Article

An Intelligent Model for Parametric Cognitive Assessment of E-Learning-Based Students

Muhammad Saqib Javed [1,2,*], Muhammad Aslam [1] and Syed Khaldoon Khurshid [1]

1 Department of Computer Science, University of Engineering and Technology, Lahore 54890, Pakistan
2 Department of Computer Science and IT, Virtual University of Pakistan, Lahore 54500, Pakistan
* Correspondence: saqibjaved@vu.edu.pk or saqibcap@gmail.com

Abstract: In an e-learning environment, question levels are based on Bloom's Taxonomy (BT), which normally classifies a course's learning objectives into diverse levels. As per the previous literature, the assessment procedure lacks accuracy and results in redundant keywords when automatically assigning Bloom's taxonomic categories using a keyword-based approach. These assessments are considered challenging as far as e-learning-based students are concerned, as the text feed is the only instrumental testing part. Student assessments are limited to multiple-choice questions and lack an evaluation of students' text-based input. This paper proposes a natural-language processing-based intelligent deep-learning model that relies on parametric cognitive assessments. By applying class labels to students' descriptive responses, the proposed approach helps classify a variety of questions mapped to BT levels. The first contribution of this work is a compiled dataset of the assessment items from 300 students, who were tested on 20 questions at each level. Each level is calculated by combining the responses from all students, resulting in 6000 questions per cognitive level for a total of 36,000 records. The second contribution is the development of an intelligent model based on a recurrent neural network (RNN), which not only predicts Bloom's question level but also learns it over further iterations. The students' text-based answers are accessed to gauge performance using a refined question pool gathered through the RNN model. The student dataset is mapped and tested using the NLP model for further classification of the students' cognitive levels. This assessment is related to the formulation of questions and the compilation of Episode 2 for assessment. The third contribution is the comparison and demonstration of the improvements in learning using a parametric cognitive-based assessment in an episodic manner. Improved classification accuracy was attained by adding more processing layers based on the iterative, RNN-based learning model to achieve the vital threshold difference. The cognitive based questions pool classification achieved by RNN results in 98% accuracy. The resulting student assessments, based on performance, increased to an accuracy ratio of 92.16% and a precision ratio of 92.36% at an aggregate level based on the Random Forest classifier. We claim that our work serves as an initiative for effective student evaluations in interactive and e-learning-based environments when handling other types of inputs, like mathematical, graphical, and multimodal inputs.

Keywords: Bloom's taxonomy (BT); artificial intelligence (AI); natural-language processing (NLP); text classification; learning objectives; deep learning; machine learning; Random Forest (RF); recurrent neural network (RNN); long short-term memory (LSTM)

1. Introduction

The ability to think in any domain is considered to be at the core of all learning activities [1]. Educational institutions evaluate the thinking process through teaching and quality assessments to maximize learning by students. In a normal educational setup, the procedure is as follows: Firstly, teaching is followed by an understanding of the process, whereby teachers design teaching materials that align with the course's learning objectives based on students' thinking abilities [2]. Secondly, a pool of questions based on quality assessments is developed, which relies on the developed test items while focusing on the levels of taxonomy [3]. Finally, assessments and testing-based evaluations are completed by conducting a written descriptive type of online examination. Addressing assessment challenges not only helps in gauging cognitive-level scales but also helps in mapping cognitive skill-set levels in a convenient amount of time and number of sessions based on the projected domain, such as an AI-based question pool for students to answer.

Ultimately, it also helps in understanding what the teacher is delivering to students, as well as the scale of the students' learning maps in their learning process [4]. COVID-19 forced academics to face the dire need to not only engage students effectively to ensure regular study online but also to implement an assessment process specific to students' cognitive behavior and learning levels.

Thinking behavior is classified into the following three concrete domains: cognitive, affective, and psychomotor. The cognitive domain mainly focuses on the scope of application [5] and can be divided into six structural levels that depict critical contemplative behavior in terms of students' learning procedures. This approach is based on the association of different keywords/action verbs to distinguish each level from each other.

The cognitive-level-based taxonomy and its names have been replaced with new keywords, such as "creating synthesis" [6]. Capability in online learning environments is measured based on metrics such as learner participation in asynchronous virtual sessions. Students in e-learning environments lack physical interaction with respect to learning and assessment. This gap between students and teachers creates hurdles to not only aligning the objectives of the course but also the measurement of learning objectives in any course domain. Gauging the completeness and correctness of instructional designs for students is a key aspect to consider and must be addressed in a timely and thorough manner so that students receive an effective delivery of content with a feedback assessment mechanism. Optimistically, this aligns course learning objectives with baseline parametric assessments of e-learning-based students to scale up their education through cognitive assessments. In usual learning spheres, students often focus on grades and course completion without fully understanding the material, which results in their cognitive levels remaining unknown. This affects not only the aggregate learning curve but the teaching method as well, making it challenging to deliver interactive content that engages students in line with CLOs and improves learning experiences from assessment results.

The assessment mode was previously manual for both conventional and technology-based students and did not focus on mapping Course Learning Outcomes (CLOs) to a projected Bloom's Taxonomy-based question bank. It is, therefore, concluded that the assessment mode is time-consuming in transparent assessment procedures, as it requires intelligent assessment and grading using text-based classification and the assignment of keywords to weights for correct procedures. Some research has been conducted that endeavors to systematize this process by means of keyword-based probing followed by machine learning and natural-language processing and procedures [7,8], but it lacks scope and time constraints to handle computational complexity.

In previous literature [7,9,10], researchers used keyword-based approaches to classify a pool of questions on the scale of BT levels. Although remarkable results have been

achieved, these approaches suffer from one major weakness, which is the redundancy of keywords based on different taxonomic cognitive levels [10].

The problem of keyword overlapping is considered to be one of the major mistakes of programmed keyword-based methods for the classification of learning-based question sets. To differentiate the CLO, which is mapped with each BT level, ML techniques are much more beneficial for ensuring not only completeness and correctness as part of the assessment process but also coherence. Therefore, topical studies [5,11,12] have devoted themselves to complementary machine-learning methods for classifying not only CLOs but also question-based datasets. However, ensuring the effectiveness of knowledge based on these studies still lacks accuracy and has data redundancy [13]. It results in deviating results since these studies employ straight machine-learning (ML) approaches, with working models not restricted to accuracy and capacity building [14].

Improved automated text-classification approaches require improvement based on the previous studies of similar areas [7]. Recently, deep-learning approaches have exhibited positive results compared with the baseline machine-learning algorithms, which are used in text-based classification [15]. The follow-up problem in this domain is the existing marked datasets of CLOs and questions based on cognitive levels. Researchers of the same domain encourage existing datasets rather than using datasets available online. In the focused research, the dataset is actual and unique, and requires the criticality of its contents to assess at its best. Hence, in this research, the dataset gathered and compiled was based on actual e-learning-based student data, specifically for BT classification.

Bloom's Taxonomy is a framework used in educational setups to classify learning objectives, which are based on their complexity and depth [16]. It is often represented with six levels detailed as follows:

Remembering: The recall of the basic facts and information.
Understanding: The explanation of ideas or models.
Application: Real-world usage along with associated knowledge in the latest situations.
Analysis: The splitting of information into parts to understand relationships or patterns.
Evaluation: Judgments based on criteria with evidence.
Creation: The combination of parts to form a new whole or propose new ideas.

Educationalists use this taxonomy for designing lessons followed by assessments that encourage intense learning credentials for simple recall in order to achieve higher-order thinking skills [17].

Natural-language processing (NLP) is a branch of artificial intelligence that mainly focuses on understanding, interpreting, and giving proper responses to human language in a meaningful and useful way. Examples of NLP in action include the following:

Speech recognition: Converting spoken words to text (like virtual assistants Alexa, etc.).
Machine translation: Translating a text from one language into another language (like Google Translate).
Sentiment analysis: Determining the tone for split text portions (like analyzing reviews).
In general, NLP combines language with computer science to handle contextual tasks like grammar, syntax, etc.

A recurrent neural network (RNN) is a kind of artificial neural network (ANN) designed for processing complex and critical sequences of data, like time-series data in the form of text. RNNs operate on loops that allow them to "memorize" inputs. It is resultantly confined to tasks where perspective matters.

Applications of RNN include the following:
Text generation: Forecasting the next expected word in a sentence.
Speech-to-text conversion: Translating core spoken language into baseline written text.

Predictive analytics: The forecasting of stock values, whether based on historical data or a flag indicator.

However, RNNs may struggle with long-term dependencies, a challenge that is addressed by the introduction of long short-term memory (LSTM) networks, which work alongside RNNs, as per our proposed research.

The present research aims to present state-of-the-art unified student data in the form of a pool of test items, with the training and testing of the same data using a ratio of 70:30, and then assessing students by projecting same-level tested questions in a two-frame sequential episodic manner. Ultimately, both episodes' assessment data are compared after making predictions through a Random Forest classifier supported by RNN at the time of test-item development for random-weight management.

The core aim is to examine learners' levels of participation in online courses for both synchronous and asynchronous communication modes, which are based on learner performance. This leads to continual clarity at the level of interaction factor identification. To inspect and collect learner participation data, the data were collected from the learning-management system (LMS) log, including occurrence and length of course, graded discussion board postings, and, ultimately, the final grades. To examine synchronous learner collaboration, conversation logs are collected from the conversational agent to guarantee learning pedagogy [18]. The research focuses on the diverse aspects of enhancing existing learning-management systems, which may lack responses as they depend on the agent to respond to the system and operate with human–computer interaction (HCI) modeling.

Research Objectives

The research objectives are as follows:

✓ Cognitive skills need to be assessed in an online environment to evaluate students.
✓ Student performance and cognitive engagement relate to one another in online education.
✓ Levels of Bloom's Taxonomy will be used to evaluate student learning states.
✓ The machine-learning model is used for the assessment of cognitive abilities through textual answers.
✓ To contribute to academia in a way that ensures distance-learning education promotion with guaranteed student presence.
✓ To inculcate effective modeling, which includes Knowledge units, Communication units, and Response units, these units work together to efficiently measure student assessment and effectively measure the responses during action.
✓ To provide possibly powerful scoring and reporting as real-time feedback mechanisms. Cognitive assessment will be at the student end for measuring student performance based on delivered content.
✓ The suggested research enhances the cognitive-evaluation process for an e-learning platform, resulting in transformed e-learning.

The research contributions are as follows:

(1) A dataset pertaining to 20 questions at 6 cognitive levels for 300 students, which makes a sum of 36,000 entries in a single episode.
(2) A custom-built LSTM classification model to ensure BT on examination questions.
(3) A parametric cognitive-based assessment, in an episodic manner, is a unique idea to diagnose and evaluate the learning using a pretrained question-pool bank.

The research paper is divided into sections as follows: Section 2 presents the literature review. Section 3 includes the proposed methodology with its construction design. Section 4 includes the experiment setup with a detailed comprehension of dataset modeling. Section 5 provides the results and discussion. Section 6 is based on the episodic results. Section 7 presents the episode result comparisons, which is an extended version scope of

this comprehensive research. Section 8 leads to the conclusion covering the extensive detail of the problem set, the research gap, and what the research achieved. Section 9 details the future work.

2. Literature Review

This section details and correlates background studies in connection with BT based on prevailing approaches for automatic cataloging of Course Learning Outcomes (CLOs) mapped to questions associated with BT and an indication of pertinent techniques by introducing and including text classification based on deep learning.

Relationship Between BT and Cognitive Domain

The detailed relevance of BT was considered along with the cognitive domain used in educational institutions for defining and redefining Course Learning Outcomes to create a relevant domain question pool [2]. The teaching material of the artificial intelligence (AI) domain is developed to assess questions pooled for finding student learning curves. A comprehensive level of research has been carried out on the amended edition of BT as it is linked with the cognitive domain, namely Remembering, Comprehension, Application, Analysis, Creation, and Evaluation) [6].

Recently, researchers have emphasized the significance of considering Bloom's Taxonomy, specifically for the classification of tweets based on pattern identification [19]. Previously conducted research showed that the cognitive domain defined versatile thinking behaviors by simply recalling memories to ensure dense logical skills in the aspect of student aggregate learning. The literature research analyzed the Course Learning Outcomes of fundamental courses, specifically those offered in African countries. The result of the study shows the first two levels of the cognitive domain (Knowledge and Comprehension), which have contributed to a 58% mapping of the composite defined CLOs. In the meantime, the defined Application level and the rest of the cognitive levels result in an aggregate percentage of 27% and 15%, respectively. Recently, researchers have used BT to analyze student questioning skills [20,21]. There was a survey published on information retrieval with natural-language processing and machine-learning-based research representing a question–answer-based community. The focus was mainly on automated text analysis, with the aim of offering a better understanding of the data to its users [22]. The comprehensively conducted survey demonstrated the emphatic importance of data-mining in text-based classification [23–25]. A study was conducted on the cognitive level by conducting an assessment of short essay-based questions from two veterinary courses at Utrecht University. The simplified classification tool was used to map taxonomy levels. Baseline classifications were made by subject-matter experts and by some non-subject-matter experts for control [26].

Emphasis has been put on the regime of innovative assessment in promoting student aggregate-level performance-based development thresholds. There exist some traditional assessment methods that often overlook the importance of physical education. The blended mode of education normally escalates cognitive, affective, and social domains. Educationalists are now rapidly moving towards new assessment strategies for comprehensive timeline-based student progress and skill-set enhancement but lack the competency testing of brain thinking, which is based on thinking ability and the ability to answer any cognitive question set in a specified time. Previous research covers this project aspect in the form of performance-based assessment, self-assessment, peer assessment, and ultimately, a technology-based enhanced level of assessment [27].

Cognitive ability testing, referred to as CogAT, mainly focuses on student problem-solving capabilities and reasoning-based abilities. Verbal measurement is followed up

by quantitative measurement with nonverbal level assessment. The CogAT system has been updated three times in the last 20 years, with its current form named CogAT Form 6 [28]. The COVID-19 pandemic emphasized the need for an online distance-learning mode of education to reduce the technology gap and provide services to students in their homes. COVID-19 prompted a revolution not only in the imparting of knowledge aided by technology but also in effective assessment by increased student participation in that mode of education.

The outcome of the study is the cognitive-level-based questions. Hence, it proves the importance of the cognitive domain in students' effective learning transformation, as demonstrated in detail in Table 1. It shows the publication year along with the results of the projected research and any research gaps that should be filled by our proposed research.

Table 1. Comprehensive literature review.

Year—Ref. No.	Methodology	Participants	Results	Research Gap
2016—[3]	NLP technique to assess student response	Online students'	Identification of online learning challenges	Required to consider student mental state
2021—[4]	ML-based text analysis for measuring conceptual knowledge	Course—Computer Signals	Automatic evaluation of learners	It needs to be tested on a higher amount of dataset
2024—[18]	Multiagent-based cognitive assessment	Undergrad students' data—SE	Accuracy of 91%	Small number of questions
2021—[21]	Intelligent agent-based system for tracking students' performance	MOODLE Platform	Innovative way to measure students' involvement	Need to determine cognitive level
2025—[27]	Timeline based on student's assessment	60	Skill-set-based results	Competency Testing Lacking
2025—[28]	CogAt Systematic Assessment	24	Effect size-based estimations	Nonverbal assessment scale
2025—[29]	Teaching-Research Correlation	Undefined	95% Confidence Level	Just medical level students tested based on questionnaires
2015—[30]	An automated approach to determine the message's cognitive level	Discussion forum messages	Enhanced the model of assessment	Need to evaluate performance
2021—[31]	Prediction model of students' performance	Random Volunteer Student Dataset	Linear regression based on Regularization	Need to focus on qualitative exams and deep-learning technique optimization.
2024—[32]	Students' mental state assessment using Roberta, and BERT model	Text data from MOOC	Overall accuracy more than 92% but on biased data	It needs to be tested on biased free higher volume of dataset
2021—[33]	Examining Blooms taxonomy using multiple choice questions	137 students tested	Achieved 74.9% Accuracy	Text-based responses of students' need to be tested
2010—[34]	Use of the KNN algorithm and RBT to classify cognitive states	100 CS Students	Accuracy is 84%	Need to compare neural network-based models
2024—[35]	ML-based Cognitive Analysis	The Dataset of the assessment data of school	Accuracy of 82.2%	Model integrator fixed parameters-based, online recommendation system

Table 1. *Cont.*

Year—Ref. No.	Methodology	Participants	Results	Research Gap
2021—[36]	WordNet similarity algorithms with NLTK and cosine similarity algorithm	Wayamba University, Sri Lanka—Random	70% Accuracy	Fixed Dataset
2022—[37]	Questions Classification based on Text-Mining Method	Exam Questions	Analyze Question Paper Using BT	Required to map cognitive level with responses
2023—[38]	Regression model for categorizing subjective questions	Computer Networks Course—Item Bank	BT-Based Level Prediction	Testing is required using course questions
2024—[39]	Use of SVM model for predictive analysis and K-Mean Clustering for descriptive analysis	University Students	Accuracy is 87%	Need to improve performance on defined dataset

The study aims to explore and establish a correlation between cognitive flexibility and research performance in medical students to cover any shortfall and to understand the status of research teaching. The prevalence calculation used is 50%, with a result of a confidence level of 95%, followed up with an acceptable error margin of 5% [29]. The relative study based on the classification of the summative assessment relies on the pool of questions established for effective data modeling based on convolutional neural networks (CNN) [40]. A classification model is used to classify exam questions based on Bloom's Taxonomy, which uses a method for classifying questions automatically depending on the feature-based extraction using IDF and word2vec [41].

E-learning models mainly work in virtual education. There are different formats, like fully online learning and hybrid learning. The latter uses a combination of in-person presence and online technological components for Massively Open Online Courses (MOOCs). MOOCs enable course engagement beyond globally recognized programs. There are various formats, such as fully online courses, blended learning models, and Massive Open Online Courses (MOOCs). These explore innovative pedagogical approaches based on popular theories and adaptive technologies to serve diverse learning styles [42]. The development and formulation of the question-generation framework are mainly based on the keyword-based phrase method for online learning. The framework constraint is that the generated questions depend upon the stated learning outcomes and skills from Bloom's Taxonomy [43].

Computerized-based adaptive testing is a form of assessment that uses artificial intelligence for cognitive diagnosis of student learning ability. By employing an adaptive embedded selection algorithm, it basically selects a question set based on the estimation of the learner's competence, with a track record of each step of the assessment [44]. Computerized adaptive tests (CAT) offer individualized questions based on student ability. Hence, it is normally designed to properly measure skills in relation to a cognitive level (e.g., Remembering, Comprehension, etc.)

The focus of previous research was primarily on objective-type papers to measure student cognitive abilities using the six BT levels [45]. To assess student cognitive skills, some researchers worked on a dataset of programming code, while others focused on diagram analysis to connect it to mental states [30]. However, there has been insignificant research on cognitive psychological evaluation using textual answers to the pool of questions developed at each stage and level of BT with the defined parametric threshold of assessment ranging from Below Average to Good student performance.

The cognitive level of the learner's analysis usually entails manual coding, which requires a high level of research capacity and proficiency from the specialist. It is considered a very hectic and time-consuming task. In addition to this, and due to the vast amount of information that needs to be filtered through various research advancements based on manual analysis of textual data, it is considered to be a tedious task [46]. Outstanding progress has been achieved with the extensive use of artificial intelligence (AI), specifically in machine-learning techniques, for performing different classification tasks. Machine learning constitutes a novel and modern technique for data analysis and has produced promising outcomes with some substantial abilities [47]. Data representation complexity can be automatically discovered and hence added to the analyzed model [48]. The purpose of this research is to not only identify intelligently the cognitive level of the question pool but also to automatically assess the student's cognitive level in online assessment mode. It, therefore, explores the factors that impact the learner's cognitive level during the learning process based on a machine-learning-based approach in order to specify the cognitive level achieved by students on a parametric weighted threshold of performance metrics in an episodic manner.

Generative AI is transforming the basic assessment of learning from traditional methods to adaptive authentic processes [49]. The transition effect of GenAI increases the potential for creating personalized and adaptive assessments by enhancing student understanding and advancement in relevant domains [50]. This enables advancements in generative agents, which are autonomous and adaptive and operate independently by ensuring continuous user interaction, illustrated by relevant tools, e.g., AutoGen (preprint) [51].

Cognitive skills are assessed in an online environment to evaluate students. Hence, e-learning mentors must be aware of their students' cognitive states in online classrooms if they want to motivate them to study. Due to physical absence, students lose concentration and are distracted easily. Addressing the environment is of utmost importance. To assess student cognitive abilities, the proposed system takes textual answers as input during an online exam, which includes subjective AI questions that fall within a range of classified categories of BT.

The concept of BT is used for estimating the accessibility and higher-order thinking skills with respect to comprehension-based questions [52,53]. The conducted research study focused on an English textbook for Grade 10, with results lacking high-order thinking competencies with a percentile of 85% in terms of reading comprehension questions as compared to far lower performance in the comprehension level of the BT. Thus, the stated studies based on discussions evidently clarified the significance of using BT in instructional design. The projected BT generated a significant quality level of questions based on some raw formless data content for the classification of tested class samples [54]. There exist various forms and formats specifying different levels of questions. To achieve the desired accuracy, a machine-learning approach provides and ensures effective high-performance classification for the extraction of feature-based classification [55].

Student performance and their cognitive engagement relate to one another in online education. Hence, residing in the relationship between conceptual knowledge and performance, there is a significant correlation. Performance can be mapped to cognitive state using BT and an intelligent system can be developed to scale student performance at respective levels iteratively by adjusting weights. Generative artificial intelligence promises to enhance effective learning experiences by mounting personalized support and learning materials by enabling timely feedback evolving versatile innovating assessment methods [56].

The levels of BT used to assess student learning state are Knowledge, Understanding, Application, Analysis, Synthesis, and Evaluation. These are the six intellectual levels

offered by BT, which advance from lower (easy) to higher (difficulty) levels. Each level has a distinct learning goal, and it is highly helpful for assessing the condition of the mind. Based on these levels, the system also evaluates cognitive and mental states.

The machine-learning model is used for the assessment of cognitive abilities through textual answers, which includes advanced machine-learning models that are needed for the complex task of evaluating an individual's mental capacity based on their written responses to questions. The Random Forest classifier is one such model, and it is extensively used in the field of NLP for text-classification problems. Before using the Random Forest (RF) classifier, features are extracted from textual answers using TF-IDF and Word2Vec.

The suggested research enhances the cognitive-evaluation process for an e-learning platform. Hence, it is quite tough to ask faculty in online classrooms to determine which learners are strong and which suffer from a conceptual understanding deficit. Through an intelligent parametric cognitive assessment of students using text analysis, this suggested study primarily enhances the method of cognitive assessment.

Additionally, it is difficult to analyze student cognitive states using virtual teaching-based evaluation methods, and an integrated system must be developed to assess student performance in a distance-learning-based environment by focusing on text semantics and analysis and mapping it to BT terms. Conceptual baseline knowledge is repeatedly extracted based on text-based answers utilizing NLP with baseline neural network-based algorithms. The primary goal of the proposed research is to examine the link between cognitive and academic success in the e-learning-based paradigm of students for integrated and precise evaluation of student cognitive ability in an episodic manner, considering the computational complexity of the assessment corpus.

Teachers need to forecast first and then analyze student performance to locate areas of weakness for timely improvement in academic standing. The concept of educational data-mining is used to coordinate computational tactics to improve workforce management and academic attainment [57]. The advent of the COVID-19 pandemic changed the education paradigm from offline- to online-based learning. This situation produced a crucial problem in the evaluation process of student-outcome-based learning. Effective evaluation is very difficult due to the lack of face-to-face interaction between students and teachers [31]. The revised bloom's taxonomy framework entails different verbs that assists in measuring course learning outcomes and individual competencies by creating customized evaluation rubrics. The recently advanced computing areas were working on the enhancements in the Bloom's Revised Taxonomy which comprises list of verbs [58]. The classification accuracy improvement based working has been concluded in the past relying just on the support vector machine but it lacks with the challenge of handling computational complexity in predicting the correct level of the tested population [59].

Researchers emphasize the need to have a unified solution to address high-performance computation power systems to operate in a controlled environment for achieving more precise measurements in terms of student cognitive parametric assessment in an e-learning environment.

3. Materials and Methods

The methodology section presents the architectural layout of the proposed research, which is mainly based on the online assessment of e-learning students. The reason for choosing distance learning is that students are more prone to the use of online computer systems for their studies as well as their assessments. This is compared to the rest of the traditional mode, where students are used to classifying CLO-based questions into discrete classes, namely Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation.

Object and action verb-based questions are considered for identifying the category of question from the pool. They act as inputs for model training to declare the cognitive level of the questions in line with the CLO.

Ponder the following example:

(1)　Outline the span and standing of artificial intelligence in everyday life applications.
(2)　Describe the elementary principles and concepts related to heuristics-based problems.

The abovementioned examples are associated with CLO statements, which belong to dual versatile levels. Example 1 is related to Understanding, and Example 2 is related to Knowledge. Nevertheless, if we examine both examples, it is observed that the examples contain identical action verbs/keywords referred to as "Define".

Examples are given as follows:

(1)　"How do you define transferable skills?"
(2)　"Define and elaborate the use of the Genetic Algorithm."

The abovementioned examples are the two projected questions belonging to two distinct classes. Example 1 belongs to Knowledge, and Example 2 fits into Analysis. However, if we investigate both question statements, it is observed that both contain the identical action verb referred to as "Define". The defined methodology is mainly divided into two modules, data pre-processing and data processing module. The defined modules are then followed up by the cognitive level prediction model (the pool-of-question level identification), which is based on RNN and students cognitive assessment model. A total of 300 students were tested on the pool of cognitive-level questions to test their cognitive level in Episode 1. A comparison of iterative episodes was used to study the learning improvement as per the CLO threshold, which was set at 70% (Good). The detail for each component is discussed in the following sections and subsections.

3.1. Data Acquisition

No standard public dataset is available associated with defined and aligned Course Learning Outcomes, which are tagged in the form of BT. The methodology contains various steps, which include sample question paper design with data sample collection, the preprocessing of data specifically for the BT-based level prediction, the design of an instruction set for recording student textual response analysis based on evaluation, and testing of the system. To train the classifier, sample artificial-based course questions are congregated from the previous research work and via web-scraping based on websites, such as Geeks for Geeks, Tutorial Point, etc. The question preprocessing is accomplished using the Natural-Language Tool Kit library. This determination creates a criterion to perform further experiments in the same problem domain. Moreover, a dataset of graduate-level regular students of artificial intelligence courses was collected using an online platform of existing virtual students. We used the stated dataset as a standard to evaluate our methodology as authors [45,58] and have also used the same type of dataset for classification into BT. Course learning outcomes were created based on the course report documents at the beginning of the semester. Hence, it assigns CLOs mapped with the questions that were part of the test item to be examined for the students. The designed system performs an alignment of course learning objectives into BT with the assurance that it is sufficient for declaring the maximum level of student learning.

Two datasets were considered for our research study in the form of episodes, where Table 2 represents the key statistics for both datasets. Dataset 1 is compiled and formulated based on RNN to intelligently predict the cognitive levels of the dataset based on random-weight assignment to action verbs. Generally, the comprehensive model clarifies the class-wise split for the stated Dataset 1. Conversely, Dataset 2 was compiled from the existing study captioned as Episode 2 data, and it was previously tagged into BT levels. Class-wise

distribution exhibits the datasets captioned as Dataset 1 and Dataset 2, where Dataset 1 is used to create a standard model for the proposed system and Dataset 2 is used to evaluate the projected system in contrast with the equivalent study on the mentioned dataset.

Table 2. Statistical dataset cast-off for this study.

Data Description	Information Source	Instances	Levels	Class Instances
Dataset for Episode 1 (300 students on test Bed)	Graduate-Level University Students (Virtual)-AI Domain	300 students × 20 Questions each level = 6000	6	Knowledge: 5000 *Assigned Weight: 15%* Comprehension: 5000 *Assigned Weight: 10%* Application: 5000 *Assigned Weight: 20%* Analysis: 5000 *Assigned Weight: 25%* Synthesis: 5000 *Assigned Weight: 10%* Evaluation: 5000 *Assigned Weight: 20%*
Dataset for Episode 2 (300 students on test Bed)	Graduate-Level University Students (Virtual)-AI Domain	300 students × 20 Questions each level = 6000	6	Knowledge: 5000 *Assigned Weight: 15%* Comprehension: 5000 *Assigned Weight: 10%* Application: 5000 *Assigned Weight: 20%* Analysis: 5000 *Assigned Weight: 25%* Synthesis: 5000 *Assigned Weight: 10%* Evaluation: 5000 *Assigned Weight: 20%*

As stated and projected, with the class instances with the weight assignment in Table 2 and graphically represented as per Figure 1, it is imperative to note that the cognitive-level question sets are set to 5000 per level. The weight assignment for ensuring uniformity and the threshold was set based on CLOs defined for the artificial intelligence domain course to gauge students for proper mapping/back-tracking of performance. The domain knowledge is based on the concept one should have, which is why it is allocated to 15% as compared to 25%, which is allocated to Synthesis, for testing student levels of creativity to develop any solution based on the acquired knowledge, as shown in the pie chart in Figure 1.
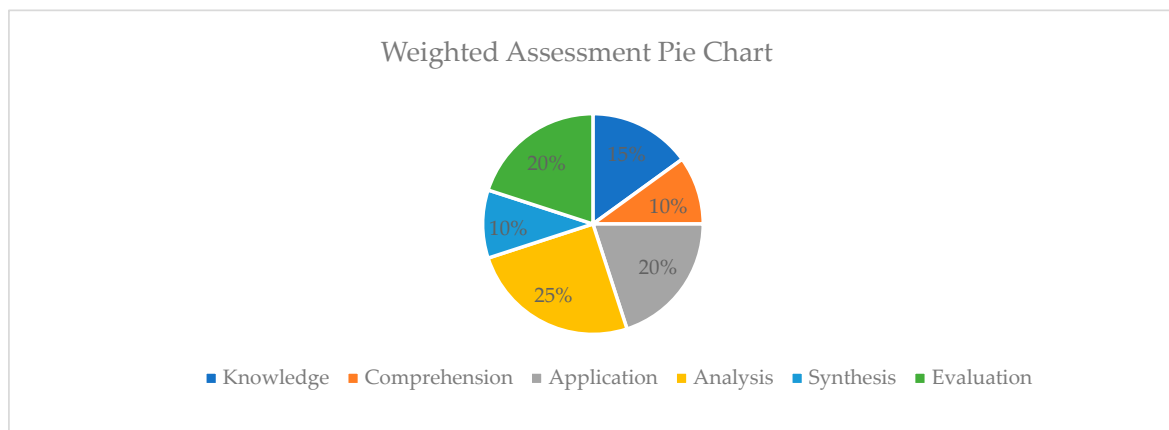
Figure 1. Weighted class-wise dataset distribution for Episode 1 and Episode 2.

3.2. Domain Understanding/Question Pool Development and Selection

The data pool of questions was gathered for a better understanding of the system and testing. The data are collected from various web scrapers, such as GeeksforGeeks, Guru99, Tutorials Point (best for aptitude tests), etc. The major purpose of conducting this test-item development is for different methods to be adopted for the classification of CLOs and examination questions in BT. The total number of participants was 300. We analyzed the responses against each question asked by the applicants once the pool inputs were retrieved by Module 1 (question-level prediction using RNN in Python 3.11 (64-bit)). The concluded major aspects of the pool of questions developed and predicted for Bloom's Taxonomy level are as follows:

The classification of questions into BT is based on an understanding of the domain.

(1) It is considered to be a crucial activity not only for the assessment of course quality but also to ensure the quality of the examination paper, which is needed to measure student learning based on outcome and vice versa.

(2) CLO-based question statement holds Bloom's keyword/action verb that overlaps different levels from neighboring words. These are examined at diverse levels. It is the words that are recognized through the family of words and for which the word-embedding library of Python is deployed to ensure the integrity of action verbs. Analyzing the results of 36,000 records from an RNN (recurrent neural network)-based assessment involves several steps, as follows:

  1. Data Preparation: This includes preprocessing the data, which may involve cleaning, normalizing, and encoding it into a format appropriate for input into the RNN.
  2. Model Training: Training the RNN model applied on a fragment of the data by creating epochs. Epochs involve feeding the input sequences into the network for computing the output predictions and comparing them with the actual targets by adjusting the network weights using procedures like backpropagation and gradient descent.
  3. Validation: Evaluating the trained model on a split fragment of the data. The split of training to test data on the ratios of 70% to 30% to generalize the model performance and to detect the overfitting.
  4. Assessment Metrics: Calculating various metrics for assessing the performance of the model at both episode levels, such as Accuracy, Precision, Recall, F1-Score (a predictive performance measure based on true positive, false positive, true negative, and false negative values), etc., depending on the assessment nature (classification, regression, etc.).
  5. Analysis of Results: This involves interpreting assessment metrics to draw conclusions about the effectiveness of the model. For example, if it is a classification task, one might analyze confusion matrices to see which classes of students/groups are struggling to achieve the parametric threshold of CLO, which was and will be set at 70–90% approx.
  6. Iterative Improvement: Based on the analysis, one might iterate on the model architecture (hyperparameters like learning rate with batch size and number of epochs to optimize training for your dataset), or data-preprocessing steps are taken to expand the performance with respect to precision and accuracy of the designed system.

(3) Question-pool training holds the task of a question pool established for students to be tested. It is prepared through the division of training and testing data operating on the layers of LSTM for batch normalization for testing different cognitive levels of

students, which are refined and then projected to students for ultimate assessment in an iterative fashion. The long short-term memory network is basically a kind of recurrent neural network, which is usually cast off for performing a sequence of modeling tasks. In our research case study, we are exercising episodic-based assessment to chain the difference between the previous and the new findings, which are used as batch normalization for feedback. This can be adapted to LSTMs to stabilize and accelerate the training process. Detailed below is the explanation of how batch normalization works in the context of LSTM networks. An LSTM cell has several gates (input, forget, output) with a memory cell that helps in maintaining information over long sequences. The basic equations governing an LSTM cell are mentioned in Figure 1, generated through Spyder IDE in Python. The flowchart of a neural network architecture is shown in Figure 2, where LSTM layers are combined with batch normalization. Here is a breakdown of how batch normalization is applied in this LSTM-based model, detailed by architectural overview.
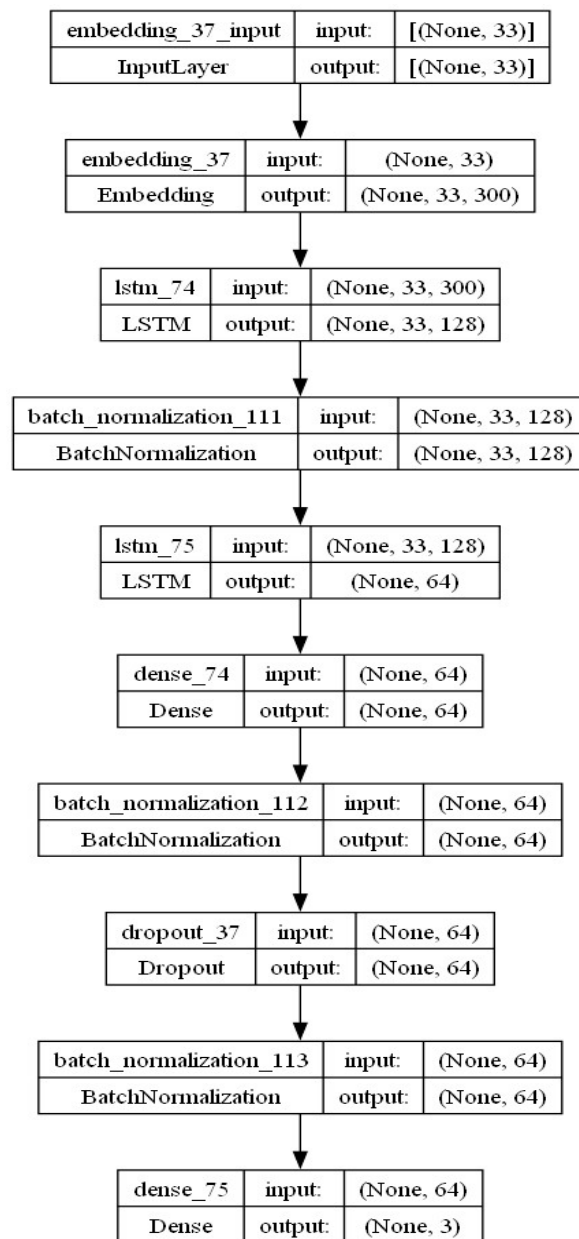
Figure 2. Iterative learning-based modular question-pool classification (word embedding—Python) epochs 70/30% training/testing.

3.3. Architecture Overview

The architectural outlay is based on certain phases and steps to follow in transforming intelligent system development. The Python-generated output weight and epoch iterative flow are given as shown below and graphically shown in Figure 2.

1. Input Layer: The input model has a shape of (None, 33), where None represents the batch size and 33 is the number of features.
2. Embedding Layer: This layer transforms the input into a higher-dimensional space, with the output shape (None, 33, 300) indicating 300-dimensional embedding for each input feature.
3. First LSTM Layer: The LSTM processes the embedded input, producing an output with the shape (None, 33, 128), meaning 128-dimensional output vectors for each time step.
4. Batch Normalization: After the first LSTM, batch normalization is applied to the LSTM output. This operation normalizes the output of LSTM across the batch, helping to stabilize and speed up training. The output shape remains (None, 33, 128).
5. Second LSTM Layer: The normalized output is passed through a second LSTM layer, reducing the dimensionality to (None, 64).
6. Dense Layer: A dense layer is applied, producing an output of shape (None, 64).
7. Second Batch Normalization: Batch normalization is applied again after the dense layer, normalizing the output across the whole batch.
8. Dropout Layer: A dropout layer develops with the same output shape (None, 64). Dropout is alleviated by preventing and overfitting in randomly setting a fraction of input units to value "0" through the training process.
9. Third Batch Normalization: Another batch-normalization layer is applied, further stabilizing the output.
10. Final Dense Layer: Finally, the model inputs a layer with 3 units, typically representing a classification into classes, as follows:

    (1) Word-based vectors by means of skip-gram-built word embedding for representing text in the form of numeric features after processing.
    (2) Lastly, the use of a BT-based level classifier is deployed for category-based questions into one of the predefined types.

Suppose raw CLO/question text is recorded in the following developed model: "Elaborate the concept of the Genetic Algorithm in artificial intelligence". For this stated text, the generated query set is given in Figure 2. It gives the high-level flow of the methodology in terms of episodes.

Query set (Prep, Embed, Level) ["Elaborate the concept of the Genetic Algorithm in artificial intelligence"), Knowledge], where Prep is Preprocessed Text, Embed is Word Vectors, and Level is BT level. The next few sections describe the construction and workings of each of these components of the proposed system.

CBOW and skip-gram are used to generate word embeddings for NLP tasks, such as semantic analysis, sentiment analysis, and machine translation in the following tasks.

Suppose the following raw question text as an input: "Elaborate the concept of the Genetic Algorithm in artificial intelligence". For this text, the generated query set is given as follows:

As per Figure 3, it gives the high-level flow of the methodology in terms of episodes. (Prep, Embed, Level) ["Elaborate the concept of the Genetic Algorithm in artificial intelligence), Knowledge].
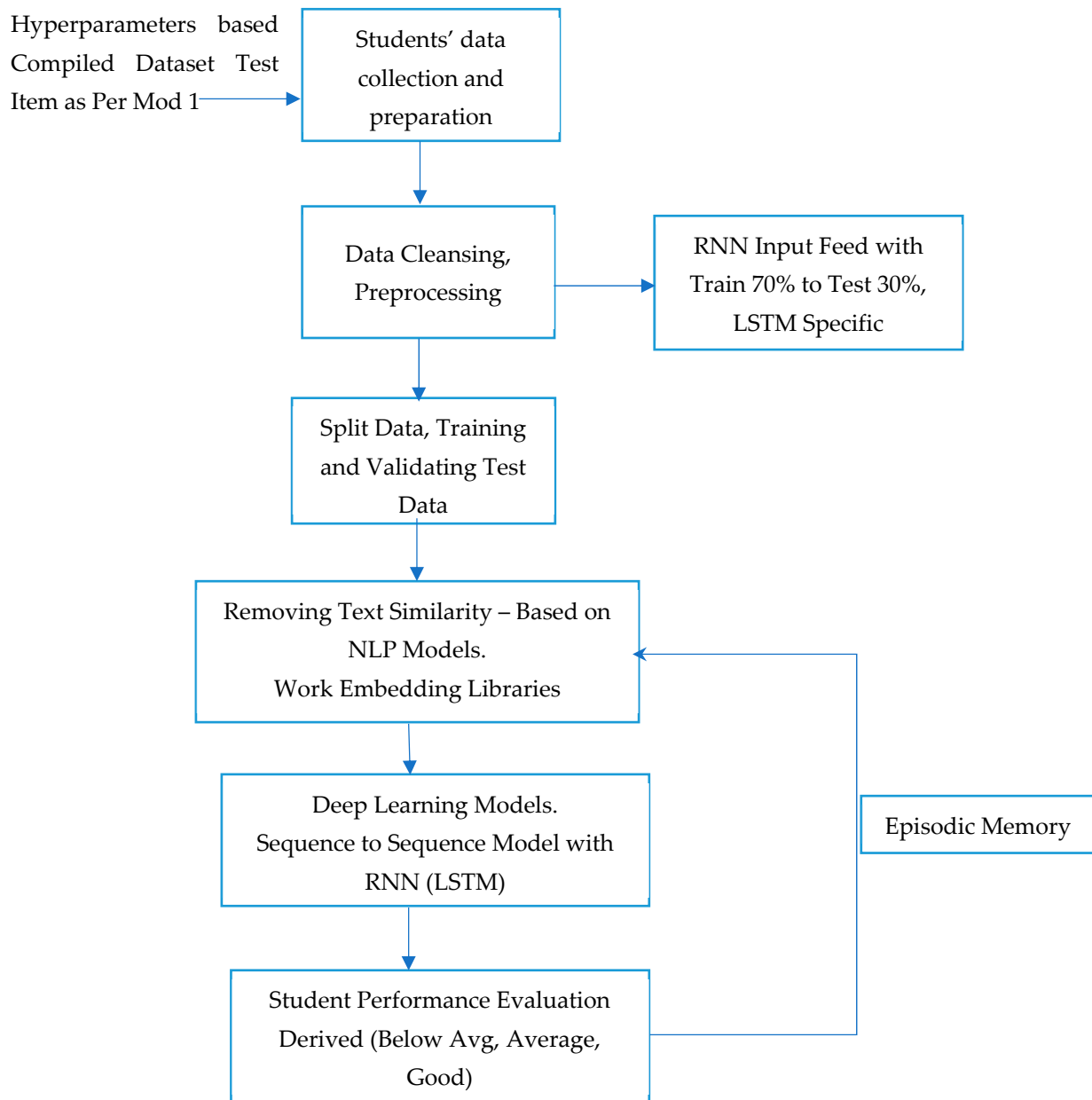
Figure 3. High-level flow of the methodology (Episode n).

Prep represents Preprocessed Text, Embed Word is based on Vectors, and Level gives the BT level. In the next sections, the assembly and workings of these components are described.

Each epoch comprises numerous iterations, where the parameters of the models keep on updating with a gradient of the associated loss function with respect to parameters. The algorithm, GloVe, is an unsupervised-learning algorithm used for acquiring vector representations for words. The vector embedding captures the semantic relationship between contextual words based on statistics across the entire corpus used for training. The GloVe model is basically trained on the global word-to-word corpus with a co-occurrence-based matrix, which essentially captures the frequency and occurrence of words in a corpus.

3.4. Proposed System Overview

The proposed system abstract level model is shown in Figure 3. The projected system takes raw question text as input to classify one of the BT levels in the cognitive domain. The proposed system executes the aforementioned tasks:

(1)   The initial step is based on text preprocessing with data cleaning, which takes an input in the form of text and then preprocesses it by altering it into baseline lower-case letters, removing stop words with punctuation marks, and then converting those words to their root words by means of lemmatization.

(2)   The next step, after text preprocessing, is to compute and calculate numeric-based word vectors on skip-gram word embedding to represent text into numeric-based feature selection.

(3)   Lastly, the BT level classifies the text into pre-distinct categories.

(4)   Text preprocessing with cleaning is the initial step, which receives the input text and then preprocesses it by adapting that given text into lower-case form by removing those stop words with the punctuation marks and then by adapting all the words in the mentioned root words in the form of lemmatization.

(5)   After the completion of text preprocessing, the succeeding step is to calculate numeric word-based vectors by means of word embedding based on skip-gram to represent text in the form of numeric features after processing.

(6)   Lastly, the BT classifier is deployed to categorize questions into predefined types.

3.5. Construction of the Proposed System

The proposed methodology portrays not only the construction but also the integration of three components implicated in the proposed system. The proposed system involves three components starting from data preprocessing, data cleaning, learning word representation using word embedding and BT classifier demonstrated in Figure 3. The detail of every module is discussed in the sections below, captioned as Episode 1. Modeling is also elaborated in terms of the statistical dataset of the study.

3.5.1. Data Preprocessing and Cleaning

The primary aspect of this phase is to measure and analyze the impact of the outcome of the research study in terms of generating impactful classification results [32]. Therefore, preprocessing techniques were applied from the collected, refined datasets to eliminate non-informative features, which are extracted from the data source. In the preprocessing phase, text-based data are precisely converted into lower-case form from the actual data. In calculation, the punctuation and stop-word removal has been completed using regular expressions along with some random pattern-matching techniques. A tokenization process is also performed, which is based on white space and WordNet-based lemmatization for baseline preprocessed text. In the tokenization stage, question text is transformed into versatile tokens/words. Then, these same words are converted into their root forms using WordNet lemmatization. The pseudocode for data preprocessing is demonstrated in Algorithm 1, which shows a sequence of numerous preprocessing steps operating on raw datasets for cleaning datasets as an output. Recurrent Neural Networks (RNNs) are in Excel-based data format, which is used for processing the sequential and recurrent data and by making it appropriate for tasks, which are used in natural-language processing based on time-series analysis. Therefore, the capability to recall previous inputs is considered to be an apparent advantage for short- to medium-range sequencing. Figures 4 and 5 exemplify the pertaining preprocessing steps based on predefined CLO and question-based data, respectively.

| **Algorithm 1.** Pseudocode of Data Preprocessing, Cleaning Algorithm |
|---|
| Input Data: Sample Query Text |

Output Data: CLO-Based Preprocessed form of Data
   Sentences extract Sentences as (INPUT)
   Prep Input = Vacant
   while sent sentence structure do
   stSentence. remove StopWords(sent) PSentenceremove Punctuation(stSentencebased) words
extract-Words-from(pSentence) PrepSentence structure
   Empty (default empty bucket)
while words do
l for word lower(word)
   Lemma Word Word.NetLemmatizer(lword) prep Sentence lemma word prep Sentence-space
   End while
Prep Input. append state (prep Sentence)
end while
Output Result ← prep Input (by tokenizing keywords and then formulating them into meaningful/useful sentences)

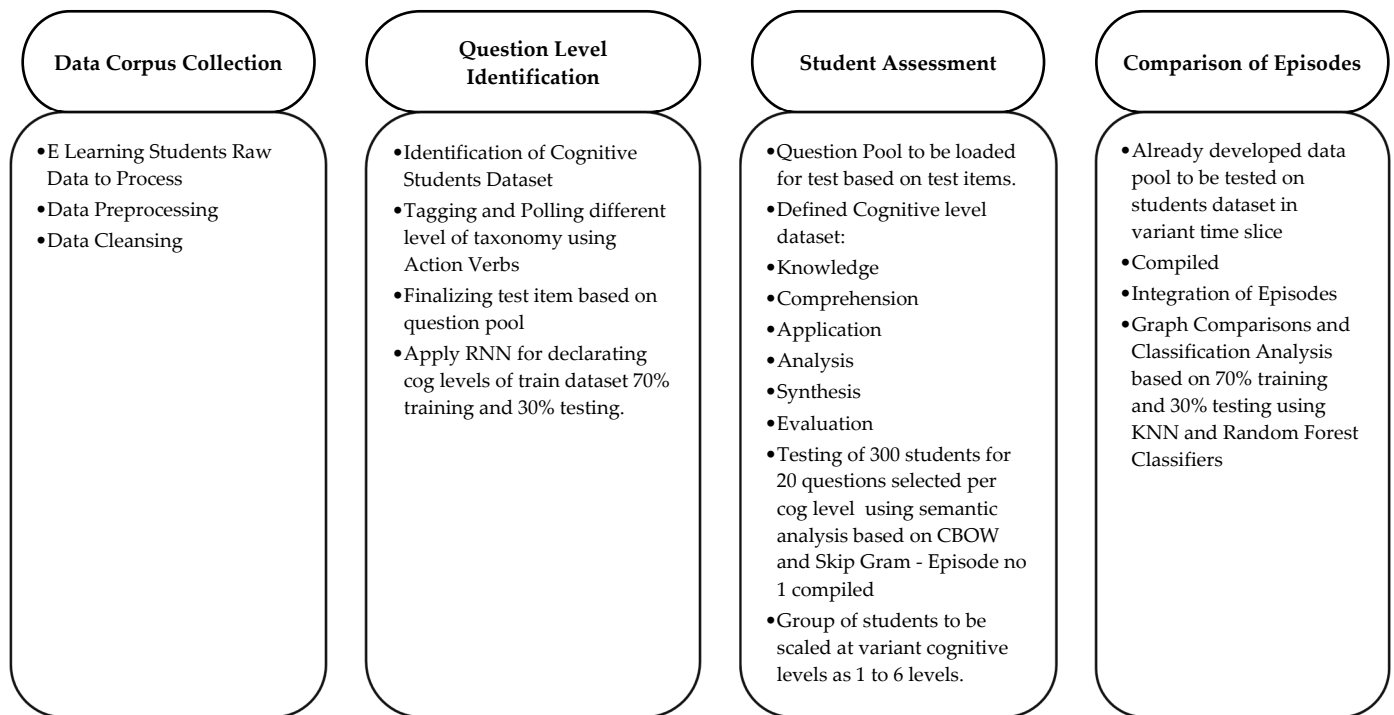### 3.5.2. Data Preparation with Splitting

The data preparation with splitting, starting from raw data to refine and classify text as shown in the flow diagram, is shown in Figure 4 as follows:
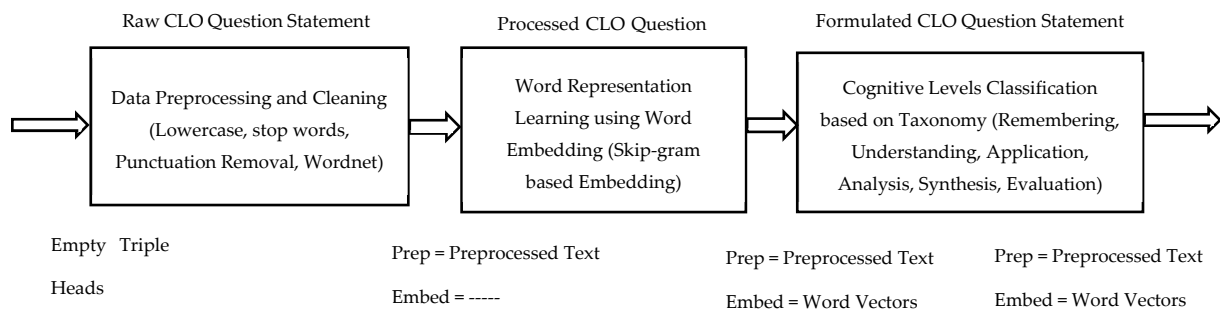
Figure 4. Proposed system overview.

The demonstration of the assessment chained modules is shown in Figure 5, which gives a clear integral solution of the effective assessment solution by chaining the corpus collection with the state-of-the-art episodic phenomenon.

After completing data preprocessing with cleaning, the follow-up step in the proposed system is basically preparation and then the splitting of data for model structuring along with evaluation. Algorithm 2 elucidates the series of steps that were performed on Dataset 1, targeted at Episode 1, and Dataset 2, targeted as Episode 2, for preparing training and test datasets for model construction and evaluation, as shown in Table 2 with proper demonstration.

Data Corpus Collection	Question Level Identification	Student Assessment	Comparison of Episodes
• E Learning Students Raw Data to Process • Data Preprocessing • Data Cleansing	• Identification of Cognitive Students Dataset • Tagging and Polling different level of taxonomy using Action Verbs • Finalizing test item based on question pool • Apply RNN for declarating cog levels of train dataset 70% training and 30% testing.	• Question Pool to be loaded for test based on test items. • Defined Cognitive level dataset: • Knowledge • Comprehension • Application • Analysis • Synthesis • Evaluation • Testing of 300 students for 20 questions selected per cog level using semantic analysis based on CBOW and Skip Gram - Episode no 1 compiled • Group of students to be scaled at variant cognitive levels as 1 to 6 levels.	• Already developed data pool to be tested on students dataset in variant time slice • Compiled • Integration of Episodes • Graph Comparisons and Classification Analysis based on 70% training and 30% testing using KNN and Random Forest Classifiers

Figure 5. Module-level preparation for an integrated assessment system.

Algorithm 2. Data Splitting with Data Preparation

INPUT: Preprocessed Input Text based purely on the Class Labels
OUTPUT: Training based on Test Data Partition
Unique Words ← unique Words ← max Length ← max Len-Test
   Size ← Test-Size
   Text ← Based sequencing (Preprocessed Input Text)
   Tags Mark ← Basic coding (Class Mark Marker)
     Data Markers ← shuffle (Text, Markers)
     Train Data = Partition (Data, Relation Ratio = 1 − Test-Size)
     Test Data = Partition (Data, Relation Ratio = Test-Size)

3.5.3. Development of BT Level Classifier

The final preparation of data based on the preprocessed phase takes input in the form of embedding, and the next step is basically to construct a baseline classification model to classify the input data split into different levels of BT. LSTM has been selected for its potential and credibility in organizing the arrangements and text for considering a classic example of sequencing. The tagged dataset is chosen. Then, the questions are tagged for chosen BT-based cognitive levels. Therefore, the proposed classification model determines and classifies questions based on desired categories based on the LSTM network.

The long-term dependencies problem is based on understanding the sequential context-based data relying on the RNN feed. The stated problem is referred to as a "Gradient Eliminator". There are different cases based on successive data episodes, which require longer sequences to understand the context effectively. The learning and data sources exist in the context of identifying the individual student level for calculating and storing the assessment results. The achievement of the model is well elaborated, which contains the contents as a learning resource, which is a pivotal part followed by the quality assessment of the same delivered content. The feedback of the assessment is then stored in a data module in a conceptual/logical aspect. The student knowledge level is based on knowledge tracing to reflect the paradox of continual metric-based learning in an e-learning environment.

3.5.4. Word Representation with Word-Embedding-Based Learning

The key feature used in the proposed system is semantic text representation utilizing the word-embedding technique. The prepared datasets are trivial, and with the trivial datasets used in deep learning, the pretrained word embedding makes it difficult to identify and classify the semantics associated with it. Therefore, we used pretrained embedding as a proper justification for effectual word representations of our planned datasets. The word embedding was solely selected for pretrained word embedding, just like "Wiki Word Vectors". This type of embedding was trained based on Wikipedia-obtained text-based scripts. The researchers have already explained this embedding in detail with predefined thresholds [50]. This embedding is completed for the task because it is projected, and it helps in obtaining semantic-based judgments of words due to the abovementioned reasons:

✓ Word-based embedding is prepared for training on Wikipedia-provided text utilizing techniques for word representation generation relying on its neighboring words.

✓ The dataset consists of the maximum number of words from the Wikipedia-provided corpus.

Long short-term memory is a kind of recurrent neural network (RNN) architecture, which is suggested to overcome the vanishing gradient problem, which arises at some point in time in traditional RNNs.

Additionally, pretrained model embedding is included, as in GloVe.6B.100D with GoogleNews-vectors-negative300, to evaluate and choose better performance from the stated tasks. The connection between the required context word and the predicted word varies with the time and complexity of the input sequence. Practically, as per the historical data, the RNN feed is not efficient enough to solve these complex cases [34,41].

LSTM neural networks are basically a special kind of RNN network that can perform based on long-term sequencing dependencies. The networks are specially designed to retain learned information for a longer time. It can make decisions regarding the effectiveness of information while processing input. It also has a gated mechanism to control the flow of input-based sequences inside the LSTM cell. Before going into the details of the Gating mechanism used in LSTM, there is a need to develop and understand the proposed sequential neural network model. The processed input sequence for the representation of weights as w1, w2, and wn is highly demonstrated in Figure 6.

The architecture of the Proposed Research Methodology is shown in Figure 5. It demonstrates and details the integrated assessment system comprehensive onset of research methodology implementation from the data-preprocessing module to the student cognitive assessment model, which relies on the subset of the text-processing module and the test-item pool developed for intelligent identification of correct question level by optimizing the hyperparameters of the RNN algorithm. A questions-pool-based dataset is based on six levels comprised of action verbs that are used for the test-item-based assessment of e-learning mode students. The cognitive levels associated with the assessment of students include Knowledge, Remembering, Comprehension, Application, Analysis, Synthesis, and Evaluation, which intelligently assess the question level to be tested on student cognitive ability in an episodic manner to identify any fluctuation in student performance, once the levels and the results are known to them. The detailed procedural work of each module is described already in Section 3, which is followed by the experimental setup in Section 4. The episodic session-based test assessment is the unique attribute of the research. Hence, there is a randomness to the learning scale of e-learning-based students for achieving continual graded assessment by applying class labels using a Random Forest classifier in deriving the scale of Below Average, Average, and Good Students. Through the procedural workings in the form of episodic iterations, student learning curves are improved. The designed model is compared with NB, KNN, Cosine Similarity, and benchmarked RNN,

and it provided fruitful results with RNN to handle an episodic improvement-gauged dataset. It shows significant performance measures of students in the AI domain.
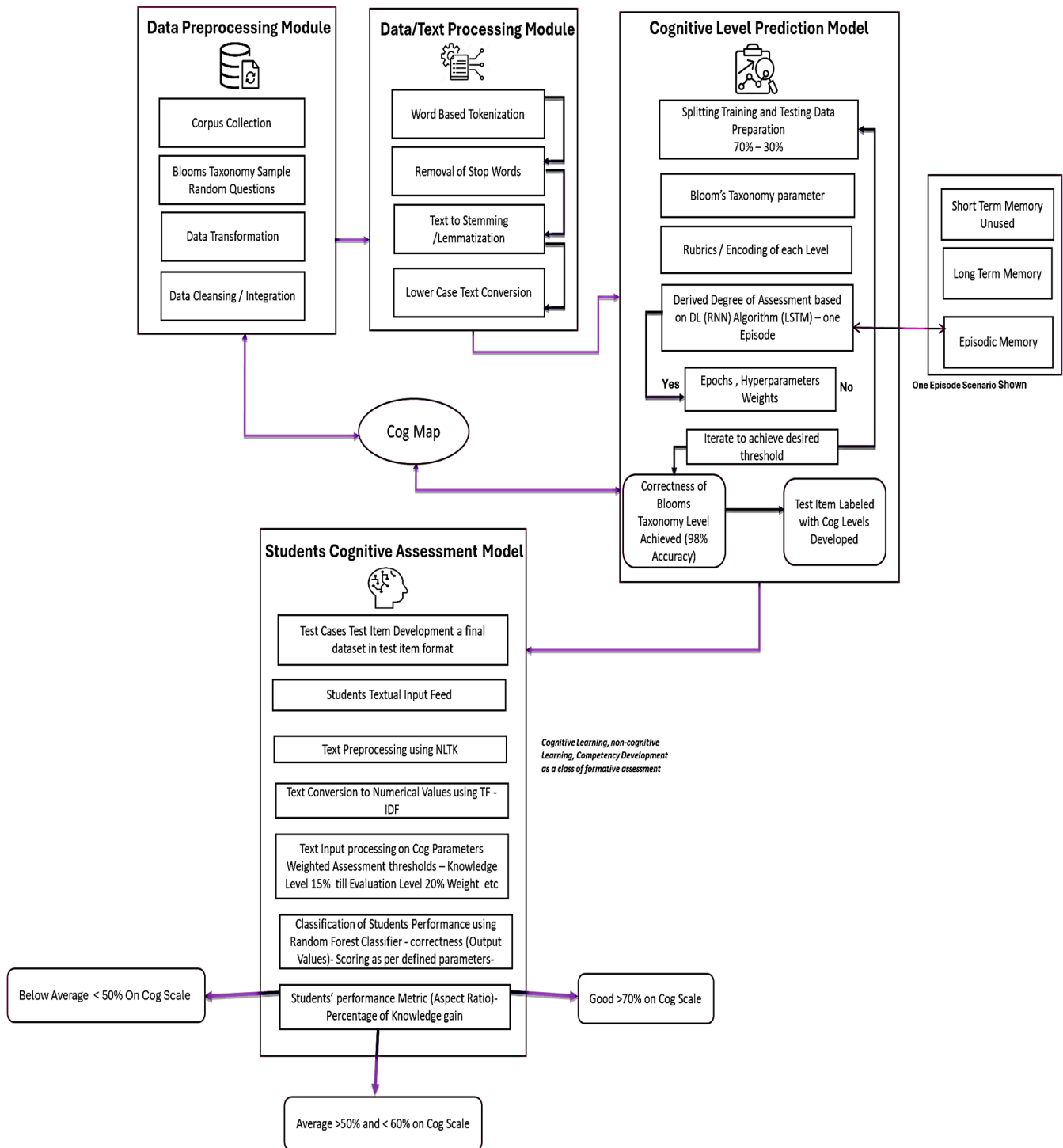
Figure 6. Architecture of proposed research methodology—Integrated assessment system.

The projected method reports the main issue of overlapping the keywords based on taxonomy-level identification. The enhanced model based on classification is an integral method extracted from literature, which shows significant developments in intelligently solving assessment problems. The major impact is towards applying the iterative assessment sessions for the identification of the cognitive scale of students. There exists an approach in historical research that influences logged student data and combines it with

machine learning (ML) and natural-language processing (NLP) methods to attain effective results. The scheme basically utilizes sentence-based semantics by representing student responses to open-ended questions [38].

4. Experimental Setup

The dataset of 300 artificial intelligence courses enrolled students at distance-learning virtual institutes has been gathered after processing threshold-based testing. BT-based questions were developed, and a pool of questions was tested on the test bed or RNN algorithm using word embedding to derive the levels, which were planned to be tested for student parametric assessment. The correctness of levels identified is concluded at 95% as per the accuracy level achieved using Python libraries. Every course CLO varies as per the requirement, importance, and threshold of any course and requires to assignment of weights at each cognitive level based on the specified threshold as allocated in Table 1.

There exist several related libraries that are used for NLP-based text evaluation performance, such as Python version 3.11, Pandas, Beautiful Soup, Natural-Language Toolkit (NLTK) with NumPy, Scikit-learn, GenSim, Matplotlib, and Seaborn. The question bank comprises a pool of questions that have been tested on student groups to assess their cognitive levels considering assigned weights of assessment without declaring the levels of questions posted for them to solve and submit. The data feed to students is split into 20 questions per level per student to be labeled as Episode 1 for each student. Similarly, for all 300 students, the same procedure was adopted and followed up to formulate the assessment process in an e-learning-based environment. We assume the student feed to be better for unknown random questions as compared with the known random questions. For Episode 2, the same process was adopted but with different question sets of the same identified levels: Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation level. The analysis was carried out not only at each student's level but also the aggregate performance at each level, as well as each student performance level and at composite Bloom's Taxonomy-based text-based question–answer level to generalize, define, and declare each group of students with the levels they achieved after two episodic iterations of the testing element. This helps not only in identifying their learning curve on the scale of below average to good but also in highlighting the content to deliver to students to improve their feed and threshold of their learning levels. The formulated steps for question-level identification were as follows:

Student assessment of the pool of the dataset derived from Step 1—300 students with 20 questions at each level to sum up to 120 questions per student at accumulated levels. The integrated dataset of 300 students with 120 questions is defined and declared as a whole record dataset assessment of 36,000 entities to be evaluated and aggregated. This actual data training and testing were done on the question-level assessment part, which was Step 1 of the whole procedure, also known as Module 1. Module 2 and Module 3 are student cognitive assessments with a comparison of both episodes for generalizing if any improvement appears for the student groups. This helps in reshaping our system. Table 3 shows the aggregate-level assessment report of the first 20 students and the last 25 students among the total of 300 students on the test bed for gauging the cognitive assessment of students projected with a dynamic pool of cognitive-based questions in a specified time slice. The result generated and shown is Episode 1 data, which leads to some fluctuation or improvement in Episode 2 of the same group of students, referred to as followed-up testing. Table 3 details the student assessment on various scales of taxonomy, having achieved the maximum threshold at each level defined as 76% on the first four levels: Knowledge, Comprehension, Application, and Analysis. The synthesis threshold is depicted at 77%, and finally, the evaluation level shows the maximum threshold achieved by students at 94%.

Table 3. Students aggregate assessment level reports on all cognitive levels (Episode 1).

Sr#	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
1	71	34	50	48	45	43
2	68	54	74	68	56	56
3	67	41	70	48	67	44
4	62	52	60	75	76	29
5	58	74	61	63	45	65
6	59	71	62	48	34	56
7	76	72	49	38	74	67
8	71	74	67	48	35	74
9	74	35	75	68	31	48
10	49	75	75	76	44	73
11	69	75	52	68	46	74
12	58	38	54	49	47	47
13	65	76	74	38	39	65
14	75	56	68	48	41	51
15	44	46	74	68	42	48
16	67	69	71	75	44	71
17	71	49	72	68	76	54
18	74	74	56	48	49	51
19	75	67	76	74	74	61
20	64	48	58	67	46	56
21	54	72	28	49	45	65
22	75	75	29	71	59	71
23	38	73	74	64	73	72
24	41	74	46	76	14	49
25	48	74	64	48	67	69
⋮	⋮	⋮	⋮	⋮	⋮	⋮
275	54	45	68	76	19	67
276	76	45	58	68	56	46
277	65	75	77	68	57	76
278	65	64	46	76	66	66
279	49	54	45	48	55	70
280	73	74	19	74	56	45
281	29	39	73	76	74	67
282	74	72	72	76	76	76
283	46	75	69	74	46	45
284	65	68	71	19	45	67
285	47	49	75	75	47	74
286	75	74	76	71	57	54
287	64	68	73	68	46	67
288	59	69	73	46	76	76
289	75	71	19	71	58	67
290	41	29	49	74	41	73
291	28	68	69	72	28	65
292	37	69	71	68	39	45
293	39	49	74	75	56	68
294	44	67	69	64	44	94
295	48	49	58	64	49	57
296	49	58	58	54	53	76
297	52	68	71	68	51	71
298	57	74	75	67	54	72
299	61	67	76	74	65	54
300	74	61	77	71	55	71

Continuing the variant pattern to reach out to the 300th student, the aggregate percentile of the cognitive scale of assessment is given with the last 275-student chunk.

This helps in reshaping our system for better improvement of students in defining and redefining their cognitive abilities and scales for improvement in their overall learning

as far as their course learning objectives are concerned, which were set at 60% to 70% as students achieved the CLO and falls below that threshold as "needs improvement" on the basis of these cognitive levels. The whole system not only routes the cause intelligently but also helps to ensure the completeness, correctness, and coherence of students in a virtual testing environment. The summarized steps are given as follows: for text analysis in Python, certain libraries are utilized, such as NLTK or SpaCy, for tasks like stop-word removal. To implement word embedding based on recurrent neural networks (RNN), the cognitive level of students was identified by following these steps:

- Preprocessing:
    - Remove stop words using NLTK or SpaCy.
    - Tokenize the text data and convert it into sequences.
- Word Embedding with RNN:
    - Use deep-learning libraries like TensorFlow or PyTorch to create an RNN model.
    - Implement an embedding layer to convert words into numeric vectors.
    - Construct an RNN model (LSTM or GRU) that takes these embeddings as input.
- Identifying Cognitive Levels:
    - Train the RNN model on a dataset labeled with different cognitive levels.
    - Use the trained model to predict the cognitive level of new text inputs.
    - The process mentioned above is considered to be a single episode followed up by Episode 2, repeating the same procedure to effectively access 300 students in the same defined parameters associated with Bloom's Taxonomy assigned with a pre-allocated percentile of assessment. The student's cognitive assessment model is based on the text-input feed with rule-based text processing based on completeness and correctness. The results are classified using a Random Forest classifier for predicting the level of correctness achieved by the students based on a defined assessment threshold as follows: Below Average <50%, Average 50%–60%, Good >60%.
    - By tuning the hyperparameters (hyperparameters like learning rate with batch size and the number of epochs to optimize training for your dataset), LSTM in RNN obtains the optimized and refined assessment results with over 90% accuracy as compared to previous models on a standard dataset.

5. Results and Discussion

The results and discussion section presents the two main phases as part of the methodology section and follows up in line with the developed system. The first phase is to channelize the semantics of test-item development and intelligently label the cognitive levels of the questions on a test bed. The second phase is the pure assessment of students on cognitive levels threshold defined as CLOs, which are performance based on what either one student or a group of students achieves in a specified range of assessments. The hyperparameters are tuned for LSTMs, which are widely used for sequence-prediction tasks like natural-language processing. The results are also divided into two sections, which are as follows: continued variant pattern to reach out to the 300-student aggregate percentile of the cognitive scale of assessment, which is given with the last 25-student chunk. The aggregate level of performance of 300 students is given in Figure 7, which spares the clear demarcation between student performance mainly at the Application and Analysis levels. These are the levels that demand students' own conceptual understanding of the domain-based questions. The margin of achieving completeness and correctness arises in this frame of reference as compared to the rest of the cognitive levels under consideration.
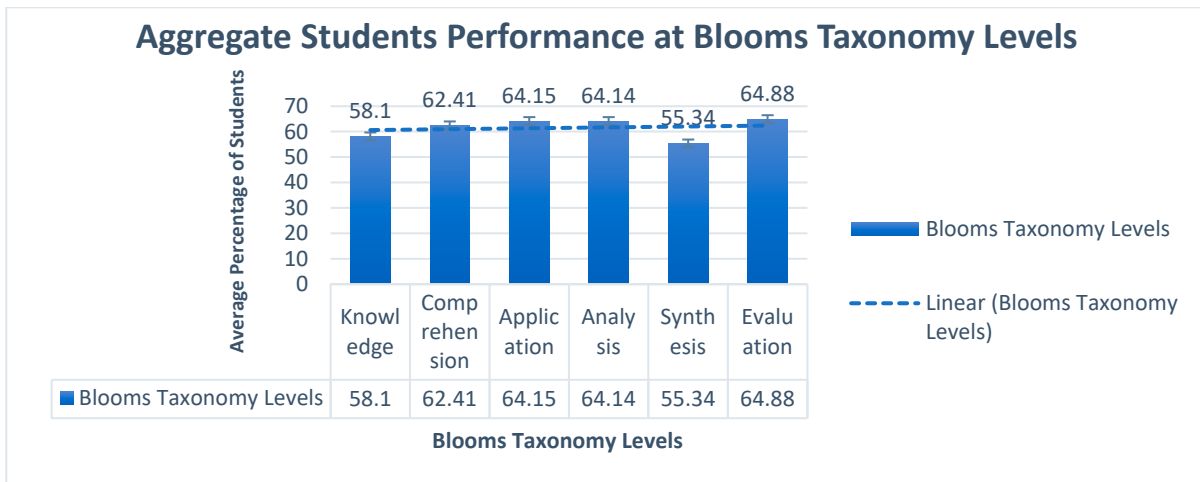
Figure 7. Aggregate student performance at Bloom's Taxonomy levels.

It is, hence, not possible to draft a whole dataset of 300 students under the assessment bed for both tables and graphs. However, the two students are shown in Table 4 as follows:

Table 4. Performance of students at each level—percentages (selected).

Student Ident	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
1	71	34	50	48	45	43
50	39	48	58	71	75	64

As shown in Figure 8, the chosen dataset of Student 1 is projected based on the results obtained from word-embedding Python. Students performed much better at the Knowledge level as compared to the rest of the levels and needed to work on the Comprehension level as per Episode 1.
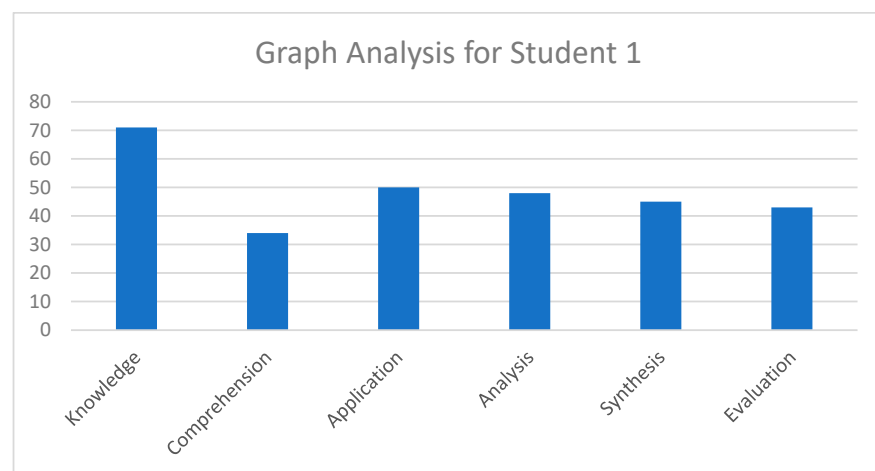
Figure 8. Assessment-based performance level—Student 1.

As shown below in Figure 9, the chosen dataset of Student 50 is projected based on the results obtained from word-embedding Python. The students performed much better at the Analysis and Synthesis levels as compared to the rest of the levels, and needed to work on the Knowledge level, as per Episode 1.
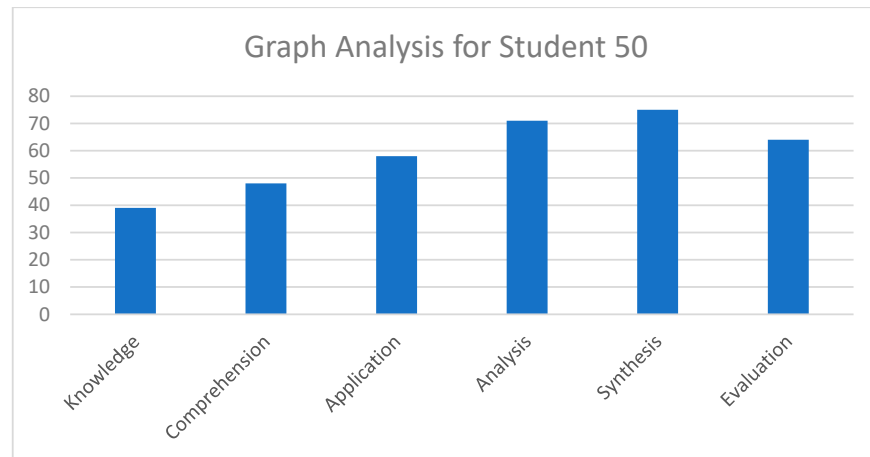
Figure 9. Assessment-based performance level—Student 50.

5.1. Comparison of Episode 1 and Episode 2

Episode 1 covers the cognitive assessment based on the anatomy of student performance for achieving the level of threshold. As per Figure 10A, the group of students performed well at the Evaluation level as compared to the Synthesis level, with 55.34% achieved by the average student. The percentage of student cognitive measurement performance increases at the Evaluation level to 80.87%, in line with the Application and Analysis level, which is at 80.18% to 80.17%, respectively. The conclusions based on this episodic feedback aimed to provide students with ample opportunity and learning to complete their studies after experiencing their performance in the first iteration of the methodological test bed.
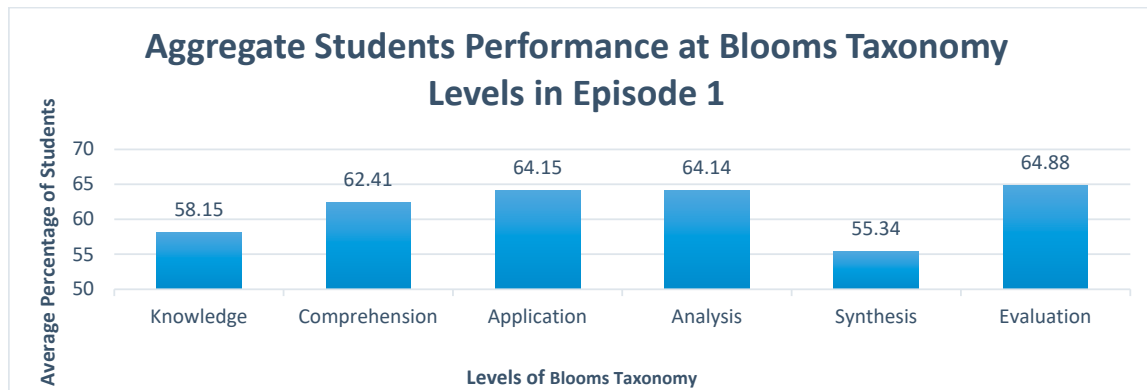
(A)
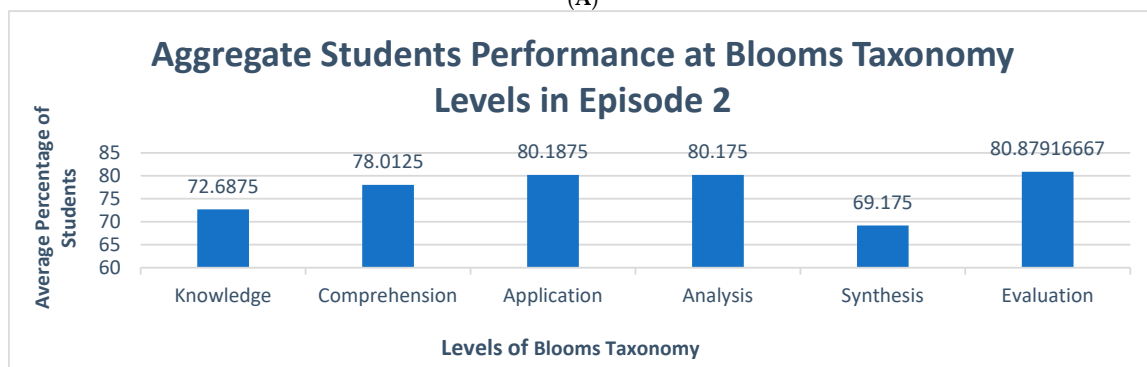
(B)

Figure 10. (**A**) Aggregate student performance at Bloom's Taxonomy levels in Episode 1. (**B**) Aggregate student performance at Bloom's Taxonomy levels in Episode 2.

As per Figure 10B, the group of students performed very well at the Evaluation level with an improved aggregate as compared to the Synthesis level, which showed 80.87% points achieved by the average students. The conclusions based on this episodic feedback aimed to provide students with ample opportunity and results in vital improvement, as shown in the graph, Figure 10B. The Synthesis level also shows improvement with 69.17% as compared with the Episode 1 assessment results, which are lagging behind at 55.34%.

5.2. Random Forest Classifier Prediction

Students at the Synthesis level did not perform well and hence achieved Below Average percentiles at the Synthesis and Evaluation levels, as shown in Table 5. After the base-level pool-of-question-level prediction, the numeric-based features of text-based responses from Word2Vec and TF-IDF are considered a gradient input for a specific instruction set, and the machine-learning model is tested based on these words. The identified words are recurrent in a text but uncommon in the student textual sentence-based answer feed, which we assumed to have rich informational value in TF-IDF. By this method, we assign a word with a weighted score based on the frequency that is shown in the text. The formula is given below labeled as Equation (1) stated as:

$$Wn\,(dn,\,tn) = T\,F\,(d,\,t)\,*\,\log \tag{1}$$

Table 5. Experimental results of Random Forest classifier.

Ranking	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
Below Average	105	71	48	66	126	39
Average	87	111	117	117	93	141
Good	108	117	135	117	81	120

The exponential terms of the TF-IDF vector slash a word by calculating multiple of the words-based Term Frequency (TF) in relation to the Inverse Document Frequency (IDF). Term Frequency, referred to as TF, calculates the occurrences and appearances of any term in some equation in a unit of time, where the term or a word is the total number of occurrences in a standard dataset if compared with the number of words in the existing prepared manuscript.

5.3. Performance Comparison Analysis of Different Classification Algorithms on Same Dataset

The cognitive assessment modeling of the student dataset was tested on the test bed of different classification models and, hence, achieved a leap of 98% accuracy as compared with the result of algorithms. The detailed results are shown in Table 6. It reflects the peak accuracy of the RNN model as compared to Graph Cosine and the rest of the performance-analysis classifiers. The results of Table 6 show the performance measurement of the same student dataset on different algorithms to scale the level of Accuracy, Precision, and Recall score. These results are graphically represented in Figure 10.

Table 6. Performance analysis of different classification algorithms.

Id	Algorithm	Accuracy	Precision	Recall	F1 Score	Execution Time (ms)	Complexity
1	Graph Cosine	75	72	78	75	15	Moderate
2	K-means	55	50	60	54	10	Low
3	Naïve Bayes	90	88	92	90	5	Low
4	Vector Lookup Table	80	82	78	80	12	Moderate
5	Recurrent Neural Network (RNN)	98	97	99	98	200	High

*5.4. Performance Comparison Analysis of Metric Base of Algorithms on Dataset*

The cognitive assessment modeling of the student dataset based on metrics (Accuracy, Recall, Training Time in (Sec), Precision, Recall, Interoperability) was tested on the test bed for the RNN model, Baseline Model, and ML Model, and hence found RNN suitable and with impressive results according to the customization already stated and defined in the methodology section. The group of two datasets considered for our research study was formulated in the form of episodes represented by metric analysis in comparison with different classification algorithms as per Table 7. Dataset 1 is compiled and formulated based on RNN, which gives 85.3% accuracy as compared to the rest of the classification algorithm, to intelligently predict the cognitive levels of the dataset based on random-weight assignment to action verbs. Accordingly, the comprehensive model clarifies the class-wise split for stated Dataset 1. Conversely, Dataset 2 was obtained from the existing study captioned as Episode 2, and it was previously tagged into different BT levels. Class-wise distribution exhibits Dataset captioned as Dataset 1 and Dataset 2 where Dataset 1 is used to create a standard model for the proposed system and Dataset 2 is used to evaluate the projected system in contrast with the equivalent study on the mentioned dataset. The baseline models serve as a benchmark in designing ML design patterns and are customizable by iteratively training models in the contextual regime. It is shown in Table 7 below:

Table 7. Metric analysis of different classifications algorithm.

Metric	RNN	Baseline Model	ML Model
Accuracy %	85.3	70.2	83.5
Training Time Sec	300	15	120
Precision	0.89	0.72	0.87
Recall	0.91	0.70	0.88
Interoperability	Low	High	Medium

5.5. Comparison of RNN Classifier

This subsection details the setup of the experiments by including the Python main libraries, which are embedded to formulate the dataset collection. Figure 11 indicates the accuracy assessments based on versatile algorithms, which range from cosine similarity, which gives a 75% result, to Naïve Bayes (90%), and, hence, it achieved the highest percentage of 98% using RNN. The extended results show the performance measure due to dataset flexibility and contextual student cognitive assessment measures at the module-level and episode-level output.
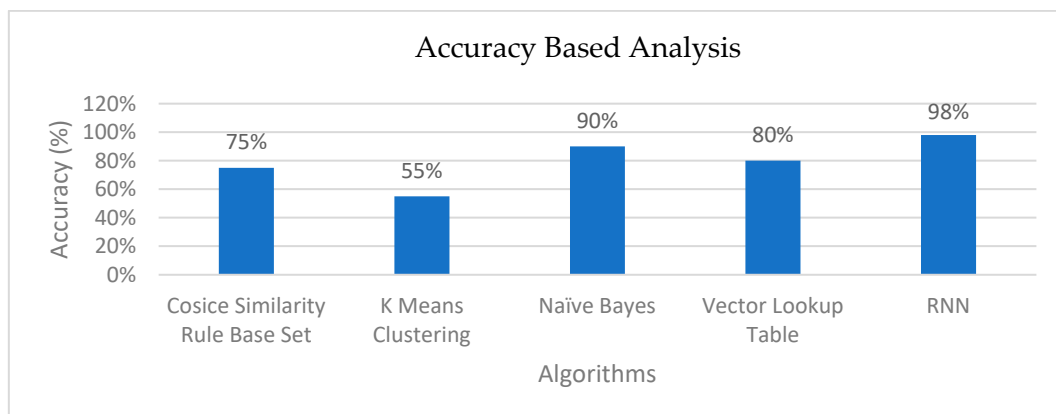
Figure 11. Comparison of RNN classifier.

Figure 12 reflects the fall of student performance at various levels of cognition, with most of the students performing well at the application level and most of the students not performing well at the Synthesis level based on the student interest and level of participation achieved through test items supported by the Random Forest classification algorithm.
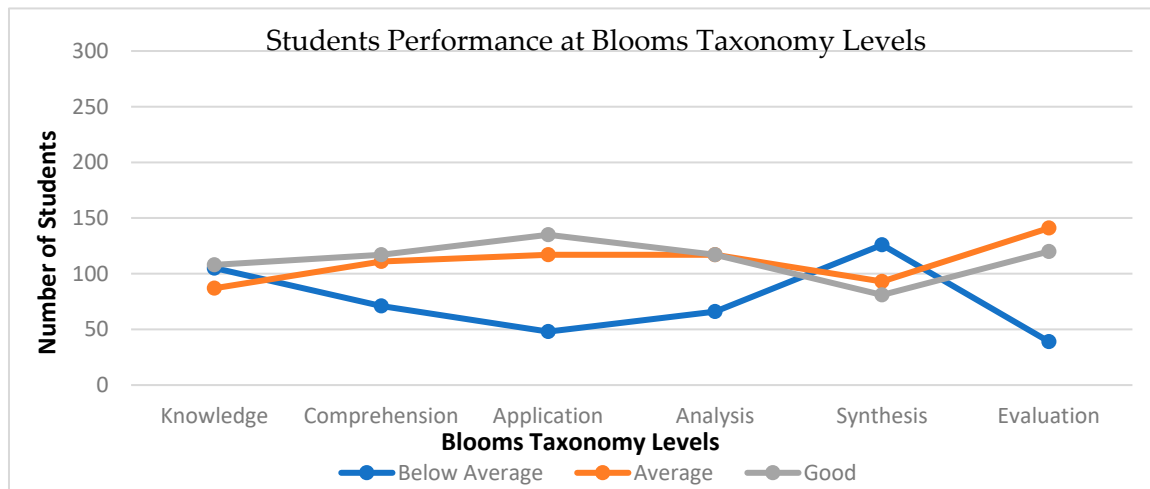

Figure 12. Student performance on cognitive levels.

5.6. Experimental Results of Random Forest Classifier

The Random Forest classifiers give aggregate-level results for identifying and mapping the cognitive scales of student assessment in relation to the Precision ratio, Recall Ratio, F1, and Accuracy. F1 measure is a predictive measurement at a metric scale used to estimate the performance of classification models in machine learning. It describes the balance ratio for imbalanced class distributions. It is the internal working of every classifier executed in Python.

The formula for calculating the F1 score is:

$$F1\ Score = \frac{TP}{TP + \frac{1}{2}\left(FP + FN\right)}$$

The detailed calculation of Random Forest based on different indicators is demonstrated in Table 8 below:

Table 8. Aggregate results of Random Forest accuracy on levels.

Model	Indicators	Taxonomy Levels Based Cognitive Scales					
		Level#1	Level#2	Level#3	Level#4	Level#5	Level#6
	Precision	89.8%	89.2%	93.3%	93%	94.16%	94.7%
Random Forest	Recall	92.0%	91.5%	92.1%	94.7%	92.95%	95.7%
	F1	90.7%	90.3%	92.7%	93.8%	93.55%	95.2%
	Accuracy			92.16%			

6. Performance Results at Each Level of Bloom's Taxonomy—Episode 1

The detailed cognitive assessment results achieved and compiled level by level for Episode 1 are shown comprehensively and labeled from Figure 13A–F. Accordingly, it shows performance measures at cognitive-level-based learning. Moderate-level performance stands at "Good" and is shown at various levels of taxonomy, as shown in the graphs below:
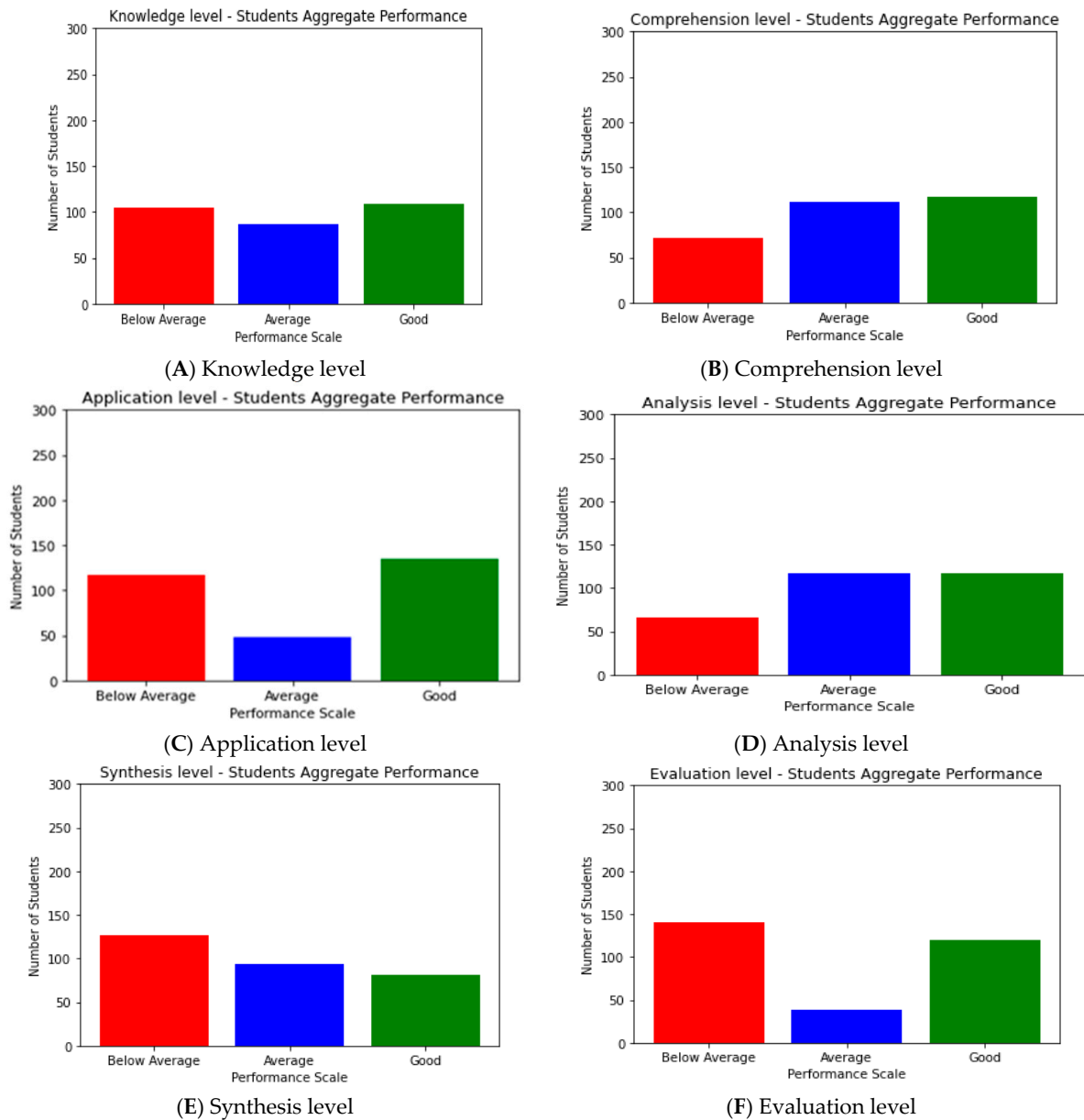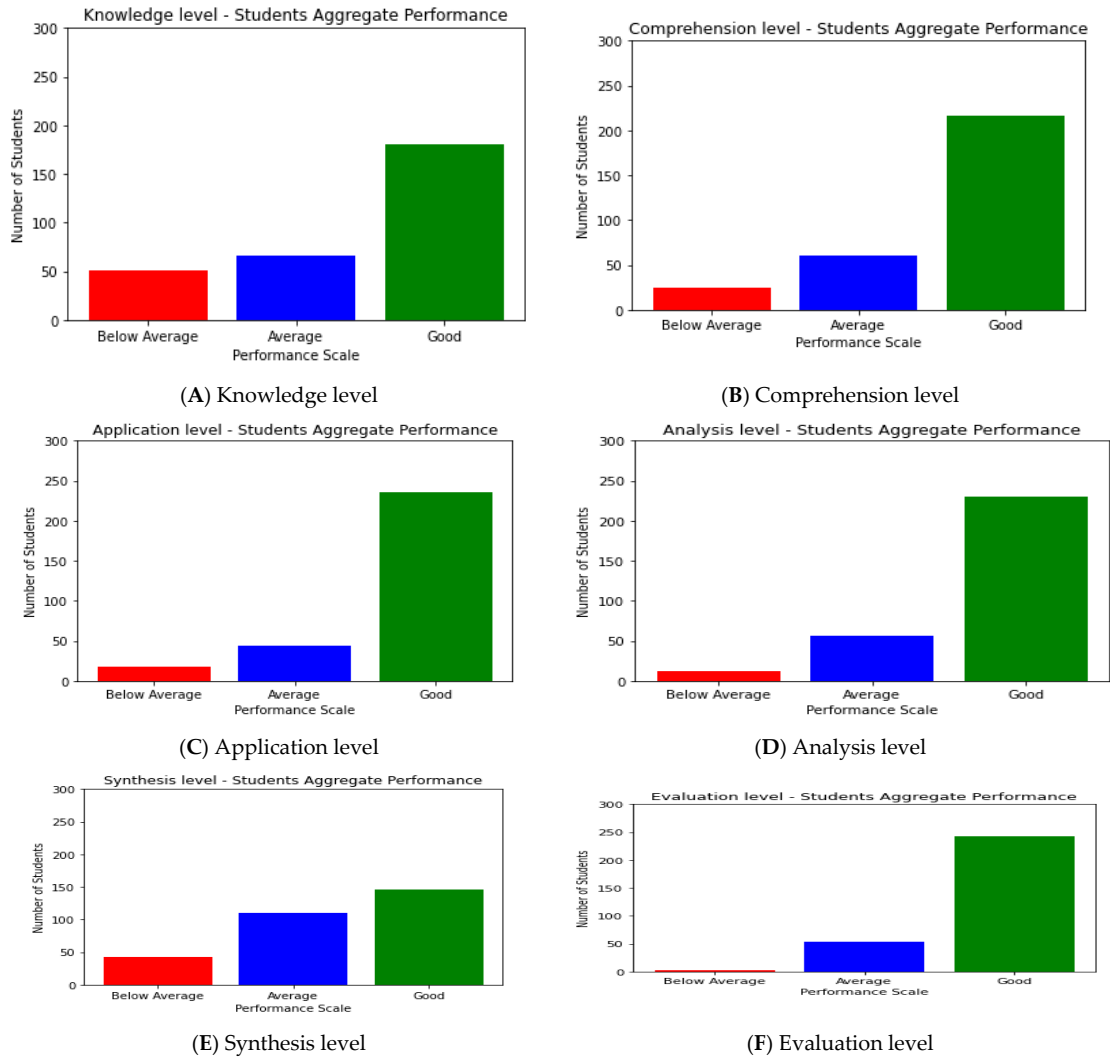
Figure 13. Students' performance results at (**A**) Knowledge level—Episode 1; (**B**) Comprehension level—Episode 1; (**C**) Application level—Episode 1; (**D**) Analysis level—Episode 1; (**E**) Synthesis level; (**F**) Evaluation level—Episode 1.

7. Performance Results at Each Level of Bloom's Taxonomy—Episode 2

The detailed cognitive assessment results achieved and compiled level by level are shown comprehensively and labeled from Figure 14A–F. Accordingly, it shows much-improved performance as compared to Episode 1 by fine-tuning different cognitive levels and by ensuring improved cognitive-level-based learning. The Application and Evaluation level shows much better results as compared to the rest of the cognitive levels.

(**A**) Knowledge level

(**B**) Comprehension level

(**C**) Application level

(**D**) Analysis level

(**E**) Synthesis level

(**F**) Evaluation level

Figure 14. Student performance results at (**A**) Knowledge level Episode 2; (**B**) Comprehension level Episode 2; (**C**) Application level Episode 2; (**D**) Analysis level Episode 2; (**E**) Synthesis level Episode 2; (**F**) Evaluation level Episode 2.

From Figure 14A–F, the student performance in Episode 2 showed much improvement, as the assessment results working as per the classifier improved results in Episode 2. The student performance has been seen dipped to Below Average, as per Figure 13A,F. It is concluded that students learned well the art of answering Evaluation-level questions, like compare, validate, evaluate, etc., and showed improved performance in terms of completeness and correctness. The parameters are tested through RNN and classified through Random Forest using Spyder, the Python IDE. The comparison shows improved performance at various levels of taxonomy as far as the learning curve of students is concerned, with achieved predicted accuracy of more than 90%.

Summarized Result of Episode 1 and Episode 2

This shows a concise preview of what students achieved after comprehensive methodology implementation and provides timely feedback for students to be prepared well on the scale of cognitive assessment as far as the second episode is concerned. It results in the optimum classification supported by the RNN algorithm and by the Random Forest classifier in predicting the students' best individual-level performances as well as student aggregate-level performances from an e-learning-based virtual student dataset. A total of 627 students fell Below Average as per Episode 1, which compares to 150 students reduced

to the same level by showing significant improvement as per Episode 2, shown in the summarized graph, Figure 15.
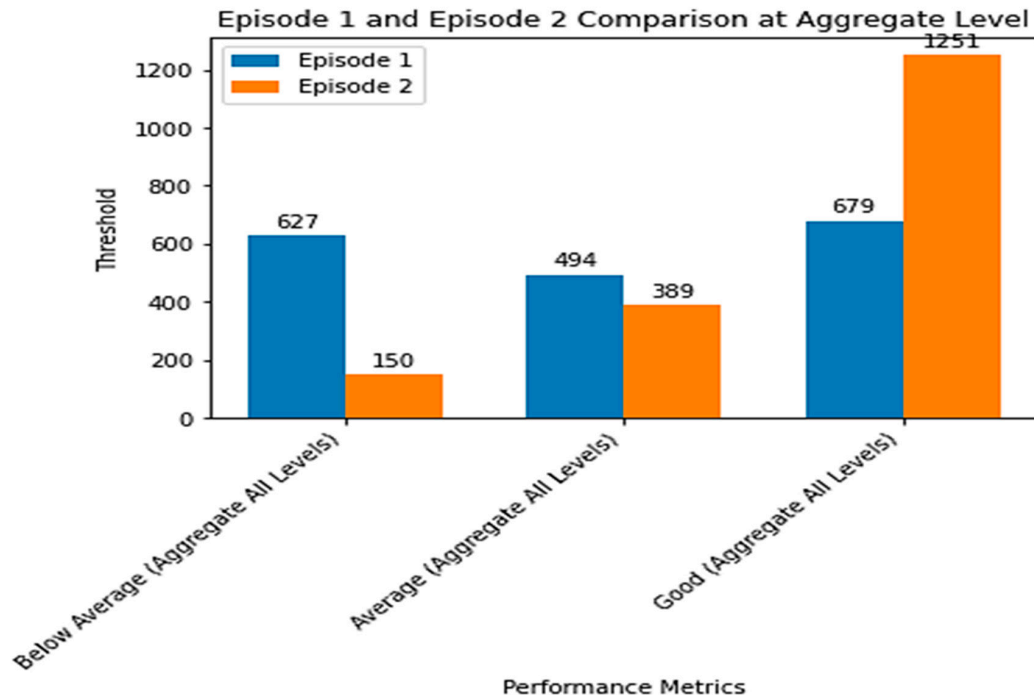

Figure 15. Summarized comparison of student performance in Episode 1 and Episode 2.

8. Conclusions

This study has the novelty of considering text-based questions using an NLP-based RNN model. It covers the shortfall of previous studies, which are limited to multiple-choice questions and keyword-based mapping approaches. The current research first perceives the level of a random pool of questions using Python libraries with a test ratio of 70:30 and then formulates a solution to group 300 students on the same test bed for assessment using word embedding and word2vector Python libraries on various scales and thresholds of weighted cognitive levels. This is not the only procedural module to conclude the study. It also leads to Episode 2, followed up by adopting the repeated procedure to test the cognitive levels of the same group of students in variant time slices to gauge the improvement in their learning performance. The results showed a 20% increase in overall learning capacity by building up using a feedback mechanism. The classification of the student's group was comprehended using Random Forest and the KNN classifier to not only refine the computational complexity but also to reduce the statistical calculation derived from two follow-up episodes in Python.

NLP-based methods are implemented to evaluate the data assembled from each student of the artificial intelligence course. The Fired dataset is used to forecast the level of BT-based questions utilizing the RNN classifier with approximately 98% accuracy. Once the taxonomy level is identified, the conforming level is analyzed for students' text-based answer feed. The Random Forest algorithm ranks students' input performance on a 3-point scale: Below Average, Average, and Good. The accuracy percentage of students' aggregate performance was 92.16%. The compiled results based on the proposed methodology reflect better results as compared to the existing studies of the same nature. There exist certain limitations to the conducted research, which do not impact the addressed problem set. These include the domain based on students being tested and the computational complexity, which may arise if the conducted and concluded methodology is applied to diversely placed

students belonging to different demographic groups. The assessment of students by the stated aspects may impact and reflect different learning curves.

9. Future Work

Future work should be based on and revolve around the health prediction and attention analysis of students in specified time slices to maximize student engagement in content. This should create a spectrum for innovating interactivity of the content not only for effective delivery but to summarize well and lead to effective assessment procedures. The role of semantic analysis plays a pivotal role in natural-language processing (NLP) by providing significant opportunities for cognitive assessment advancements. The concept of cognitive health is derived from future research plans on already concluded cognitive assessments of the e-learning student model.

Abbreviations

AI	Artificial Intelligence
CogAT	Cognitive Ability Testing
LSTM	Long Short-Term Memory
GENAI	Generative artificial intelligence
RF	Random Forest
ML	Machine Learning
NLP	Natural-Language Processing
BT	Blooms Taxonomy
COG	Cognition
KNN	K Nearest Neighbor
RNN	Recurrent Neural Network
CNN	Convolutional Neural Network
MOOCS	Massively Open Online Courses
IDF	Inverse Document Frequency
NLTK	Natural-Language Toolkit
DL	Deep Learning
LO	Learning Objective
CLO	Course Learning Objective

References

1. Omar, N.; Haris, S.S.; Hassan, R.; Arshad, H.; Rahmat, M.; Zainal, N.F.A.; Zulkifli, R. Automated analysis of exam questions according to BT. *Procedia Social Behav. Sci.* **2012**, *59*, 297–303. [CrossRef]
2. Chang, W.-C.; Chung, M.-S. Automatic applying BT to classify and analysis the cognition level of english question items. In Proceedings of the Joint Conferences on Pervasive Computing (JCPC), Taipei, Taiwan, 3–5 December 2009; pp. 727–734.
3. OSMAN; Yahya, A.A. Classifications of exam questions using natural language syntatic features: A case study based on Bloom's taxonomy. In Proceedings of the 3rd International Arab Conference on Quality Assurance in Higher Education (IACQA'2016), Zarqa, Jordan, 9–11 February 2016; pp. 1–8.
4. Monrad, S.U.; Bibler Zaidi, N.L.; Grob, K.L.; Kurtz, J.B.; Tai, A.W.; Hortsch, M.; Gruppen, L.D.; Santen, S.A. What faculty write versus what students see? Perspectives on multiple-choice questions using Bloom's taxonomy. *Med. Teach.* **2021**, *43*, 575–582. [CrossRef] [PubMed]
5. Krathwohl, D.R.; Anderson, L.W. Merlin C. Wittrock and the revision of BT. *Educ. Psychol.* **2010**, *45*, 64–65. [CrossRef]
6. Bloom, B.S. Taxonomy of educational objectives: The classification of educational goals. Cognit. *Domain* **1981**, *51*, 441–453.
7. Chyung, S.-Y.; Stepich, D. Applying the 'congruence' principle of BT to designing online instruction. *Q. Rev. Distance Educ.* **2003**, *4*, 317–330.
8. Kowsari, K.; Meimandi, K.J.; Heidarysafa, M.; Mendu, S.; Barnes, L.; Brown, D. Text classification algorithms: A survey. *Information* **2019**, *10*, 150. [CrossRef]
9. Haris, S.S.; Omar, N. A rule-based approach in BT question classification through natural language processing. In Proceedings of the 2012 7th International Conference on Computing and Convergence Technology (ICCCT), Seoul, Republic of Korea, 3–5 December 2012; pp. 410–414.
10. Dwivedi, S.K.; Arya, C. Automatic text classification in information retrieval: A survey. In Proceedings of the 2nd International Conference on Information and Communication Technology for Competitive Strategies (ICTCS), Udaipur India, 4–5 March 2016; p. 131.
11. Das, S.; Mandal, S.K.D.; Basu, A. Classification of action verbs of BT cognitive domain: An empirical study. *J. Educ.* **2021**, *201*, 002205742110021.
12. Young, T.; Hazarika, D.; Poria, S.; Cambria, E. Recent trends in deep learning based natural language processing. *IEEE Comput. Intell. Mag.* **2018**, *13*, 55–75. [CrossRef]
13. Bengio, Y.; Senecal, J.-S. Adaptive importance sampling to accelerate training of a neural probabilistic language model. *IEEE Trans. Neural Netw.* **2008**, *19*, 713–722. [CrossRef]
14. Khairuddin, N.N.; Hashim, K. Application of BT in software engineering assessments. In Proceedings of the 8th WSEAS International Conference on Applied Computer Science (ACS'08), Venice, Italy, 21–23 November 2008; pp. 66–69.
15. Zhang, J.; Wong, C.; Giacaman, N.; Luxton-Reilly, A. Automated classification of computing education questions using BT. In Proceedings of the 23rd Australasian Computing Education Conference, Virtual, 2–4 February 2021; pp. 58–65.
16. Abduljabbar, D.A.; Omar, N. Exam questions classification based on BT cognitive level using classifiers combination. *J. Theor. Appl. Inf. Technol.* **2015**, *78*, 447.
17. Osadi, K.; Fernando, M.G.N.A.S.; Welgama, W. Ensemble classifier-based approach for classification of examination questions into BT cognitive levels. *Int. J. Comput. Appl.* **2017**, *975*, 8887.
18. Shahzad, R.; Aslam, M.; Al-Otaibi, S.; Javed, M.S.; Khan, A.R.; Bahaj, S.A.; Saba, T. Multi-Agent System for Students Cognitive Assessment in E-Learning Environment. *IEEE Access* **2024**, *12*, 15458–15467. [CrossRef]
19. Badjatiya, P.; Gupta, S.; Gupta, M.; Varma, V. Deep learning for hate speech detection in tweets. In Proceedings of the 26th International Conference on World Wide Web Companion, Perth, Australia, 3–7 April 2017; pp. 759–760.
20. Swart, J.; Daneti, M. Analyzing learning outcomes for electronic fundamentals using BT. In Proceedings of the 2019 IEEE Global Engineering Education Conference (EDUCON), Dubai, United Arab Emirates, 9–11 April 2019; pp. 39–44.
21. Rahmatih, A.N.; Indraswati, D.; Gunawan, G.; Widodo, A.; Maulyda, M.A.; Erfan, M. An analysis of questioning skill in elementary school pre-service teachers based on BT. *J. Phys. Conf. Ser.* **2021**, *1779*, 012073. [CrossRef]
22. Hoogeveen, D.; Wang, L.; Baldwin, T.; Verspoor, K.M. Web forum retrieval and text analytics: A survey. *Found. Trends Inf. Retr.* **2018**, *12*, 1–163. [CrossRef]
23. Manning, C.D.; Raghavan, P.; Schutze, H. *Introduction to Information Retrieval*; Cambridge University Press: Cambridge, UK, 2008; Chapter 20; pp. 405–416.
24. Buckley, C. *Implementation of the Smart Information Retrieval System*; Tech. Rep. TR85-686; Cornell University: New York, NY, USA, 1985. Available online: https://ecommons.cornell.edu/server/api/core/bitstreams/9f55bb4c-70a1-49b8-b78d-8914a8f51e62/content (accessed on 24 November 2024).
25. O'Riordan, C.; Sorensen, H. Information filtering and retrieval: An overview. In Proceedings of the 16th Annual International Conference of the IEEE, Atlanta, GA, USA, 4–7 May 1997; pp. 28–31.

26. van Hoeij, M.J.W.; Haarhuis, J.C.M.; Wierstra, R.F.A.; van Beukelen, P. Developing a classification tool based on Bloom's taxonomy to assess the cognitive level of short essay questions. *J. Vet. Med. Educ.* **2004**, *31*, 261–267. [CrossRef]

27. Tariq, M.U.; Sergio, R.P. Innovative Assessment Techniques in Physical Education: Exploring Technology-Enhanced and Student-Centered Models for Holistic Student Development. In *Global Innovations in Physical Education and Health*; IGI Global: Hershey, PA, USA, 2025; pp. 85–112.

28. Ozen, Z.; Pereira, N.; Karatas, T.; Castillo-Hermosilla, H.; Maeda, Y. A Meta-Analytic Evaluation: Investigating Evidence for the Validity of the Cognitive Abilities Test. *Gift. Child Q.* **2025**, *69*, 3–15. [CrossRef]

29. Almutawa, S.S.; Alshehri, N.A.; AlNoshan, A.A.; AbuDujain, N.M.; Almutawa, K.S.; Almutawa, A.S. The influence of cognitive flexibility on research abilities among medical students: Cross-section study. *BMC Med. Educ.* **2025**, *25*, 7. [CrossRef]

30. Jayakodi, K.; Bandara, M.; Perera, I. An automatic classifier for exam questions in engineering: A process for BT. In Proceedings of the 2015 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), Zhuhai, China, 10–12 December 2015; pp. 195–202.

31. Yamasari, Y.; Rochmawati, N.; Putra, R.E.; Qoiriah, A.; Yustanti, W. Predicting the students' performance using regularization-based linear regression. In Proceedings of the 2021 Fourth International Conference on Vocational Education and Electrical Engineering (ICVEE), Surabaya, Indonesia, 2–3 October 2021; pp. 1–5.

32. Singla, A. Roberta and BERT: Revolutionizing Mental Healthcare through Natural Language. *Shodh Sagar J. Artif. Intell. Mach. Learn.* **2024**, *1*, 10–27. [CrossRef]

33. Stringer, J.K.; Santen, S.A.; Lee, E.; Rawls, M.; Bailey, J.; Richards, A.; Perera, R.A.; Biskobing, D. Examining Bloom's taxonomy in multiple choice questions: Students' approach to questions. *Med. Sci. Educ.* **2021**, *31*, 1311–1317. [CrossRef]

34. Yusof, N.; Hui, C.J. Determination of Bloom's cognitive level of question items using artificial neural network. In Proceedings of the 2010 10th International Conference on Intelligent Systems Design and Applications, Cairo, Egypt, 29 November–1 December 2010; pp. 866–870.

35. Chanaa, A. A cognitive level evaluation method based on machine learning approach and Bloom of taxonomy for online assessments. *J. Educ. Learn. (EduLearn)* **2024**, *18*, 553–560. [CrossRef]

36. KJayakodi; Bandara, M.; Perera, I.; Meedeniya, D. Wordnet and cosine similarity-based classifier of exam questions using BT. *Int. J. Emerg. Technol. Learn.* **2016**, *11*, 142–149. [CrossRef]

37. Hong, S.; Kim, J.; Yang, E. Automated text classification of maintenance data of higher education buildings using text mining and machine learning techniques. *J. Archit. Eng.* **2022**, *28*, 04021045. [CrossRef]

38. Botelho, A.; Baral, S.; Erickson, J.A.; Benachamardi, P.; Heffernan, N.T. Leveraging natural language processing to support automated assessment and feedback for student open responses in mathematics. *J. Comput. Assist. Learn.* **2023**, *39*, 823–840. [CrossRef]

39. Gupta, S.; Gulia, P. Comparative Analysis of Predictive Algorithms for Performance Measurement. *IEEE Access* **2024**, *12*, 33949–33958. [CrossRef]

40. Mohammed, M.; Omar, N. Question classification based on BT cognitive domain using modified TF-IDF and word2vec. *PLoS ONE* **2020**, *15*, e0230442. [CrossRef]

41. Mikolov, T.; Karafiát, M.; Burget, L.; Černocky, J.; Khudanpur, S. Recurrent neural network-based language model. In Proceedings of the 11th Annual Conference of the International Speech Communication Association, Chiba, Japan, 26–30 September 2010; pp. 1–24.

42. Sharma, S.; Saxena, D.K. Understanding the Cognitive Constituents of E-Learning. In *Best Practices and Strategies for Online Instructors: Insights from Higher Education Online Faculty*; IGI Global Scientific Publishing: Hershey, PA, USA, 2025; pp. 277–312.

43. Wijanarko, B.D.; Heryadi, Y.; Toba, H.; Budiharto, W. Question generation model based on key-phrase, context-free grammar, and BT. *Educ. Inf. Technol.* **2021**, *26*, 2207–2223. [CrossRef]

44. El Msayer, M.; Aoula, E.S.; Bouihi, B. Artificial intelligence in computerized adaptive testing to assess the cognitive performance of students: A Systematic Review. In Proceedings of the 2024 International Conference on Intelligent Systems and Computer Vision (ISCV), Fez, Morocco, 8–10 May 2024; pp. 1–8.

45. Shaikh, S.; Daudpotta, S.M.; Imran, A.S. Bloom's learning outcomes' automatic classification using lstm and pretrained word embeddings. *IEEE Access* **2021**, *9*, 117887–117909. [CrossRef]

46. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. In Proceedings of the Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013, Lake Tahoe, NV, USA, 5–8 December 2013; pp. 3111–3119.

47. Socher, R.; Perelygin, A.; Wu, J.; Chuang, J.; Manning, C.D.; Ng, A.; Potts, C. Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, WA, USA, 18–21 October 2013; pp. 1631–1642.

48. Vasilomanolakis, E.; Karuppayah, S.; Mühlhäuser, M.; Fischer, M. Taxonomy and survey of collaborative intrusion detection. *ACM Comput. Surv.* **2015**, *47*, 55:1–55:33. [CrossRef]

49. Song, D.; Rice, M.; Oh, E.Y. Participation in online courses and interaction with a virtual agent. *Int. Rev. Res. Open Distrib. Learn.* **2019**, *20*. [CrossRef]

50. Swiecki, Z.; Khosravi, H.; Chen, G.; Martinez-Maldonado, R.; Lodge, J.M.; Milligan, S.; Selwyn, N.; Gašević, D. Assessment in the age of artificial intelligence. *Comput. Educ. Artif. Intell.* **2022**, *3*, 100075. [CrossRef]

51. Wu, Q.; Bansal, G.; Zhang, J.; Wu, Y.; Zhang, S.; Zhu, E.; Li, B.; Jiang, L.; Zhang, X.; Wang, C. Autogen: Enabling nextgen llm applications via multi-agent conversation framework. *arXiv* **2023**, arXiv:230808155.

52. Atiullah, K.; Fitriati, S.W.; Rukmini, D. Using revised BT to evaluate higher order thinking skills (hots) in reading comprehension questions of english textbook for year X of high school. *English Educ. J.* **2019**, *9*, 428–436. [CrossRef]

53. Laddha, M.D.; Lokare, V.T.; Kiwelekar, A.W.; Netak, L.D. Classifications of the summative assessment for revised BT by using deep learning. *arXiv* **2021**, arXiv:2104.08819.

54. Febri, A. Analysis of Students' Critical Thinking Skills at Junior High School in Science Learning. *J. Phys. Conf. Ser.* **2019**, *1397*, 012018. [CrossRef]

55. Zhang, D.; Lee, W.S. Question classification using support vector machines. In Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Toronto, ON, Canada, 28 July–1 August 2003; pp. 26–32.

56. Yan, L.; Greiff, S.; Teuber, Z.; Gašević, D. Generative Artificial Intelligence and Human Learning. *arXiv* **2024**, arXiv:2408.12143.

57. Musa, M.; Ahmadu, A.S.; Williams, C. Comparative Analysis of K-Means and Naïve Bayes Algorithms for Predicting Students' Academic Performance. *Int. J. Dev. Math. (IJDM)* **2024**, *1*, 196–208. [CrossRef]

58. Servin, C.; Tang, C.; Geissler, M.; Stange, M.; Tucker, C. Enhanced verbs for BT with focus on computing and technical areas. In Proceedings of the 52nd ACM Technical Symposium on Computer Science Education, Virtual, 13–20 March 2021; p. 1270.

59. Yahya, A.; Osman, A. *Automatic Classification of Questions into Bloom's Cognitive Levels Using Support Vector Machines*; Tech. Rep. 2914; Naif Arab University for Security Sciences: Riyadh, Saudi Arabia, 2011.