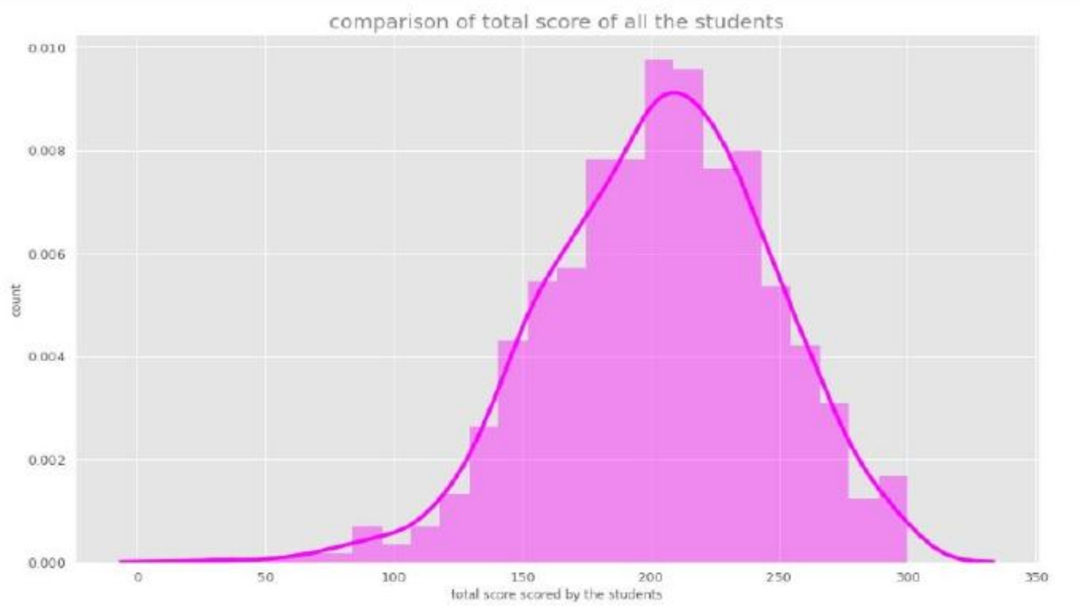


**Figure 11.** Range of total score of students

The overall percentage gathered by each student will play a very crucial role in this analysis. According to this percentage factor, students will be able to choose different courses for studies. Figure 12 demonstrates the overall percentage range scored by the students across the different courses.

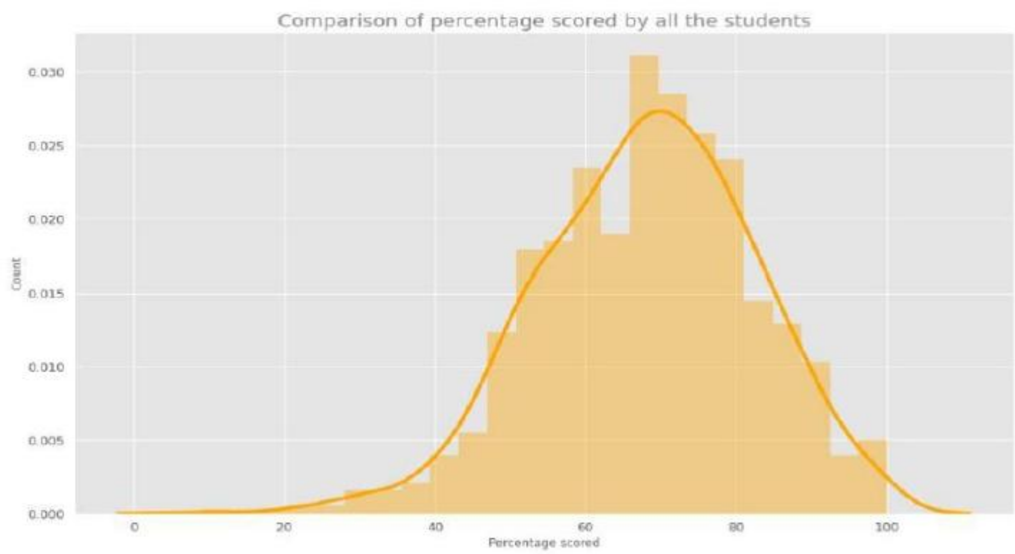


**Figure 12.** Range of Percentages obtained by the students

In this research, the following Machine Learning Models are implemented, analyzed and compared to obtain the best possible results related to students' performance.

1) Decision Tree Classifier depicts the performance analysis where the following features are considered:
   - Time Spent on Modules
   - Quiz Score (Average)
   - Prior Knowledge Level
   - Number of Forum Interactions

All these factors are classified according to the Importance Score. Figure 13 depicts the analysis.

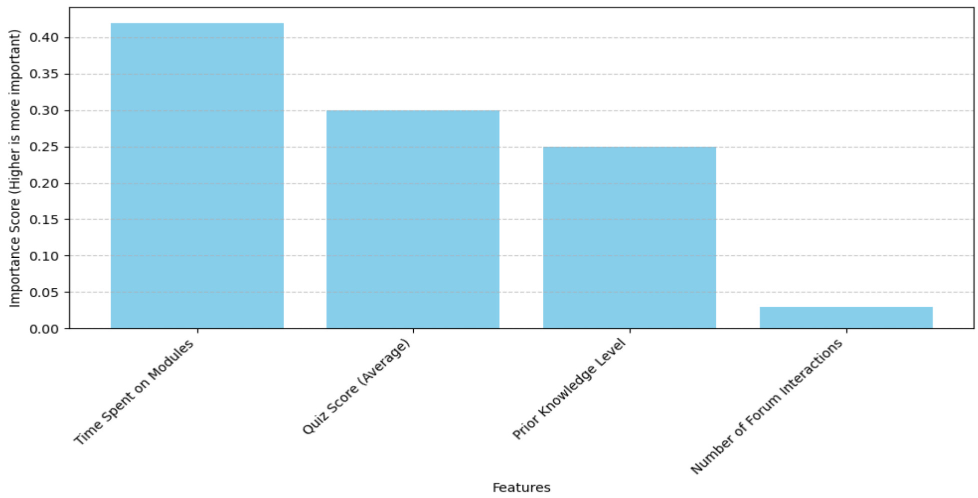


**Figure 13.** Decision Tree Algorithm illustrations of E-Learning Dataset

2) Random Forest Algorithm depicts the performance analysis where the following features are considered:
   - Time Spent on Modules
   - Quiz Score (Average)
   - Prior Knowledge Level (Self-reported)
   - Number of Forum Posts
   - Clicks on Additional Resources

All these factors are classified according to the Importance Score. Figure 14 depicts the analysis.

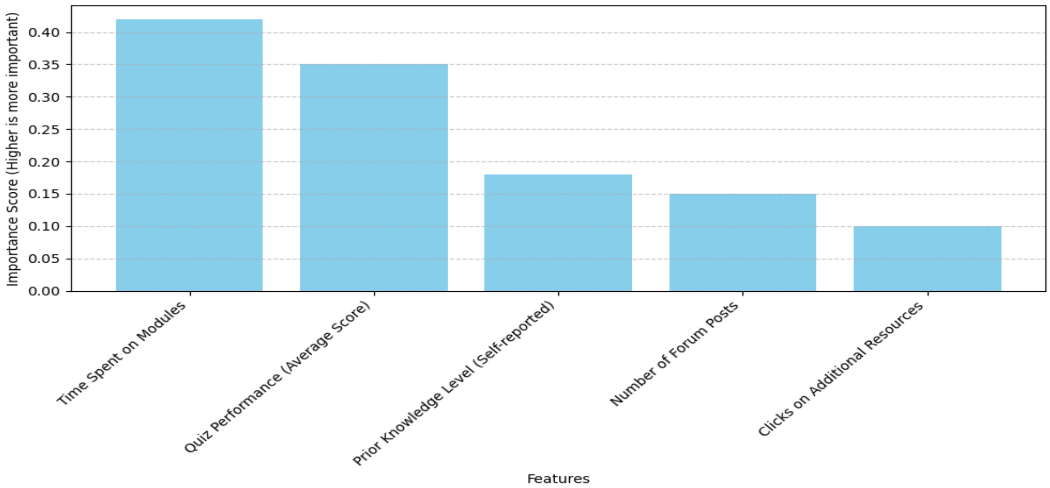


**Figure 14.** Random Forest Algorithm illustrations of E-Learning Dataset

The confusion matrix and their accuracy have been obtained based on Decision Tree and Random Forest Algorithm. By using Decision Tree Classifier 83% accuracy has been found and Random Forest Classifier provides 89% accuracy. Figure 15 depicts the whole illustration.

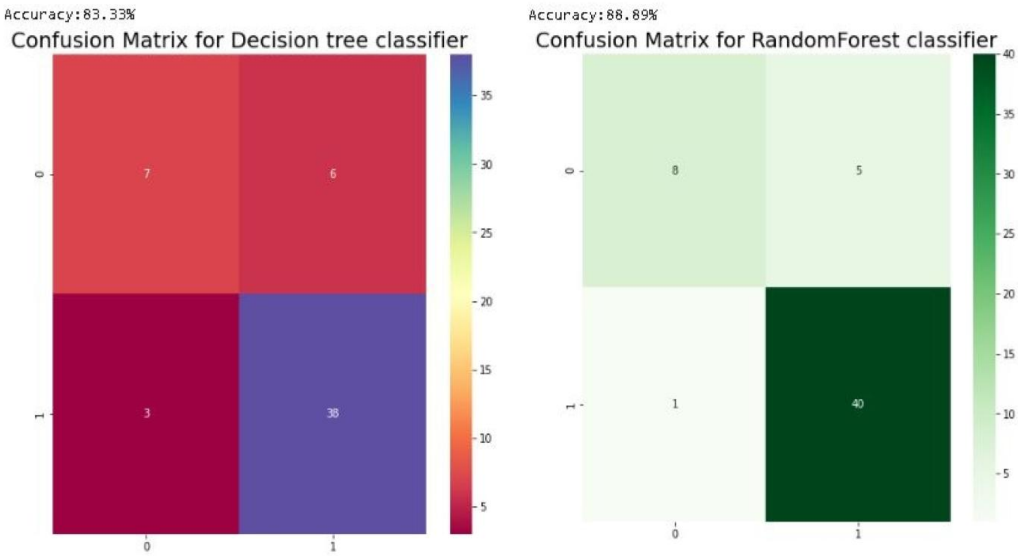


**Figure 15.** Comparison of Confusion Matrix based on Decision Tree Classifier and Random Forest Classifier

The CatBoost Model demonstrates good overall performance, as most points cluster near the diagonal line. However, there are a few outliers:

- Above the line: Learners whose actual performance was better than the model predicted (potentially due to uncaptured factors like motivation or external help).

- Below the line: Learners whose actual performance was worse than predicted (possibly due to factors like technical issues or disengagement).

Figure 16 depicts all illustrations of the CatBoost Model.

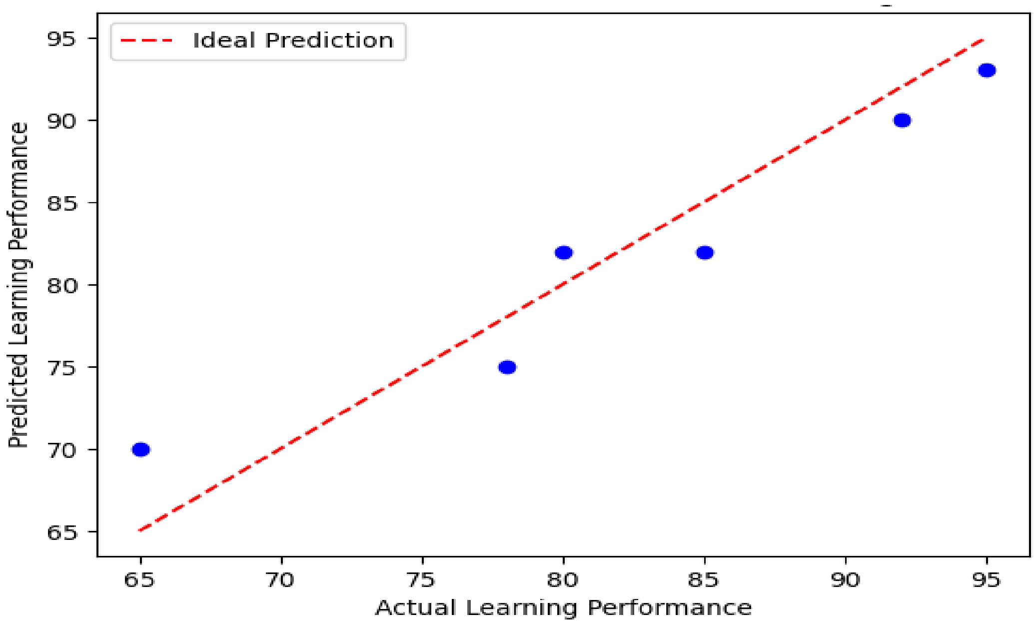


**Figure 16.** CatBoost Model performance in E-Learning

XGBoost model performance is analyzed as per the following aspects, and Figure 17 shows the detailed analysis of correlation of the above-mentioned activities.

- Final course grades
- Assessment scores
- Completion rates
- Time spent on activities
- Number of forum interactions

The confusion matrix and their accuracy have been obtained based on CatBoost and XGBoost Algorithm. By using CatBoost 91% accuracy has been found and XGBoost algorithm provides 89% accuracy. Figure 18 depicts the whole illustration.

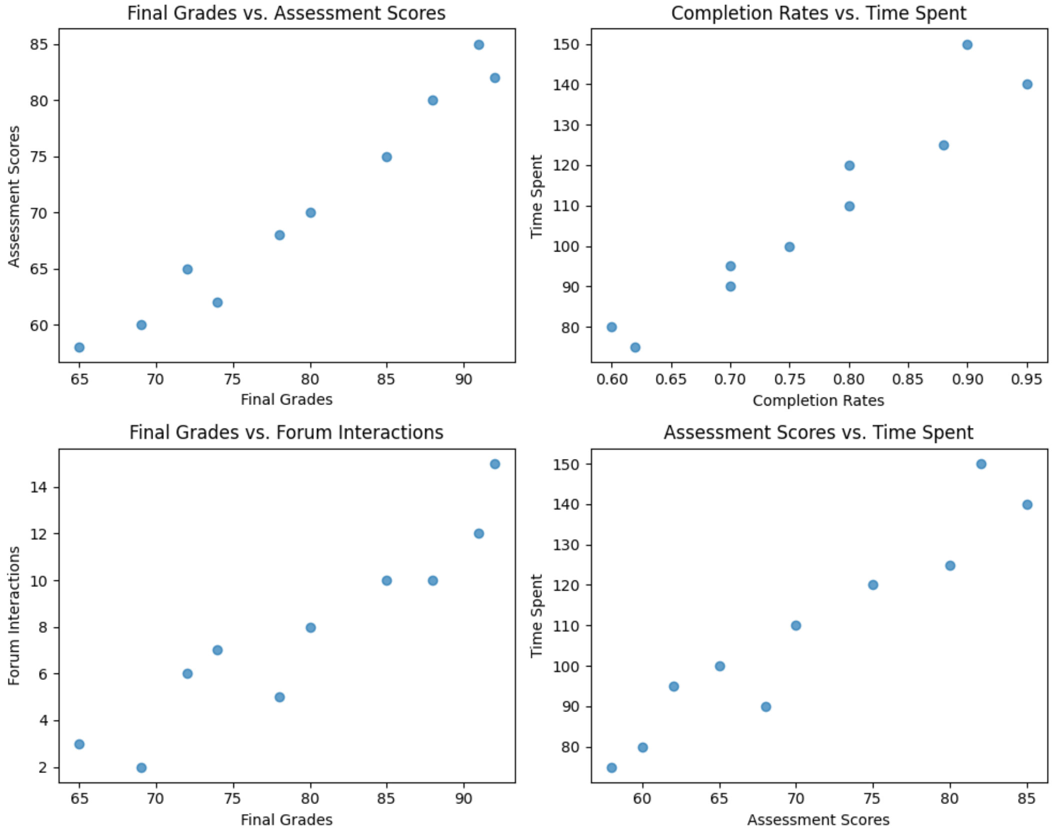


**Figure 17.** XGBoost Model performance in E-Learning

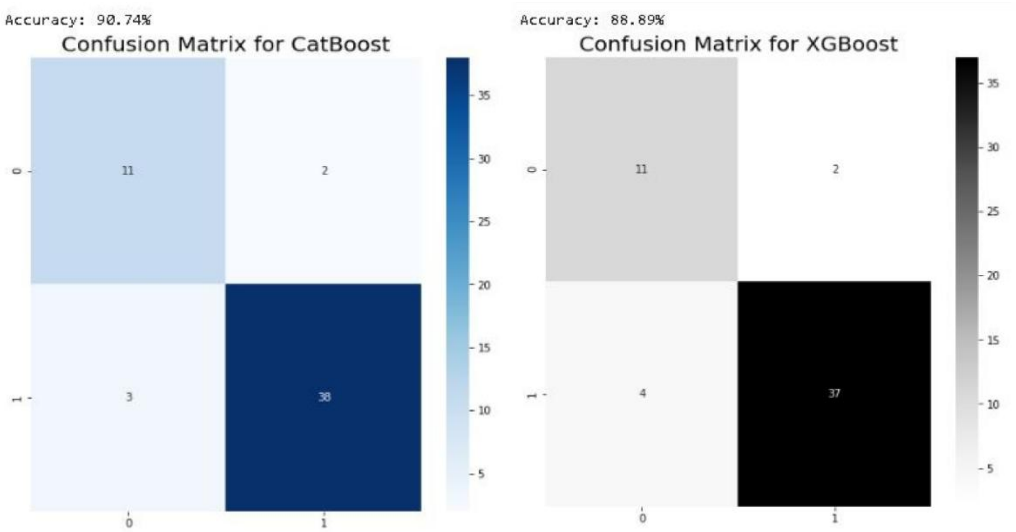


**Figure 18.** Comparison of Confusion Matrix based on CatBoost Classifier and XGBoost Classifier

Our research indicates that our model with the CatBoost may achieve a 91% accuracy rate. The other algorithms, however, are nearly at 90%. Furthermore, a lot of universities across the world adhere to the alternate report concept, so extracting a district insightful training dataset will be necessary to obtain a dataset that summarizes the entire idea. The hypothesis that a student can do better if they come from a decent family background is addressed by the research. Furthermore, it was seen that the launch element also significantly influenced the prediction. Based on these findings, it can be concluded that students in excellent health are more likely to do well, and that those who are able to access a conducive learning environment are also likely to achieve academic success.

## 6. CONCLUSION

A relatively recent addition to learning investigation (LA), open learning examination (OLA) emerged from the growing need for self-organized, deeply ingrained, and self-composed learning opportunities. OLA is able to process data collected from diverse learning environments and contexts, examined using a range of examination methods, and for multiple partners with varying rates and locations. This diverse range of OLA components is a test that should be addressed by adopting a PLA paradigm. The way that LA is currently implemented relies on a predetermined set of questions and cues, which is inappropriate for OLA, because the markers are unpredictable. The goal-question-indicator (GQI) strategy for PLA is presented in this research together with the rationale, framework, execution, and assessment nuances of the pointer motor section.

## REFERENCES

[1] Sonali Agarwal, GN Pandey, and MD Tiwari. **Data mining in education: data classification and decision tree approach**. *International Journal of e-Education, e-Business, e-Management and e-Learning*, vol. 2, no. 2, pp.140, 2012.

[2] D Fatima, Sameen Fatima, and AV Krishna Prasad. **A survey on research work in educational data mining**. *IOSR Journal of Computer Engineering (IOSR-JCE)*, vol. 17, no. 2, pp. 43–49, 2015.

[3] L Pandeeswari and K Rajeswari. **Student academic performance using data mining techniques**. *International Journal of Computer Science and Mobile Computing, IJCSMC*, vol. 3, no. 10, pp. 726–731, 2014.

[4] Rajni Jindal and Malaya Dutta Borah. **A survey on educational data mining and research trends**. *International Journal of Database Management Systems*, vol. 5, no. 3, pp. 53, 2013.

[5] ABED Ahmed and Ibrahim Sayed Elaraby. **Data mining: A prediction for student's performance using classification method**. *World Journal of Computer Application and Technology*, vol. 2, no. 2, pp. 43–47, 2014.

[6] Amjad Abu Saa. **Educational data mining & students' performance prediction**. *International journal of advanced computer science and applications*, vol. 7, no. 5, 2016.

[7] Arham Muslim, et al., **A Rule-Based Indicator Definition Tool for Personalized Learning Analytics**. *Smart Learning Environments*, 2018.

[8] Marie Bienkowski et. Al, **Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief**, *U.S. Department of Education Office of Educational Technology*, 2012

[9] Filippo Sciarrone et. Al. **Machine Learning and Learning Analytics: Integrating Data with Learning**, IEEE, 2015

[10] Fedor Duzhin et. Al. , **Machine Learning-Based App for Self-Evaluation of Teacher-Specifific Instructional Style and Tools**, *Educ. Sci.* 2018

[11] Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A., & Schroeder, **Design and Implementation of a Learning Analytics Toolkit for Teachers**. *Educational Technology & Society*, 2015.

[12] Daniel Spikol et. Al. **Using Multimodal Learning Analytics to Identify Aspects of Collaboration in Project-Based Learning**, *CSCL 2017 Proceedings*, 2017.

[13] Muslim, A., Chatti, M., Mughal, M. and Schroeder, U. **The Goal - Question - Indicator Approach for Personalized Learning Analytics**, *In Proceedings of the 9th International Conference on Computer Supported Education (CSEDU 2017)*, 2017.

[14] Daniel Spikol et. Al. **Supervised machine learning in multimodal learning analytics for estimating success in project-based learning**, *J Comput Assist Learn*. 2018.

[15] Dezhao Song, **Natural Language Question Answering and Analytics for Diverse and Interlinked Datasets**, *Proceedings of NAACL-HLT* 2015

[16] Pardo et. Al. **Using learning analytics to scale the provision of personalised feedback**, *British Educational Research Association*, 2017

[17] A. Zafra and S. Ventura, **Predicting student grades in learning management systems with multiple instance learning genetic programming**, *in Educational Data Mining - EDM 2009*, Cordoba, Spain, July 1-3, 2009.

[18] Zhai C. **Interactive information retrieval: Models, algorithms, and evaluation**. *In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2444-2447, Jul 25, 2020.

[19] Suhartono D, Khodirun K. **System of Information Feedback on Archive Using Term Frequency-Inverse Document Frequency and Vector Space Model Methods**. *International Journal of Information Systems*. vol. 3, no. 1, pp. 36 – 42, 2020.

[20] Mustapha Riad, Soukaina Gouraguine, Mohammed Qbadou, and Es-Sˆaadia Aoula. **Towards a new adaptivee-learning system based on learner's motivation and machine learning**. *In 2023 3rd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, pp. 1–6, 2023.

[21] Samina Amin, M. Irfan Uddin, Ala Abdulsalam Alarood, Wali Khan Mashwani, Abdulrahman Alzahrani, and Ahmed Omar Alzahrani. **Smart e-learning framework for personalized adaptive learning and**

**sequential path recommendations using reinforcement learning**. *IEEE Access*, vol. 11, pp. 89769–89790, 2023.

[22] Lam TY, Dongol B. **A blockchain-enabled e-learning platform**. *Interactive learning environments*. vol. 30, no. 7, pp. 1229-1251, 2022.

[23] Riza LS, Piantari E, Junaeti E, Permana IS. **Implementation of the Gamification Concept in the Development of a Learning Management System to Improve Students' Cognitive In Basic Programming Subjects Towards a Smart Learning Environment**. *ADI Journal on Recent Innovation*, vol. 5, no.1, pp. 43-53, 2023.

[24] Hjørland B. **Information retrieval and knowledge organization: A perspective from the philosophy of science**. *Information*, vol. 12, no. 3, pp. 135, 2021.

[25] Azad HK, Deepak A. **Query expansion techniques for information retrieval: a survey**. *Information Processing & Management*, vol. 56, no. 5, pp. 1698-1735, 2019.

[26] Guo J, Fan Y, Pang L, Yang L, Ai Q, Zamani H, Wu C, Croft WB, Cheng X. **A deep look into neural ranking models for information retrieval**. *Information Processing & Management*. 2020 Nov 1;57(6):102067.

[27] Martinez-Rodriguez JL, Hogan A, Lopez-Arevalo I. **Information extraction meets the semantic web: a survey**. *Semantic Web*, vol. 11, no. 2, pp. 255-335, 2020.