Project Report On

Exam Question Classification Based on Blooms Taxonomy
A Hybrid Approach Using Traditional Machine Learning and Neural Networks

Submitted By
Chetan Ingle 433034
Akash Gulge 434022
Shreyas Chandurkar 434044

Under the Guidance of

Prof.Chaitali Shewale

In Partial fulfilment of
Bachelor of Technology in Information Technology [2023-2024]

At

Department of Information Technology

Vishwakarma Institute of Information Technology, Pune 411048

Affiliated To

Savitribai Phule Pune University, Pune

Bansilal Ramnath Agarwal Charitable Trust's

Vishwakarma Institute of Information Technology, Pune

CERTIFICATE

This is to certify that the work entitled "**Exam Question Classification Based On Blooms Taxonomy: A Hybrid Approach Using Traditional Machine Learning And Neural Network**" is a bonafide work carried out by Mr. Chetan Ingle , Mr. Akash Gulge, Mr. Shreyas Chandurkar in partial fulfilment of the award of Bachelor of Technology in Information Technology, Savitribai Phule Pune University, Pune, during the year 2023-24 The project report has been approved as it satisfies the academic requirements in respect of the project work prescribed for the Bachelor of Technology Degree.

Prof. Chaitali Shewale

Project Guide

Prof. Pravin Futane

Head, IT Department

Dr. Vivek Deshpande

Director, VIIT PUNE.

Date:

Examiner: 1 ……………. 2. …………….

Place: Pune

Project Allocation Letter

UniApps
University exam ERP

CIN : U72900PN2011PTC138954

To
Prof. Chaitali Shewale
Department of Information Technology
Vishwakarma Institute Information Technology

Subject: Sanction of Technical Sponsorship for "Classification of Questions Based on Bloom's Taxonomy: A Hybrid Approach Using Traditional Machine Learning and Neural Network"

Dear Prof. Chaitali,

We are pleased to inform you that [Company/Organization Name] has decided to grant technical sponsorship to the project titled "Classification of Questions Based on Bloom's Taxonomy: A Hybrid Approach Using Traditional Machine Learning and Neural Network," led by students Chetan Ingle (ID: 22010416), Akash Gulge (ID: 22010608), and Shreyas Chandurkar (ID: 22120264) from [Your Institution/Organization Name].

After a thorough review of your proposal and considering the innovative nature and potential impact of your project on the educational sector, we are excited to support your efforts to advance the application of machine learning and neural networks in education. We believe that your project aligns with our company's mission to foster technological innovation and education.

The technical sponsorship provided by WeshineTech Pvt Ltd will include the following support:

Computational Resources: We will provide access to necessary computational resources and software tools to aid in the development and testing of your machine learning models and neural networks.

Technical Mentorship: Our team of experts in machine learning and educational technology will be available for consultation to offer guidance, insights, and feedback to refine your project methodologies and objectives.

Please note that this sponsorship is contingent upon the receipt of regular progress reports and a final project report upon completion, detailing the methodologies, results, and potential implications of your research. Additionally, [Company/Organization Name] should be acknowledged as a sponsor in all project-related publications, presentations, and promotional materials.

We are enthusiastic about the potential of your project and look forward to seeing the impact of your work on enhancing educational practices through the innovative use of technology. Congratulations once again on securing this sponsorship, and we wish you the best in your research endeavours.

Sincerely,

Dr. Sandeep Kadam
Voce President
Date : 20-03-2024

WESHINE TECH PVT. LTD. PUNE

Address
Office No. 104, Mantri Alpine, Mumbai - Bangalore Highway,
above Crystal Honda, Bavdhan, Pune - 411021,
Ph - 020 67700500,
sales@weshinetech.biz | www.weshinetech.in | www.uniapps.in

WeShineTech

Letter Of Completion

UniApps
University exam ERP

CIN : U72900PN2011PTC138954

To
Prof. Chaitali Shewale
Department of Information Technology
Vishwakarma Institute Information Technology

Subject: Letter of Completion

Dear Prof. Chaitali,

We are pleased to inform you that the project titled "Classification of Questions Based on Bloom's Taxonomy: A Hybrid Approach Using Traditional Machine Learning and Neural Network," is completed by students Chetan Ingle (ID: 22010416), Akash Gulge (ID: 22010608), and Shreyas Chandurkar (ID: 22120264) from Vishwakarma Institute Information Technology.

The work done by students is satisfactory and we appreciate the efforts taken by you and Students throughout the completion of project.

Sincerely,

Dr. Sandeep Kadam
Vice President

Date : 11-05-2024

WESHINE TECH PVT. LTD. * PUNE

Address
Office No. 104, Mantri Alpine, Mumbai - Bangalore Highway,
above Crystal Honda, Bavdhan, Pune - 411021,
Ph - 020 67700500,
sales@weshinetech.biz |

Acknowledgement

We are deeply grateful and would like to convey our heartfelt thanks to Professor Pravin Futane, Head of the Department, for his exemplary leadership and invaluable guidance throughout the duration of this project. His unwavering support and provision of essential resources were instrumental in the successful completion of this project report.

Our heartfelt appreciation extends to Professor Chaitali Shewale, our esteemed project guide, for her mentorship, insightful feedback, and constant encouragement. Her expertise and dedication significantly contributed to the quality and depth of our work.

We extend our thanks to all the faculty members of the Department of Information Technology at VIIT, Pune. Their valuable time, support, comments, suggestions and motivation played a pivotal role in shaping this project.

We acknowledge and appreciate the collaborative environment fostered by the Department and the institute, which has been conducive to our academic and project endeavours.

Chetan Ingle 433034
Akash Gulge 434022
Shreyas Chandurkar 434044

ABSTRACT

This study addresses the task of categorizing exam questions according to Bloom's Taxonomy, a crucial element in educational assessment. Manual classification is time-consuming and susceptible to subjective judgment. To address this challenge, a novel hybrid approach is introduced, integrating both traditional machine learning methods and Neural Networks.

The primary objective is to automate the classification of exam questions, providing educators with a more efficient and unbiased assessment tool for evaluating students' cognitive abilities. By harnessing machine learning techniques, the aim is to enhance the accuracy and speed of question categorization. The hybrid methodology combines the interpretability of traditional models with the capacity of Neural Networks to identify complex patterns within the data.

This integration is designed to improve the precision of classifying questions across Bloom's Taxonomy levels. The potential applications span various educational domains, including automated grading, personalized learning, and detailed performance analysis. In the realm of educational technology, the hybrid model holds promise for contributing to adaptive learning platforms, improved question generation tools, and tailored feedback mechanisms for learners.

In summary, this hybrid approach addresses the challenges associated with manual classification while advancing educational technology and research methodologies. It provides a standardized and effective tool for educators, fostering a deeper understanding of students' cognitive development. The model's automation capabilities aim to streamline assessment processes, offering a valuable resource for educators and researchers alike.

CONTENTS

List Of Figures

List Of Tables

1. INTRODUCTION

1.1 What is Blooms Taxonomy?

Bloom's Taxonomy is a framework that classifies various educational objectives and cognitive skills into a structured hierarchy. It was developed by educational psychologist Benjamin Bloom in the 1950s and later revised by Anderson and Krathwohl in 2001. It organizes cognitive skills into six levels, from low thinking skills to high thinking skills:

1. **Remembering**: Retrieving facts, data, or concepts.
2. **Understanding**: Clarifying ideas or concepts and showing comprehension.
3. **Applying**: Implementing knowledge or information in different contexts or situations.
4. **Analysing**: Decomposing information into parts, recognizing patterns, and grasping relationships.
5. **Evaluating**: Forming judgments based on criteria and standards.
6. **Creating**: Producing new ideas, products, or perspectives.

Bloom's Taxonomy gives a framework for peoples to design learning objectives , assessments that target specific cognitive skills, aiding in the evaluation and improvement of students' understanding and application of knowledge.

1.2 Aim or Purpose of Project:

The purpose of the project is to create a system that automates the classification of exam questions which are based on Bloom's Taxonomy. The automation aims to bring efficiency and objectivity to the grading process, mitigating the time-consuming nature of manual classification and reducing the influence of personal opinions. The overarching goal is to increase the accuracy of question classification using a combination of traditional machine learning methods and Neural Networks.

1.3 Brief Introduction of Project:

The project focuses on leveraging technology, specifically machine learning, to classify the exam questions according to Bloom's Taxonomy. By combining traditional machine learning models and Neural Networks, the system aims to provide a reliable and precise classification mechanism. This approach addresses the challenges associated with manual grading, such as time constraints and subjective biases. The project not only seeks to streamline the grading process but also aligns with broader educational goals, contributing to the improvement of educational technology and research methods.

1.4 How Proposed System Will Help Users:

The proposed system offers several benefits to users in the education domain:

- **Efficient Grading:** The system enables automated grading, saving time for educators and facilitating a quicker turnaround for assessments.

- **Consistency and Objectivity:** By utilizing machine learning, the system reduces personal biases in question classification, ensuring more consistent and objective grading across different assessors.

- **Personalized Learning Experiences:** The model's accurate classification can contribute to creating personalized learning experiences for students, aligning assessments with their cognitive abilities and learning styles.

- **Performance Analysis:** The system provides detailed performance analyses, offering insights into students' strengths and weaknesses, which can inform targeted interventions and improvements in teaching strategies.

- **Educational Technology Enhancement:** In the realm of educational technology, the system supports adaptive learning tools, improves question creation processes, and offers personalized feedback, enhancing the overall learning experience.

- **Contribution to Research and Development:** Beyond immediate applications, the system's insights can contribute to advancements in educational data analysis, the development of new assessment methods, and a deeper understanding of cognitive aspects related to educational content.

2. LITERATURE SURVEY

Various studies focused on automating the classification of the exam questions which are based on Bloom's Taxonomy using ML techniques. These studies address the challenges associated with manual classification, aiming to streamline the assessment process and provide educators with efficient tools for evaluating students' cognitive skills.

Sr No	Paper Name	Author	Year and Publication	Survey
1	Classifying Question Papers with Bloom's Taxonomy Using Machine Learning Techniques	Author Minni Jain, Rohit Beniwal, Aheli Ghosh, Tanish Grover, and Utkarsh Tyagi	SpringerLink 2019	[1] The authors employed a methodology that involved using various Machine Learning techniques used to classify question papers into different levels. They collected a dataset of 1024 questions and developed a web app to evaluate approach. The research approach included the use of nine different ML techniques, such as Logistic Regression, Linear Discriminant Analysis, K-Nearest Neighbors, Random Forest, Decision Trees, Support Vector Machine (SVM), Neural Network, and others, to classify the question papers. The implementation details included feature extraction, ML techniques used, and experimental setup. The authors also discussed related work, the work flow, coding approach, and the experimental setup in the paper. They got the result that the accuracy of 83.3% was achieved with Logistic Regression and Linear Discriminant Analysis ML techniques.

				Additionally, the authors suggested using Nature-Inspired algorithms to further improve accuracies.
2	Question classification based on Bloom's taxonomy cognitive domain using modified TF-IDF and word2vec	Manal Mohammed and Nazlia Omar. Manal Mohammed	PLoS One. 2020	[2] The authors interpret their findings as demonstrating the effectiveness of their proposed method for classifying exam questions in Bloom’s taxonomy cognitive domain. They emphasize that their model, which utilizes modified TF-IDF and word2vec features, achieved significant results in classifying questions from various fields based on Bloom’s taxonomy. The study used two datasets, one 141 questions and the other having 600 questions, and evaluated the results using three different classifiers: KNN, Logistic Regression,SVM. The results showed that the method outperformed traditional TF-IDF, getting F1-measures from 71.1% to 89.7% for the different classifiers and datasets. Additionally, the authors highlighted the potential applications of their model in educational settings, such as aiding educators and lecturers in analyzing exam questions to complete the requirements for various levels for education. They also suggested potential

				applications in automatic test generation systems, intelligent tutoring systems, and serious games. Furthermore, the authors noted that further study could improve, enhance, or extend their work by expanding the dataset to include different types of questions and providing a public benchmark dataset for question classification
3	Analysis of Bloom Taxonomy-Based Examination Data Using Data Mining	Amit Kumar, Dr. Dinesh Singh, Dr. M. S. Dhankhar	IJISAE 2023	[3] The authors interpret their findings by presenting a comprehensive analysis of the research conducted on the application of technique-data mining to the assessment of student test scores, particularly in the context of utilizing Bloom's Taxonomy for question classification. They utilize many machine learning algorithms, like BiLSTM , CNN, for feature extraction and model training and testing. The authors compare the performance of their proposed model with other models, comparing metrics like recall, precision,F1-score. They also provide a detailed evaluation of the outcomes, including the accuracy, precision, recall, and confusion matrix of the proposed model. Additionally, the authors present the results of their

				research in a tabular format, outlining the comparison criteria applied to the outcomes and the mathematical equations used for evaluation. They also discuss the implications of their findings, highlighting the usefullness of the model in predicting question difficulty levels and its potential for enhancing the assessment of students' comprehension and learning progress. Overall, the authors interpret their findings by demonstrating the superiority of their proposed model and its contribution to the field of automated assessment systems.
4	Exam Questions Classification Based on Bloom’s Taxonomy: Approaches and Techniques	Monika Yogendra Gaikwad, Vaishnavi Ravindra Kute, Gayatri Subhash Pawar, Mrunal Manikchand Magare	ICCIS 2020	[4] The authors interpret their findings by proposing an autonomous question paper-generation system that utilizes Bloom’s Taxonomy and machine learning to assign marks to questions and generate question papers. They emphasize the significance of leveraging Bloom’s Taxonomy for question categorization and mark assignment, providing a structured approach to streamline the process of question paper generation. The proposed system architecture, algorithm, and technology for question paper generation

				are outlined, with a focus on Bloom's Taxonomy for categorizing questions, text pre-processing, rule development, and randomization algorithms. The authors conclude by highlighting the potential of their proposed system to accurately assign marks to questions and generate multiple question papers by using Bloom's Taxonomy and machine learning. Additionally, the document provides a thorough review of existing literature on exam question classification which is based on Bloom's six level Taxonomy, offering insights into the research landscape and the evolution of automated question paper generation systems.
5	CLASSIFYING EXAM QUESTIONS BASED ON BLOOM'S TAXONOMY USING MACHINE LEARNING APPROACH	Nidaa Ghalib and Dhiyaa Salih Hammad from Al-Furat Al-Awsat Technical University	Technologies for the development of information systems tris-2019	[5] The authors of the document utilized a machine learning approach which is used to classify exam questions based on Bloom's Taxonomy. The methodology involved several phases, including data set planning and compiling, tokenizing and stop words elimination, feature selection, classification, and evaluation. They employed supervised machine learning approaches, specifically Naïve Bayes and K-Nearest Neighbour classifiers, to mainly

				classify exam questions. The data set which was used for the purpose of training and testing was takenfrom a programming questions bank for the Bachelor of Information Technology Year 1 to 3, as well as questions from various sites related to Bloom's taxonomy studies in programming subjects. Additionally, the authors experimented with various feature selection methods, such as Mutual Information, Odd Ratio, and Chi-Square, to improve performance of the machine learning models for the classification. The results indicated that feature selection methods positively contributed to the performance of the classifiers. Overall, the authors' methodology involved using of machine learning techniques and feature selection methods to automate the classification of exam questions which is mainly based on Bloom's Taxonomy.
6	Exam questions classification based on Bloom’s taxonomy cognitive level using classifiers combination	Dhuha Abdulhadi Abduljabbar and Nazlia Omar	Journal of Theoretical and Applied Information Technology 2015.	[6] The paper contributes new knowledge by proposing a combination that efficiently solve the problem of cognitive category finding for programming questions and achieves satisfactory results. The study

				introduces a method uses a combination of strategy which is based on a voting algorithm, integrating three machine learning classifiers: Support Vector Machine , Naïve Bayes , and k-Nearest Neighbour . This approach aims to classify exam questions automatically according to the cognitive levels of Bloom’s taxonomy, addressing the challenge of classifying short text questions. The paper also explores the use of feature selection process, such as Mutual Information, Odd Ratio, and Chi-Square, to enhance the classification process. Additionally, the study evaluates the performance of the proposed mixup model and discusses the potential for further experimentation with other techniques to obtain enhanced results. Overall, the paper contributes to the advancement of automated question classification systems and provides insights into the effective integration of various feature selection methods and classification algorithms to achieve more accurate question classification.

3. SYSTEM REQUIREMENT SPECIFICATION

The system represents a significant leap forward in the realm of exam question classification according to Bloom's Taxonomy, aiming to confront and overcome the inherent challenges linked with manual assessment practices. At its core, this project is dedicated to harnessing the power of an innovative hybrid methodology, which intricately merges the strengths of Neural Network architectures with the robustness of traditional machine learning techniques. By seamlessly integrating these two complementary paradigms, the objective is to engineer an automated question classification system that not only streamlines the arduous task of categorization but also enhances the precision and reliability of the assessment process.

3.1 System Components:

3.1.1. Data Collection and Preprocessing:

- Gathering a diverse dataset of labelled exam questions containing various cognitive levels which are there in Bloom's Taxonomy.
- Pre-processing the data to extract relevant features, including linguistic patterns, keywords, and contextual information.

3.1.2. Traditional Machine Learning Model:

- Employ traditional machine learning algorithms like decision trees, SVM, or logistic regression, to harness the interpretability and explainability of these models.
- Utilize extracted features from the dataset to train the traditional model for initial question categorization.

3.1.3. Neural Network Integration:

- Implement a Neural Network model, possibly using deep learning frameworks like TensorFlow or PyTorch.
- Leverage the Neural Network's capacity to capture intricate relationships within the data, enhancing the model's ability to discern complex patterns associated with Bloom's Taxonomy levels.

3.1.4. Hybrid Model Fusion:

- Combine the outputs of the traditional machine learning model and the Neural Network to create a hybrid classification system.
- Develop a mechanism to dynamically adjust the contribution of each model which depends on the characteristics of the input question, ensuring a balanced and accurate categorization.

3.1.5. Training and Validation:

- Train the hybrid model on the labeled dataset, optimizing for both precision and recall across Bloom's Taxonomy levels.

- Validate the model's performance using cross-validation techniques to ensure robustness and generalizability.

3.1.6. Implementation and Integration:

- Implement the hybrid model within an educational technology platform, allowing seamless integration with existing assessment systems.
- Ensure the system is scalable and adaptable to varying educational contexts and question types.

3.1.7. Applications and Extensions:

- Explore applications beyond question classification, including automated grading, personalized learning recommendations, and detailed performance analysis.
- Consider extending the system to contribute to adaptive learning platforms, question generation tools, and personalized feedback mechanisms.

3.1.8. Evaluation and Iteration:

- Regularly evaluate the system's performance in real-world educational settings, gathering feedback from educators and users.
- Iteratively refine the model based on insights and evolving educational needs.

4. METHODOLOGY / ALGORITHMS USED

4.1 Methodology:

The methodology for automating the classification of exam questions which are based on Bloom's Taxonomy involves a hybrid approach, combining traditional machine learning and Neural Network methods.

4.1.1. Dataset Preparation:

- Curate a labeled dataset of exam questions with categories corresponding to Bloom's Taxonomy levels.
- Include diverse question types and difficulty levels to ensure model robustness.

4.1.2. Feature Extraction:

- Extract relevant features from the exam questions for traditional machine learning, such as word frequency, sentence structure, and linguistic patterns.
- Utilize natural language processing techniques to represent questions in a format suitable for machine learning algorithms.

4.1.3. Traditional Machine Learning Model:

- Implement a traditional machine learning algorithm, possibly using scikit-learn in Python.
- Train the model on the labelled 0dataset, optimizing for accuracy in classifying questions into Bloom's Taxonomy categories.

4.1.4. Neural Network Integration:

- Develop a Neural Network architecture, leveraging frameworks LSTM with BERT transform
- Use word embeddings or other techniques to capture semantic relationships and intricate patterns within the exam questions.

4.1.5. Hybrid Model Fusion:

- Integrate the predictions from the traditional machine learning model and the Neural Network model.
- Establish a weighting mechanism to combine the strengths of interpretability from traditional models with the pattern recognition capabilities of Neural Networks.

4.1.6. Model Evaluation and Optimization:

- Assess the performance of the hybrid model using accuracy metrics accuracy, precision, recall, and F1 score.
-Fine-tune the given model parameters and architecture inorder to improve accuracy and generalization.

4.2 Algorithms Used:

4.2.1. Traditional Machine Learning Algorithm:

1. K-Nearest Neighbours (KNN)

- KNN is a straightforward and intuitive algorithm which is used for classification and regression tasks.

-In classification, when presented with a new data point, KNN identifies the K nearest data points in the training set and assigns the new point the class label that is most common among them.

-In regression, KNN predicts value which is for a new data point by taking the average of the values of its K nearest neighbours.

-The algorithm depends on distance metric, e.g. Euclidean distance, which is used to evaluate the similarity between different data points.

2. Support Vector Machine (SVM)

- SVM is a supervised learning algorithm mainly used for classification and regression tasks.

-For classification purpose, SVM determines the optimal hyperplane which separates classes in the feature space.

-It can manage both linearly separable and non-linearly separable data through various kernel functions, such as linear, polynomial, and radial basis function (RBF).

-SVM focuses on maximizing the margin between support vectors of different classes, enhancing its robustness to outliers.

3. Naive Bayes (NB)

- NB, or Naive Bayes, is a algorithm based on Bayes' theorem, which assumes independence among features.

- NB frequently performs remarkably well in classification tasks, particularly with high-dimensional data.

-It computes the probability of an instance belonging to each class which is based on the probabilities for the individual features.

4. Random Forest (RF)

-RF, or Random Forest, is an ensemble learning method that utilizes decision trees for both classification as well as regression tasks.

-It builds multiple decision trees in training and provides the mode of classes (use for classification) or the prediction (for regression) from the each individual trees.

-Each tree in the random forest is mainly trained on subset which is randomly selected from the training data and features, which helps reduce overfitting and enhances robustness.

-RF is recognized for its high accuracy, scalability, and capability to manage large datasets with high dimensionality.

.

5. FLOW DIAGRAM:

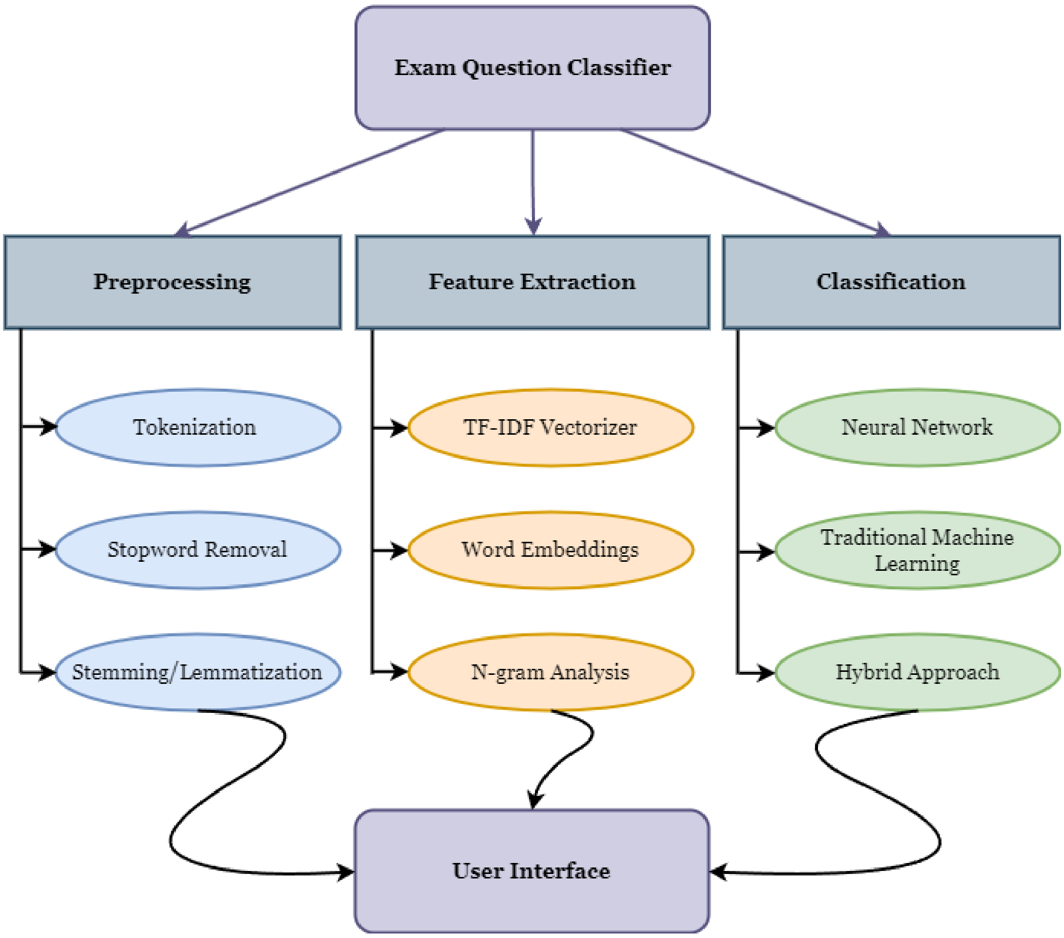


Fig. 5.0

6. MODULE SPLIT-UP:

The diagram represents the modules of an "Exam Question Classifier" system, which contains of three main components: Preprocessing, Feature Extraction, and Classification. Each of these components has further sub-modules.

6.1Preprocessing:

- Tokenization: It break down the text into small small words or tokens.
- Stop Word Removal: Involves removing same words which do not add significant meaning to the text (e.g., "the", "and", "is").
- Stemming/Lemmatization: Reduces words to their base or root form (e.g., "running" to "run").

6.2Feature Extraction:

- TF-IDF Vectorizer: Converts text into numerical values using Term Frequency-Inverse Document Frequency, indicating the importance of a word in a document relative to its frequency across all documents.
- Word Embeddings: Represents words as vectors in a high-dimensional space based on their semantic meaning.
- N-gram Analysis: Analyzes sequences of N consecutive words or characters to capture context and word order.

6.3Classification:

- Neural Network: Neural Network is a machine learning model which similar to the human brain's neural network for pattern recognition and classification tasks.
- Traditional Machine Learning: Standard machine learning algorithms like as decision trees and support vector machine.
- Hybrid Approach: A combination of neural networks and traditional machine learning techniques for improved classification performance.

6.4User Interface:

This module provides an interface for users to interact with the system, inputting exam questions and receiving classified results.

7. DESIGN - UML DIAGRAMS:

7.1 Class Diagram:

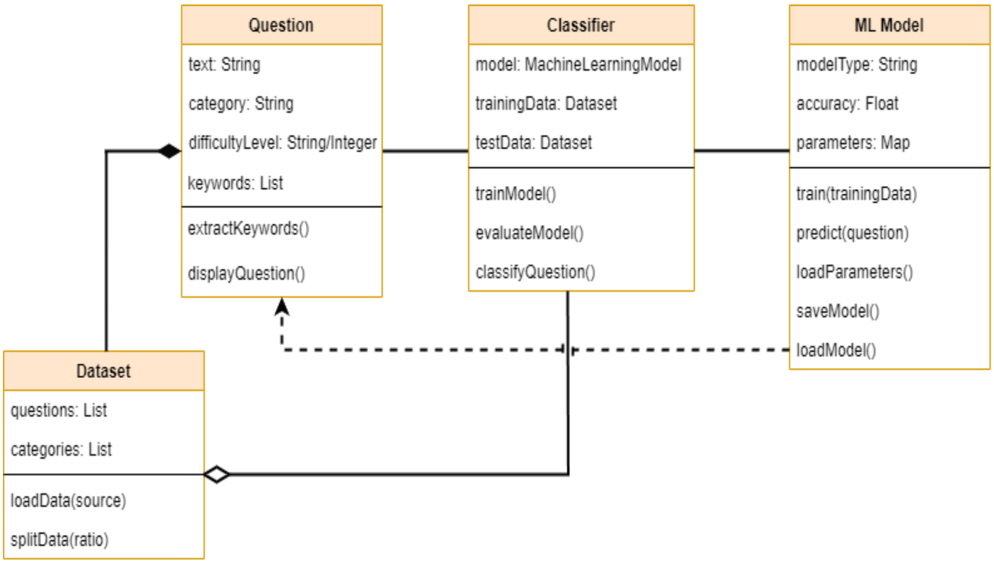


Figure No.7.1

7.1.1: Split-Up:

7.1.1.1 **Class Diagram Overview:**

- The diagram represents a system or software application using Unified Modelling Language (UML).
- It visually depicts the overall structure of the software by showing its classes, attributes, operations , and also the relationships among various objects.
-

7.1.1.2 **Components in the Diagram:**

- **Class "Question":**
  - Represents an individual question in the examination.
  - Attributes:
    - ✓ text: String (stores the question text).
    - ✓ category: String (indicates the question category).
    - ✓ difficultyLevel: String/Integer (denotes the question's difficulty level).
    - ✓ keywords: List (holds relevant keywords for the question).
  - Methods:
    - ✓ extractKeywords(): Extracts keywords from the question.
    - ✓ displayQuestion(): Displays the question.
- **Class "Dataset":**
  - Contains a list of questions and categories.
  - Attributes:
    - ✓ questions: List (stores questions).
    - ✓ categories: List (stores question categories).
  - Methods:

- ✓ loadData(source): Loads data from a source.
- ✓ splitData(ratio): Splits data based on a given ratio.

- **Class "Classifier":**
  - Manages machine learning models and data.
  - Attributes:
    - ✓ model: MachineLearningModel (holds the trained model).
    - ✓ trainingData: Dataset (contains labeled training data).
    - ✓ testData: Dataset (holds unseen test data).
  - Methods:
    - ✓ trainModel(): Trains the model.
    - ✓ evaluateModel(): Evaluates model performance.
    - ✓ classifyQuestion(): Classifies a question.
- **Class "ML Model":**
  - Represents a machine learning model.
  - Attributes:
    - ✓ modelType: String (type of model, e.g., neural network).
    - ✓ accuracy: Float (model accuracy).
    - ✓ parameters: Map (stores model parameters).
  - Methods:
    - ✓ train(trainingData): Trains the model.
    - ✓ predict(question): Makes predictions.
    - ✓ loadParameters(): Loads model parameters.
    - ✓ saveModel(): Saves the model.
    - ✓ loadModel(): Loads a saved model.

7.1.1.3. **Relationships:**

- Lines connecting classes represent associations:
  - For example, the Classifier uses the ML Model for classification.
  - The Dataset provides data to the Classifier.
  - The Question interacts with the Classifier.

7.2 Object Diagram:

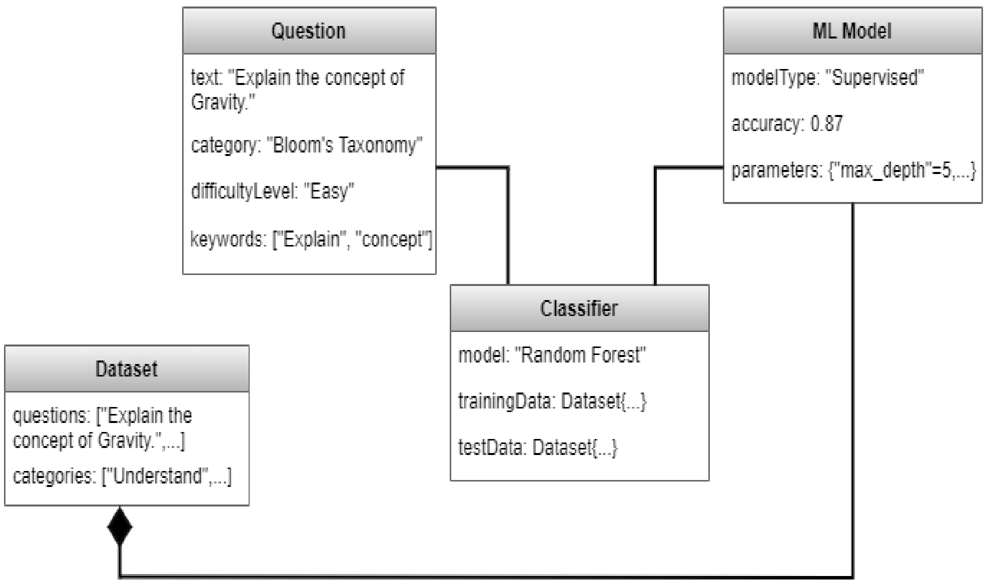


Figure No. 7.2

7.2.1: Split-Up:

7.2.1.1. **Question Object:**
- Represents a specific question.

7.2.1.2. **Dataset Object:**
- Contains a collection of questions and their associated categories.

7.2.1.3. **Classifier Object:**
- Utilizes a machine learning model (here, “Random Forest”) to classify questions.

7.2.1.4. **ML Model Object:**
- Represents the machine learning model used for classification.

7.2.1.5. **Relationships:**
- The Dataset is linked to the Question object, indicating that it consists of multiple questions.
- The Classifier object is connected to both the Dataset and the ML Model, showing that it uses data from the Dataset to train/test the ML Model.

This object diagram illustrates specific instances of classes and their associated attribute values. The relationships demonstrate how objects interact within the system.

7.3 Use Case Diagram:

Figure No. 7.3

7.3.1**: A. Split-Up:**

This use case diagram outlines the relation between Examinees, Instructors, and a System Administrator in the context of exam questions classification using a hybrid approach (combining neural networks and traditional machine learning).

Let's break it down:

7.3.1.1. **Actors:**

- Examinees:
  - They represent students or users who interact with the system indirectly.
  - They have two primary use cases:
    - ✓ View Result: Examinees can view the results of their exams.
    - ✓ View Exam: Examinees can view the exam questions.
- Instructors (Subject Instructors):
  - They are responsible for creating and managing exam questions.
  - Their use cases include:
    - ✓ Uploads Set of Questions: Instructors upload sets of exam questions.
    - ✓ Provide Feedback: Instructors provide feedback on the classification results.
    - ✓ Feature Engineering (Indirect): Instructors contribute to training the ML model by providing features for classification.
- System Administrator:
  - Responsible for managing the system and ensuring its smooth operation.
  - Their use cases include:
    - ✓ Evaluate Confidence: The administrator evaluates the confidence level of the classification results.
    - ✓ Manage Model: The administrator manages the ML model (e.g., retraining, updates).
    - ✓ Assign Bloom's Taxonomy: If confidence is high, the administrator assigns a Bloom's Taxonomy category to the question.
    - ✓ Monitor Activity: The administrator monitors system activity.
    - ✓ View Model: The administrator can view details about the ML model.

7.3.1.2. **Relationships:**

- Include Relationship:
  - The "Assign Bloom's Taxonomy" use case includes the "Traditional Machine Learning" use case.
  - This indicates that traditional ML techniques are applied when confidence is not high.
- Extend Relationship:

- The "Evaluate Accuracy" use case extends the "Classify Question" use case.
- If confidence is not high, the system extends the classification process to evaluate accuracy.

7.3.1.3. **ML Model:**

- Central to the system, the ML model is involved in classification, feature extraction, and evaluation.

This diagram provides a high-level view of the system's functionality and interactions, emphasizing the roles of different actors and their interactions with the ML-based classification system.

7.4 Activity Diagram:

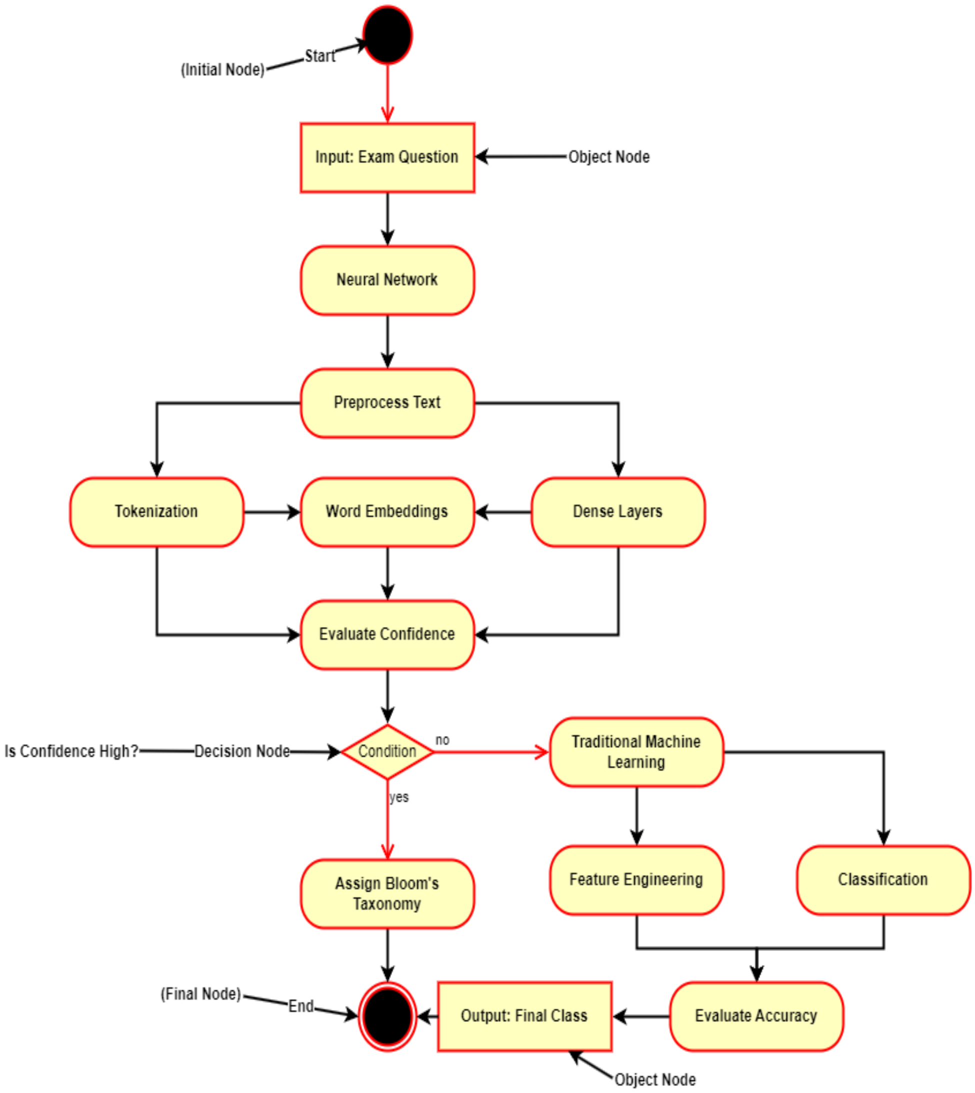


Figure No. 7.4

7.4.1.**: A. Split-Up:**

This Activity Diagram outlines the process of classifying an exam question using a combination of neural networks and traditional machine learning. Let’s break it down:

7.4.1.1. **Input: Exam Question:**

- Represents the starting point.
- The user submits an exam question for classification.

7.4.1.2. **Neural Network:**

- A series of steps involving neural network processing:
    - ✓ Preprocess Text: The question text is preprocessed (e.g., removing stop words, stemming).
    - ✓ Tokenization: The text is split into tokens (words or subwords).
    - ✓ Word Embeddings: The tokens are embedded into vector representations.
    - ✓ Dense Layers: The vectors pass through dense layers for feature extraction.

7.4.1.3. **Evaluate Confidence:**

- A decision point:
    - ✓ If the confidence in the neural network classification is high, proceed to the next step.
    - ✓ If not, move to traditional machine learning.

7.4.1.4. **Traditional Machine Learning:**

- Feature engineering and classification:
    - ✓ Feature Engineering: Additional features are extracted from the question.
    - ✓ Classification: The question is classified using traditional ML techniques.
    - ✓ Evaluate Accuracy: The accuracy of the classification is assessed.

7.4.1.5. **Assign Bloom’s Taxonomy:**

- If confidence is high or traditional ML classification is successful:
    - ✓ The question is assigned a category from Bloom’s Taxonomy (e.g., “Understand,” “Apply,” etc.).

7.4.1.6. **Output: Final Class:**

- Represents the end of the process.
- The classified question is the final output.

This diagram illustrates how both neural networks and traditional ML contribute to question classification, ensuring accurate categorization based on Bloom’s Taxonomy.

7.5 Sequence Diagram:

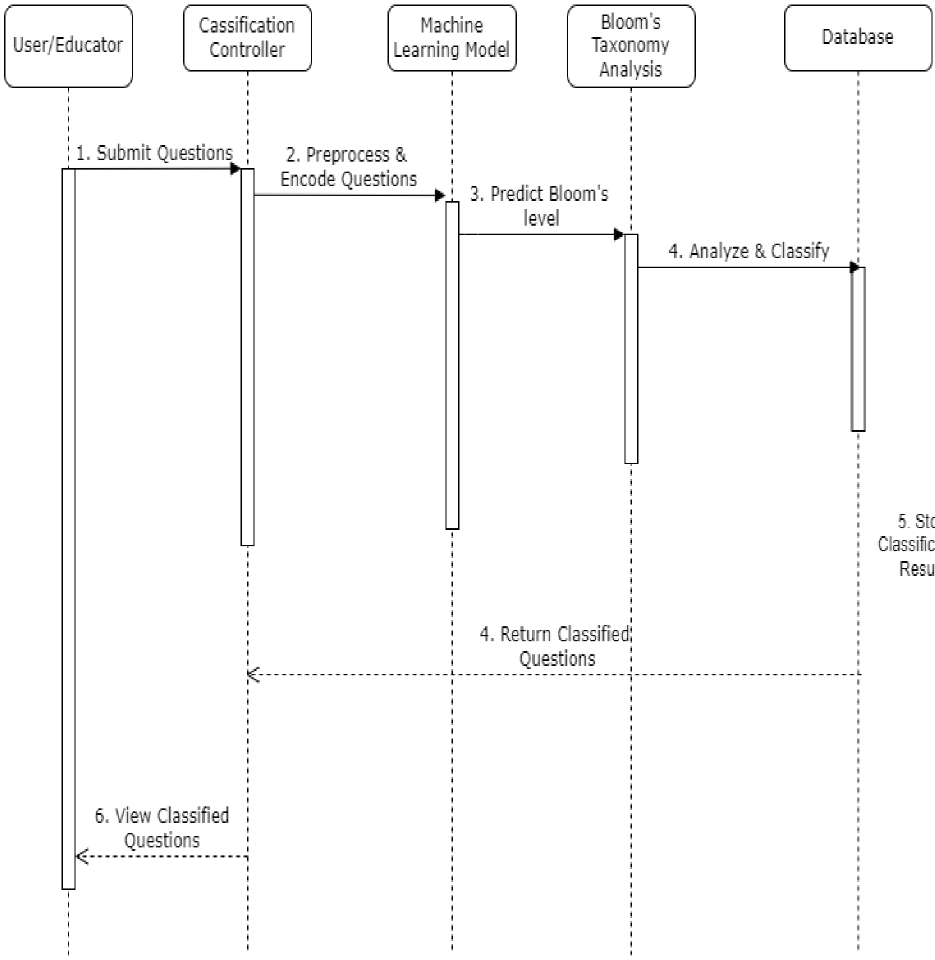


Figure No. 7.5

7.5.1.**: A. Split-Up:**

- Represents the person interacting with the system.
- Initiates the process by submitting exam questions for classification.

7.5.1.2. **Classification Controller:**

- Responsible for handling the classification process.
- Receives the submitted questions.
- Preprocesses and encodes the questions (e.g., tokenization, feature extraction).

7.5.1.3. **Machine Learning Model:**

- Predicts the Bloom's level for each question.
- Utilizes a trained neural network or other machine learning techniques.
- The model has learned from labelled training data.

7.5.1.4. **Bloom's Taxonomy Analysis:**

- Analyses the predicted Bloom's levels.
- Classifies the questions into one of the six levels: 1) Knowledge, 2) Comprehension, 3) Application, 4) Analysis, 5) Synthesis, 6) Evaluation.
- Bloom's Taxonomy provides the framework for this analysis.

7.5.1.5. **Database:**

- Stores the results of the classification.
- Classified questions are associated with their predicted Bloom's levels.

7.5.1.6. **View Classified Questions:**

- Allows the user/educator to access the classified questions.
- The user can review the questions categorized based on the Bloom's Taxonomy.

This sequence diagram outlines the flow of information and interactions between components during the classification process. It demonstrates how machine learning and Bloom's Taxonomy are combined to enhance question classification in educational contexts.

8. SOFTWARE TOOLS / TECHNOLOGIES USED:

The project outlined in the provided content involves the classification of exam questions using a hybrid approach of Neural Networks and traditional machine learning for educational technology. The following software tools and technologies may be utilized:

8.1. **Programming Languages:**

- Python: As a versatile language, Python is commonly used in machine learning projects. It provides various libraries and frameworks suitable for both traditional machine learning and Neural Network implementations.

8.2**. Machine Learning Libraries:**

- Scikit-learn: A comprehensive library which is there traditional machine learning algorithms, offering tools for data preprocessing, model training, and evaluation.
- TensorFlow: These deep learning frameworks are crucial for implementing Neural Networks. They provide high-level abstractions for building and training complex neural models.

8.3**. Database Management:**

- SQLite, MySQL, or PostgreSQL: Depending on project scale and requirements, a relational database management system may be used for storing and managing the labeled dataset of exam questions.

8.4. **Data Analysis and Visualization:**

- Pandas and NumPy: These Python libraries are essential for data manipulation and numerical operations, facilitating data preparation before feeding it into the machine learning models.
- Matplotlib or Seaborn: For visualizing data distributions and model performance, these libraries are commonly employed in the Python ecosystem.

8.5. **Development Environment:**

- Jupyter Notebooks or Visual Studio Code: These development environments offer an interactive and collaborative space, enabling iterative development and easy experimentation with machine learning models.

8.6. **Version Control:**

- Git: Git is mainly Utilized for version control, for collaboration, and keeping the track on changes made in the project codebase.

8.7**. Documentation:**

- Sphinx or Jupyter Notebooks: Tools for creating documentation that explains the project's methodologies, code structure, and results.

8.8. **Collaboration and Project Management:**

- GitHub or GitLab: Platforms for hosting the project repository, facilitating collaboration among team members.
- Trello or Asana: Project management tools to organize tasks, milestones, and team communication.

9. OUTCOMES:

The proposed outcomes of this project encompass several key advancements in the field of educational technology and assessment. Through the innovative hybrid approach integrating Neural Network and traditional machine learning methods for the classification of exam questions which are based on Bloom's Taxonomy, the following outcomes are anticipated:

9.1. **Automation of Exam Question Classification:**

- Successful implementation of the hybrid model is expected to automate and streamline the classification of the exam questions, reducing the reliance on manual efforts and subjective judgment.

9.2. **Enhanced Accuracy and Efficiency:**

- The integration of traditional machine learning features and Neural Network capabilities aims to significantly improve the accuracy and speed of question categorization, providing educators with a more efficient tool for assessing students' cognitive skills.

9.3. **Standardized Assessment Processes:**

- The project seeks to establish a standardized and objective methodology for classifying exam questions, reducing variability in assessment outcomes and promoting fairness in evaluating students' cognitive development.

9.4. **Application in Educational Technology:**

- The outcomes extend to various educational technology applications, including automated grading systems, personalized learning platforms, and in-depth performance analysis tools, contributing to a more adaptive and efficient educational environment.

9.5. **Contribution to Research and Development:**

- The project outcomes are anticipated to support research activities, such as educational data mining and the development of new assessment methods. The model's analysis of cognitive demands in educational materials can give useful insights to ongoing research in the field.

9.6. **Integration into Adaptive Learning Platforms:**

- The hybrid model is expected to contribute to the development of adaptive learning platforms, offering tailored educational experiences for students. Additionally, it may enhance question generation tools and provide personalized feedback to learners.

9.7. **Resource for Educational Researchers:**

- The automated classification capabilities of the model can serve as a valuable resource for educational researchers, facilitating activities such as analyzing cognitive development trends, assessing the effectiveness of teaching methodologies, and refining educational materials

10. Test Cases:

Question No.	Question	Actual Bloom's Level	Calculated Bloom's Level	Result (Pass/Fail)
1	Match the correct data structure to its time complexity for searching (e.g., Hash Table - O(1))?	Remember	Remember	Pass
2	Examine the syntax of a lambda function in Python?	Remember	Remember	Pass
3	Discuss the principles of cybersecurity, explaining concepts such as encryption, authentication, and intrusion detection	Understand	Understand	Pass
4	Compare the different types of file systems (e.g., FAT32, NTFS, ext4) used in operating systems, discussing their features and limitations.	Understand	Understand	Pass
5	Solve a hypothetical scenario involving a data breach in a healthcare organization and propose strategies for mitigating the impact on patient privacy.	Apply	Apply	Pass
6	Analyse the potential risks and benefits of using facial recognition technology for law enforcement and surveillance purposes.	Apply	Apply	Pass
7	A transportation agency is planning a road widening project to accommodate increasing traffic volumes and improve traffic flow. Analyse the traffic management strategies (e.g., lane reconfiguration, signal synchronization, roundabout implementation) and recommend solutions for minimizing disruptions to traffic during construction and maximizing safety for road users.	Analyze	Analyze	Pass
8	Your team is designing a sustainable building envelope for a high-rise office tower to enhance energy efficiency and occupant comfort. Compare the thermal	Analyze	Understand	Fail

	insulation materials (e.g., spray foam insulation, cellulose insulation, rigid foam boards) and recommend solutions for achieving thermal performance targets and reducing heating and cooling loads			
9	Judge the reliability of a finite element analysis (FEA) model in predicting the structural behaviour of reinforced concrete buildings.	Evaluate	Evaluate	Pass
10	Assess the effectiveness of flood mitigation measures in reducing the risk of urban flooding in coastal cities	Evaluate	Evaluate	Pass
11	You are tasked with developing a sustainable transportation infrastructure for a metropolitan city to promote multimodal mobility and reduce traffic congestion. Create an integrated transportation network that includes public transit systems, bicycle lanes, pedestrian walkways, and smart mobility solutions such as ride-sharing and electric scooters, incorporating traffic simulation models, travel demand forecasting, and accessibility analysis for efficient and equitable urban transportation planning.	Create	Create	Pass
12	You are designing a green building certification system for promoting sustainable building practices and reducing environmental impacts in the construction industry. Create a certification framework that evaluates building performance across various sustainability development, incorporating life cycle assessment, green building standards, and stakeholder engagement for holistic building sustainability assessment and certification	Create	Create	Pass

11. Result and Analysis

11.1 Performance Analysis On the basis of F1, Recall, Support, precision and Accuracy. By using TF-IDF Feature Extraction Technique

Sr.no	Ref no	Blooms Taxonomy Level	Precision	Recall	F1 Score	Support
1	**SVM**	1	0.98	0.97	0.97	58
		2	0.93	0.95	0.94	57
		3	0.78	0.89	0.83	44
		4	0.88	0.91	0.90	66
		5	0.98	0.85	0.91	61
		6	0.98	0.98	0.98	65
		Accuracy			**0.93**	**351**
		Macro Avg	**0.92**	**0.92**	**0.92**	**351**
		Weighted Avg	**0.93**	**0.93**	**0.93**	**351**
2	**NB**	1	0.97	0.97	0.97	58
		2	0.94	0.82	0.88	57
		3	0.65	0.70	0.67	44
		4	0.82	0.74	0.78	66
		5	0.80	0.77	0.78	61
		6	0.82	0.95	0.88	65
		Accuracy			**0.83**	**351**
		Macro Avg	**0.83**	**0.83**	**0.83**	**351**
		Weighted Avg	**0.84**	**0.83**	**0.83**	**351**
3	**KNN**	1	0.18	0.97	0.31	58
		2	0.50	0.02	0.03	57
		3	0.86	0.14	0.24	44
		4	0.60	0.27	0.37	66
		5	0.00	0.00	0.00	61
		6	1.00	0.02	0.03	65
		Accuracy			**0.23**	**351**
		Macro Avg	**0.52**	**0.23**	**0.16**	**351**
		Weighted Avg	**0.52**	**0.23**	**0.16**	**351**
4	**RF**	1	0.95	0.98	0.97	58
		2	0.96	0.89	0.93	57
		3	0.80	0.84	0.82	44
		4	0.89	0.89	0.89	66
		5	0.95	0.90	0.92	61
		6	0.94	0.98	0.96	65
		Accuracy			**0.92**	**351**
		Macro Avg	**0.92**	**0.92**	**0.92**	**351**
		Weighted Avg	**0.92**	**0.92**	**0.92**	**351**

Table 11.1

11.2 Accuracy Comparison of Model Using TDF-IDF

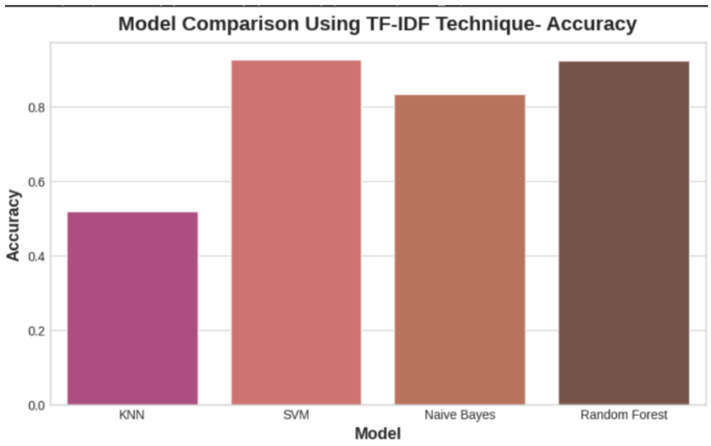


Fig.11.2

11.3 Comparison between model based on evaluation metrices using TF-IDF

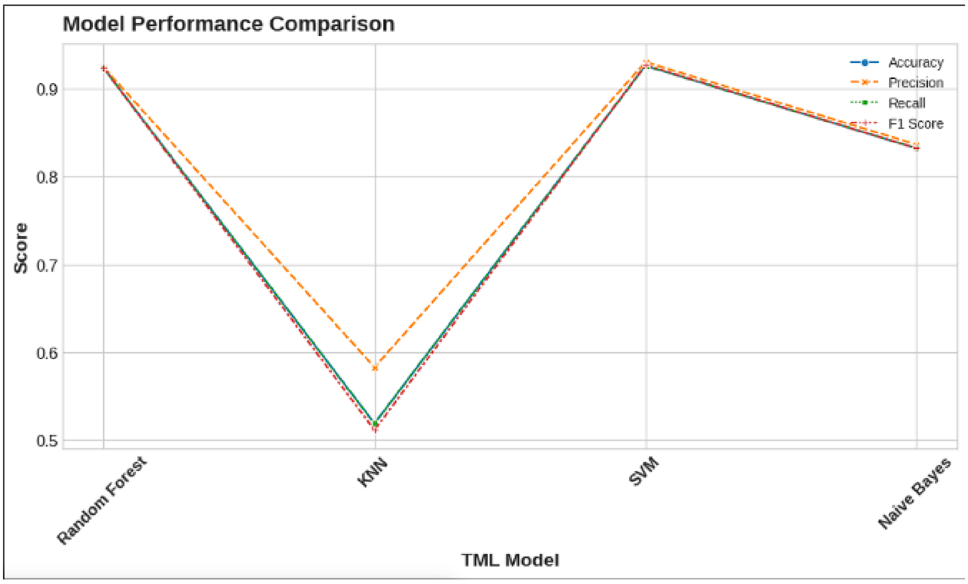


Fig.11.3

11.4 Performance Analysis On the basis of F1, Recall, Support, precision and Accuracy. By using Word-Embeddings and N-gram Feature Extraction Technique

Sr.no	**Ref no**	**Blooms Taxonomy Level**	**Precision**	**Recall**	**F1 Score**	**Support**
1	**SVM**	1	0.29	0.63	0.40	128
		2	0.18	0.25	0.21	123
		3	0.00	0.00	0.00	110
		4	0.00	0.00	0.00	118
		5	0.00	0.00	0.00	107
		6	0.14	0.30	0.19	116
		Accuracy			**0.21**	**702**
		Macro Avg	**0.10**	**0.20**	**0.13**	**702**
		Weighted Avg	**0.11**	**0.21**	**0.14**	**702**
2	**NB**	1	0.87	0.71	0.78	128
		2	0.73	0.64	0.68	123
		3	0.61	0.50	0.55	110
		4	0.63	0.60	0.61	118
		5	0.62	0.68	0.65	107
		6	0.60	0.88	0.72	116
		Accuracy			**0.67**	**702**
		Macro Avg	**0.68**	**0.67**	**0.67**	**702**
		Weighted Avg	**0.68**	**0.67**	**0.67**	**702**
3	**KNN**	1	0.92	0.71	0.80	128
		2	0.61	0.48	0.54	123
		3	0.63	0.65	0.64	110
		4	0.41	0.43	0.42	118
		5	0.44	0.66	0.53	107
		6	0.69	0.63	0.66	116
		Accuracy			**0.59**	**702**
		Macro Avg	**0.62**	**0.59**	**0.60**	**702**
		Weighted Avg	**0.62**	**0.59**	**0.60**	**702**
4	**RF**	1	0.83	0.76	0.79	128
		2	0.48	0.50	0.49	123
		3	0.48	0.50	0.49	110
		4	0.45	0.51	0.48	118
		5	0.57	0.39	0.46	107
		6	0.40	0.47	0.43	116
		Accuracy			**0.53**	**702**
		Macro Avg	**0.53**	**0.52**	**0.52**	**702**
		Weighted Avg	**0.54**	**0.53**	**0.53**	**702**

Table 11.4

11.5 Accuracy Comparison Of Models Using Word Embeddings and N-gram.

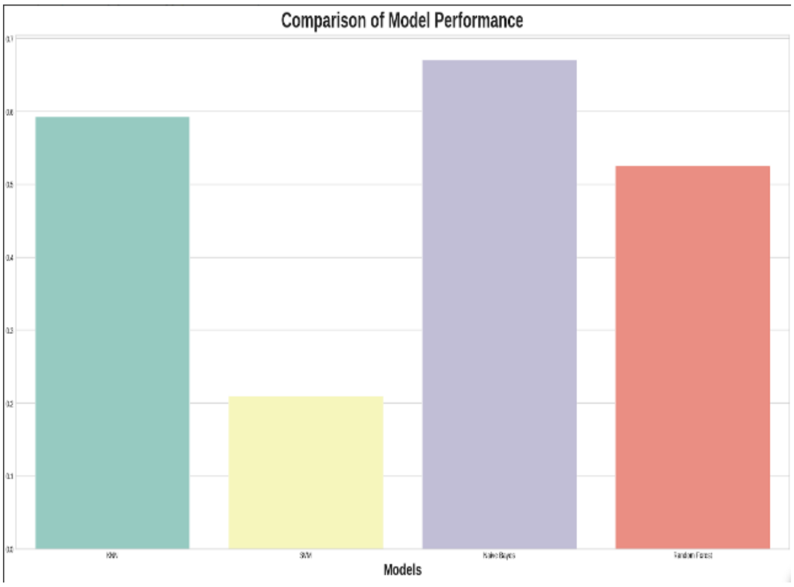


Fig.11.5

11.6 Comparison between model on the basis of evaluation metrices using word embedding & n-gram

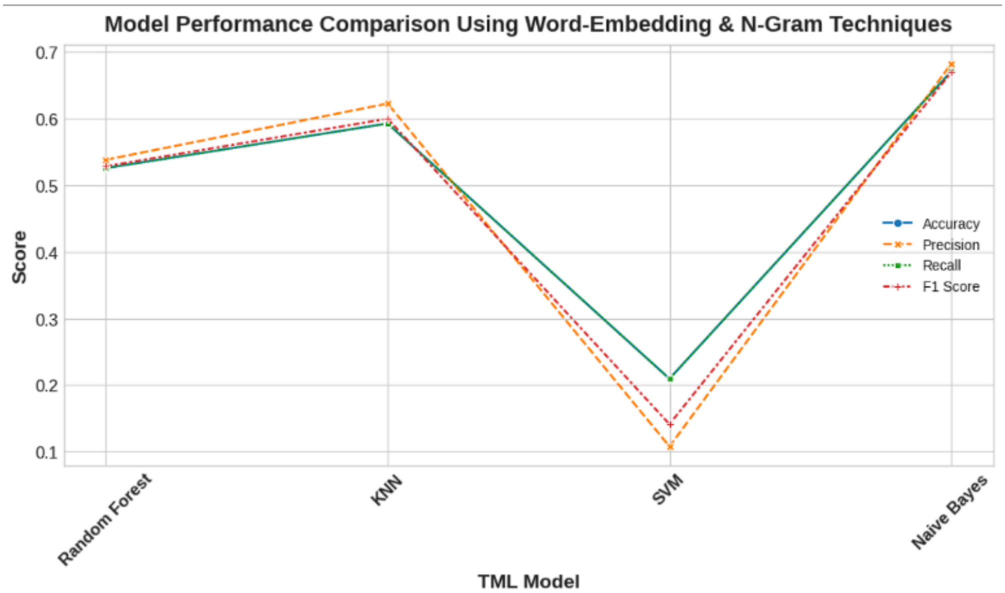


Fig11.6

11.7 Bert Model Results

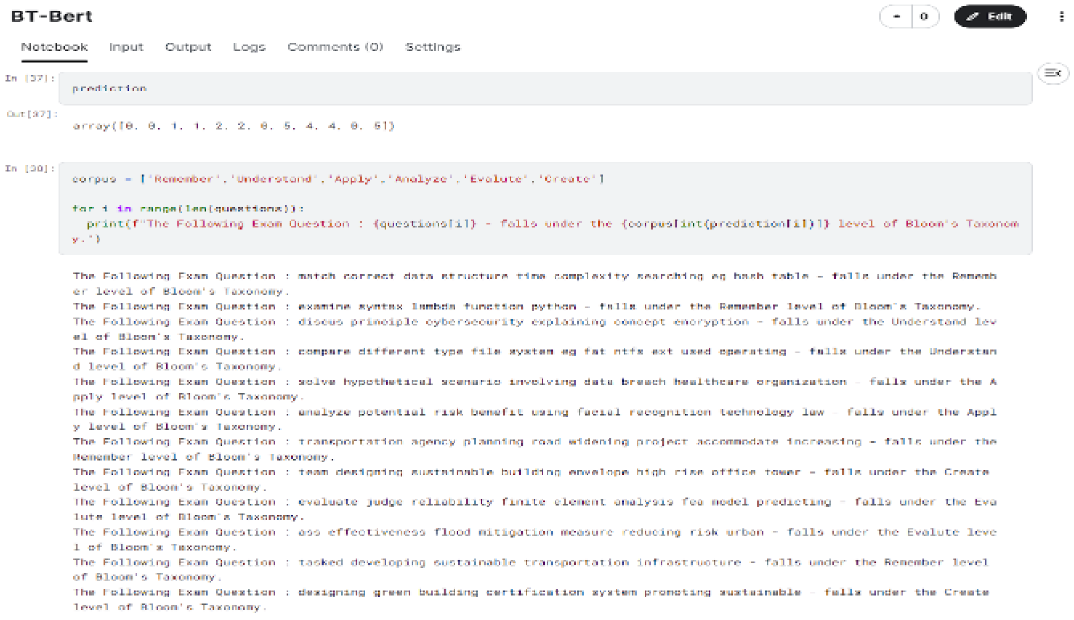


Fig 11.7

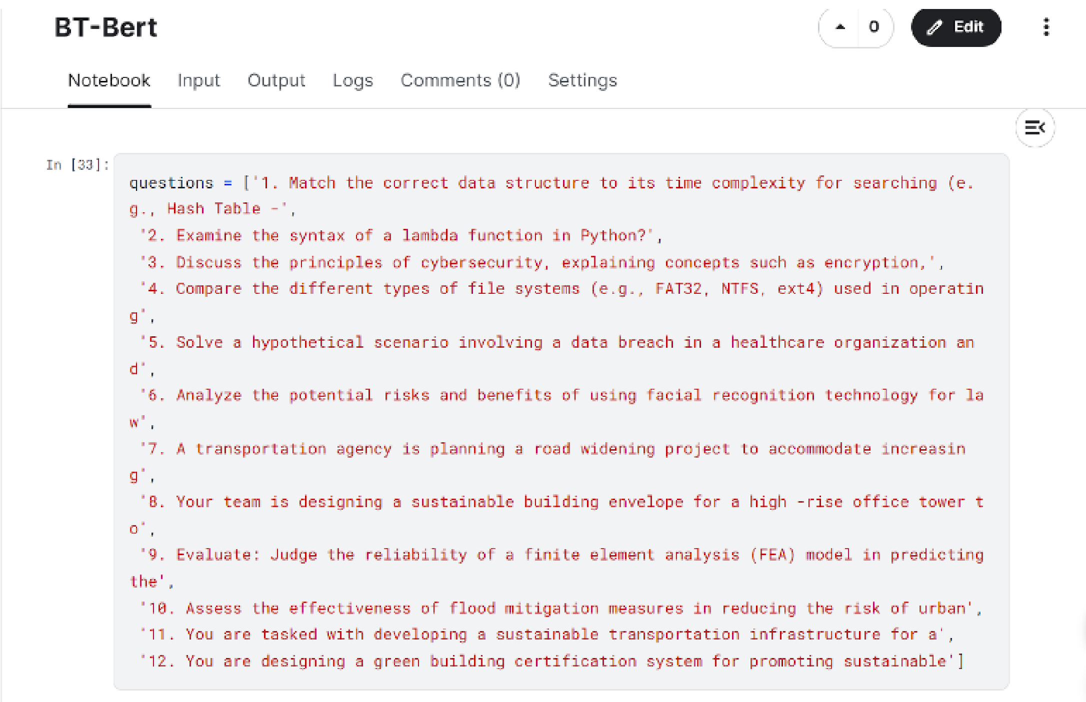


Fig.11.7

11.8 Extracted Questions From Input PDF

```
['1. Match the correct data structure to its time complexity for searching e.g., Hash Table -',
'2. Examine the syntax of a lambda function in Python?',
'3. Discuss the principles of cybersecurity, explaining concepts such as encryption,',
'4. Compare the different types of file systems (e.g., FAT32, NTFS, ext4) used in operating',
'5. Solve a hypothetical scenario involving a data breach in a healthcare organization and',
'6. Analyze the potential risks and benefits of using facial recognition technology for law',
'7. A transportation agency is planning a road widening project to accommodate increasing',
'8. Your team is designing a sustainable building envelope for a high -rise office tower to',
'9. Judge the reliability of a finite element analysis (FEA) model in predicting the structural',
'10. Assess the effectiveness of flood mitigation measures in reducing the risk of urban',
'11. You are tasked with developing a sustainable transportation infrastructure for a',
'12. You are designing a green building certification system for promoting sustainable',
"13. Examine and construct MST using Prim's Algorithm. Assume the following graph as an",
'14. Identify the shortest path from Node 0 to all other Nodes in the following graph using']
```

Fig.11.8

11.9 Data Before and After Processing

index	Labels	Exam Questions	text_clean
0	1	State the significance of the keyword "volatile" in Java programming?	state significance keyword volatile java programming
1	1	Recognize the purpose of the "extern" keyword in C programming?	recognize purpose extern keyword programming
2	1	Examine the difference between static and dynamic linking in software development?	examine difference static dynamic linking software development
3	1	Identify the characteristics of a Singleton design pattern?	identify characteristics singleton design pattern
4	1	Recall the syntax for declaring a constant pointer to a constant integer in C++?	recall syntax declaring constant pointer constant integer
5	1	Reproduce the steps to implement a depth-first search algorithm for a graph?	reproduce steps implement depthfirst search algorithm graph
6	1	Quote the definition of a semaphore in operating systems?	quote definition semaphore operating systems
7	1	Describe the purpose of the "this" pointer in C++?	describe purpose pointer
8	1	List the advantages of using a linked list over an array in certain scenarios?	list advantages using linked list array certain scenarios
9	1	Enumerate the differences between a shallow copy and a deep copy of an object?	enumerate differences shallow copy deep copy object
10	1	Recognize the potential issues caused by race conditions in multithreaded programming?	recognize potential issues caused race conditions multithreaded programming
11	1	Recall the purpose of the "virtual" keyword in C++?	recall purpose virtual keyword
12	1	Tabulate the differences between a heap and a stack in memory allocation?	tabulate differences heap stack memory allocation
13	1	Retell the steps to implement a binary search tree traversal without recursion?	retell steps implement binary search tree traversal without recursion
14	1	State the output of a given regular expression applied to a string?	state output given regular expression applied string
15	1	Identify the differences between an interface and an abstract class in Java?	identify differences interface abstract class java
16	1	Describe the purpose of the "volatile" keyword in concurrent programming?	describe purpose volatile keyword concurrent programming
17	1	Match the correct data structure to its time complexity for searching (e.g., Hash Table - O(1))?	match correct data structure time complexity searching eg hash table
18	1	Examine the syntax of a lambda function in Python?	examine syntax lambda function python
19	1	Recall the function of the "super" keyword in Java inheritance?	recall function super keyword java inheritance

Fig.11.9

11.10 Classification Results of Extracted Question

```
input_data_features = loaded_vectorizer.transform(questions)
# print(input_data_features.shape)
predict = []

for i in questions:
  prediction = loaded_model.predict(input_data_features)
  predict.append(prediction[0])

print(prediction)

corpus = ['Remember','Understand','Apply','Analyze','Evalute','Create']

for i in range(len(questions)):
  print(f"The Following Exam Question : {questions[i]} - falls under the {corpus[int(prediction[i])-1]} level of Bloom's Taxonomy.")
```

```
[1 1 2 2 3 3 4 2 5 5 6 6 3 3]
The Following Exam Question : 1. Match the correct data structure to its time complexity for searching (e.g., Hash Table - - falls under the Remember level of Bloom's Taxonomy.
The Following Exam Question : 2. Examine the syntax of a lambda function in Python? - falls under the Remember level of Bloom's Taxonomy.
The Following Exam Question : 3. Discuss the principles of cybersecurity, explaining concepts such as encryption, - falls under the Understand level of Bloom's Taxonomy.
The Following Exam Question : 4. Compare the different types of file systems (e.g., FAT32, NTFS, ext4) used in operating - falls under the Understand level of Bloom's Taxonomy.
The Following Exam Question : 5. Solve a hypothetical scenario involving a data breach in a healthcare organization and - falls under the Apply level of Bloom's Taxonomy.
The Following Exam Question : 6. Analyze the potential risks and benefits of using facial recognition technology for law - falls under the Apply level of Bloom's Taxonomy.
The Following Exam Question : 7. A transportation agency is planning a road widening project to accommodate increasing - falls under the Analyze level of Bloom's Taxonomy.
The Following Exam Question : 8. Your team is designing a sustainable building envelope for a high -rise office tower to - falls under the Understand level of Bloom's Taxonomy.
The Following Exam Question : 9. Judge the reliability of a finite element analysis (FEA) model in predicting the structural - falls under the Evalute level of Bloom's Taxonomy.
The Following Exam Question : 10. Assess the effectiveness of flood mitigation measures in reducing the risk of urban - falls under the Evalute level of Bloom's Taxonomy.
The Following Exam Question : 11. You are tasked with developing a sustainable transportation infrastructure for a - falls under the Create level of Bloom's Taxonomy.
The Following Exam Question : 12. You are designing a green building certification system for promoting sustainable - falls under the Create level of Bloom's Taxonomy.
The Following Exam Question : 13. Examine and construct MST using Prim's Algorithm. Assume the following graph as an - falls under the Apply level of Bloom's Taxonomy.
The Following Exam Question : 14. Identify the shortest path from Node 0 to all other Nodes in the following graph using - falls under the Apply level of Bloom's Taxonomy.
```

Fig 10.10

12. References

1. https://link.springer.com/chapter/10.1007/978-981-13-9942-8_38
2. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7081997/
3. https://ijisae.org/index.php/IJISAE/article/view/3860
4. https://www.researchgate.net/publication/347254153_Exam_Questions_Classification_Based_on_Bloom%27s_Taxonomy_Approaches_and_Techniques
5. https://www.researchgate.net/publication/339775553_CLASSIFYING_EXAM_QUESTIONS_BASED_ON_BLOOM%27S_TAXONOMY_USING_MACHINE_LEARNING_APPROACH
6. https://www.jatit.org/volumes/Vol78No3/15Vol78No3.pdf

13.Annexure 1

13.1 Plagiarism Report

turnitin

Similarity Report ID: oid:28480:59674908

PAPER NAME

blackbookv4.pdf

WORD COUNT 8820 Words	CHARACTER COUNT 54553 Characters
PAGE COUNT 50 Pages	FILE SIZE 1.7MB
SUBMISSION DATE May 21, 2024 2:22 PM GMT+5:30	REPORT DATE May 21, 2024 2:23 PM GMT+5:30

13% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

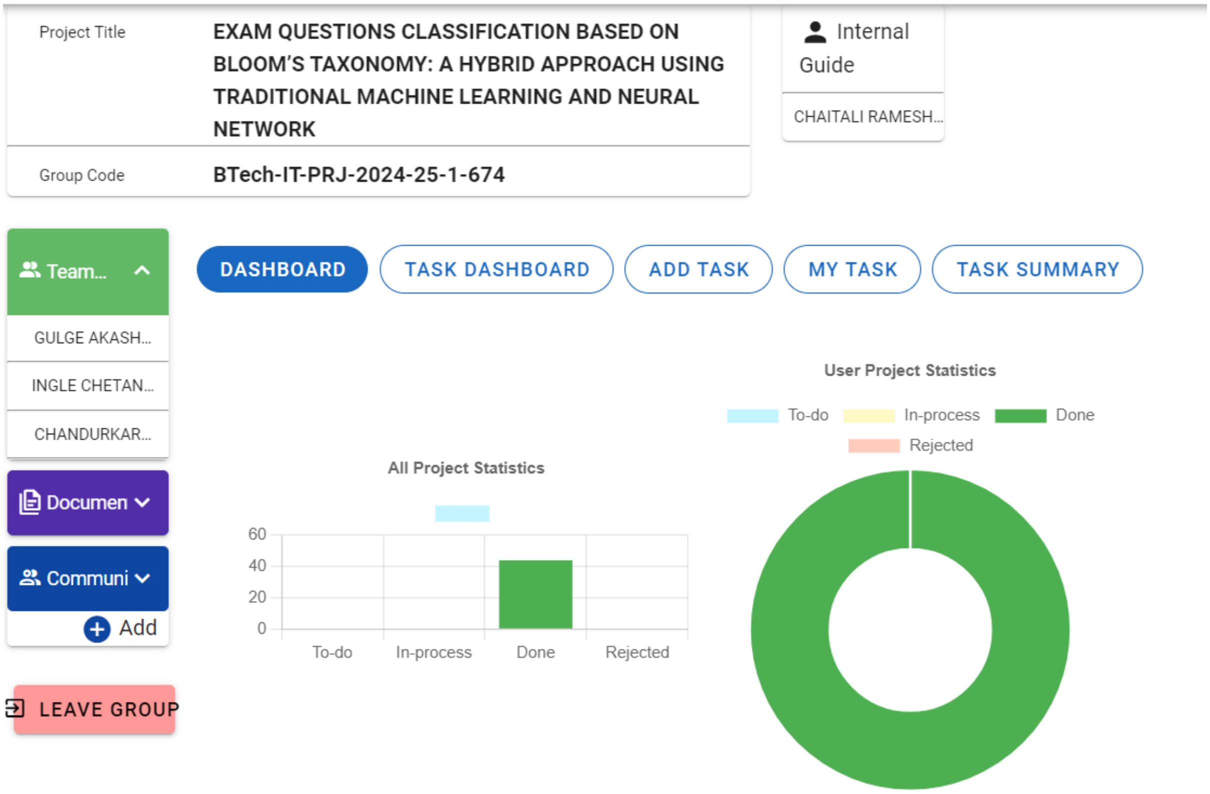
- 9% Internet database
- 6% Publications database
- Crossref database
- Crossref Posted Content database
- 10% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Quoted material
- Cited material
- Small Matches (Less then 10 words)

Summary

14.Annexure -II



15.Research Paper

EXAM QUESTIONS CLASSIFICATION BASED ON BLOOM'S TAXONOMY: A HYBRID APPROACH USING TRADITIONAL MACHINE LEARNING AND NEURAL NETWORKS

[1]Mr. Shreyas N. Chandurkar
Department of Information Technology
BRACT Vishwakarma Institute of Information Technology Pune, India
shreyas.22120264@viit.ac.in

[2]Mr. Chetan G. Ingle
Department of Information Technology
BRACT Vishwakarma Institute of Information Technology Pune, India
chetan.22010416@viit.ac.in

[3]Mr. Akash G. Gulge
Department of Information Technology
BRACT Vishwakarma Institute of Information Technology Pune, India
akash.22010608@viit.ac.in

[4]Prof. Chaitali Shewale
Prof. Department of Information Technology
BRACT Vishwakarma Institute of Information Technology Pune, India
chaitali.shewale@viit.ac.in

Abstract – This study presents a groundbreaking solution to the arduous task of classifying exam questions according to Bloom's Taxonomy, a cornerstone of educational assessment. Manual classification, prone to subjectivity and time constraints, is revolutionized through the introduction of an innovative hybrid approach, amalgamating Neural Network and traditional machine learning techniques. The primary objective is to automate question categorization, furnishing educators with a more efficient and impartial means to evaluate students' cognitive abilities. Leveraging machine learning augments the precision and expediency of this categorization process.

The hybrid model intricately intertwines the interpretability of traditional machine learning models with the nuanced pattern recognition capabilities of Neural Networks. Methodologically, the model is trained on a meticulously labeled dataset of exam questions, harnessing features extracted from traditional machine learning algorithms such as K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Naive Bayes (NB), and Random Forest. Notably, the Random Forest algorithm demonstrates superior accuracy among the ensemble.

The implications of this hybrid model reverberate across diverse educational domains. Its applications span from automated grading systems to personalized learning platforms and detailed performance analytics. Within educational technology, the model's integration promises advancements in adaptive learning environments, question generation tools, and tailored feedback mechanisms. Furthermore, its utility extends to research and development endeavors, facilitating tasks such as educational data mining, the formulation of novel assessment methodologies, and the scrutiny of cognitive demands within educational materials.

In essence, this hybrid approach transcends the limitations of manual classification, heralding a new era in educational technology and research methodologies. It furnishes educators with a standardized and efficacious tool, fostering a deeper comprehension of students' cognitive progression. The model's automation capabilities streamline assessment procedures, furnishing a valuable resource for educators and researchers alike, thereby revolutionizing the educational landscape.

I. INTRODUCTION

1.1 What is Blooms Taxonomy?:

Bloom's Taxonomy is a hierarchical framework that classifies educational objectives and cognitive skills into a structured hierarchy. It was developed by educational psychologist Benjamin Bloom in the 1950s and later revised by Anderson and Krathwohl in 2001. The taxonomy organizes cognitive skills into six levels, ranging from lower-order thinking skills to higher-order thinking skills:

1. **Remembering**: Recalling facts, information, or concepts.
2. **Understanding:** Explaining ideas or concepts and demonstrating comprehension.
3. **Applying:** Using information or knowledge in a new situation or context.
4. **Analyzing**: Breaking down information into components, identifying patterns, and understanding relationships.
5. **Evaluating:** Making judgments based on criteria and standards.

English (United States)