Assistive tool for the Evaluation of Online Exam Papers in Tertiary Education

Nuwan Perera
Faculty of Graduate Studies and Research,
Sri Lanka Institute of Information Technology

Dasuni Nawinna
Department of Computer Systems Engineering,
Sri Lanka Institute of Information Technology

ABSTRACT

Setting exam papers is a strenuous and time-consuming task for academics in the higher education. Teachers must consider a range of aspects such as the academic year of class, the learning objectives of the course, the duration of exam, level of difficulty in questions, and the scoring system to ensure there is no bias, discrimination, or prejudice. At present, the assessments are more frequently conducted through online platforms. To ensure the quality of assessment, higher education institutions enforce standards that are evaluated through the moderation process. Currently, the moderation of exam papers is carried out manually by senior academics. This is time consuming task as it requires referring several documents to verify quality aspects of the exam paper. This paper presents the development of an automated system to facilitate the educators to evaluate their online exam paper's quality. Initially, an online survey has been carried out among the educators. The system is based on the Bloom's Revised Taxonomy model, Natural Language Processing, and Python for developing task and HTML and CSS for template creating. The system makes the process of paper setting and moderation more efficient and facilitate producing more quality assessments in higher education domain.

General Terms

Natural Language Processing, eLearning

Keywords

eLearning, Education, Assessment, Quality, Paper setting, Moderation, NLP, Blooms Taxonomy

1. INTRODUCTION

The COVID-19 pandemic has completely shifted the nature of higher education. According to current world situation digital teaching and learning help for every lecture and every student. Digital teaching and learning are based on computerized environment with new technologies in the world. There are lot of support things (technologies, methods, tools) to them. For instance, google classroom, zoom app, various content management systems and learning management systems, websites, flexible data quota, smart devices and strong network and speed internet facilities. The digital examinations are very important thing for students and educators in various way. In above mentioned situation the possibility of exam cancellations and postponements has had a significant negative effect on the traditional method of student evaluation. Because educators had to transform to the online method. Even though in higher education of around the world, online exams are frequently used to evaluate students. However, exam paper creator should be covered related learning outcomes through the exam paper. In this study Researcher has found the manual processes used by educators for check covered content parts of the course unit and using models for question patterns. By this study the automation system of online exam questions evaluation tool has been

derived. Researchers use bloom's revised taxonomy model as quality measure standard and use Natural Language Processing (NLP) for reach the project outcomes. Researcher selected python as main programming language with Django framework and use HTML/CSS for create interfaces. Propose framework has two functions in the development stage. First function of this system reveals the bloom's revised taxonomy level of each question. It will be helpful to educator to decide suitable questions or not at the considered semester. Then find covered percentage of the course module by the exam paper. This may be helpful for educator to reconsider the exam paper. This system automatically checks the grammar correction of each question. Here researcher used Natural Language Processing to fulfill the above functions. System architecture will be explained with the coding.[1]. This paper is structured as follows: Section II provides details on the literature survey. Section III explains the proposed framework and system. In Section IV the details about using technologies as methodology part. In Section V Challenges of the System VI followed conclusions about the Research. Finally mentioned Acknowledgement and References.

2. LITERATURE REVIEW

The study of D. Dayananda, K. Chathumini and S. Vasanthapriyan investigated under the topic of A Novel Framework for Online Exams during the Pandemic of COVID-19: Evaluation Methods, Students' Priorities and Academic Dishonesty in Online Exams. This study suggests a thorough framework for online exams as a timely solution for students' evaluation. This impacts how online tests are graded, students' priorities, and academic dishonesty cite [1]. They have represented four modules with ARENA simulation, which are Exam Registration Module, Identity Verification Module, Student Attend in Online Exam Module and Continuous Monitoring Module [1]. With the use of ARENA simulation, the suggested framework's modules were assessed to see if they could be implemented and if they would accurately reflect the validity. Their suggested framework can deliver proactive responses for online exams that offer thorough assistance for organizing an online exam with more protection over the dependability and integrity of online platforms. Additionally, they have revealed the online examination methods that pupils prefer to attend in this format. Finally, they say that the major focusing areas can be separately investigated to deliver more comprehensive solutions.

R. Rajesh and R. Kanimozhi said by their paper at the present time evaluating an exam paper and declaring result in a restricted period is a difficult task for educational schools, colleges, institutions, departments and Universities [15]. Thus, manual exam paper correction becomes more difficult, and many fraudulent activities are happened today. To make it easier and more accurate the proposed aim is to develop a software for automatic exam paper evaluation and grading

system, the system works by scanning the handwritten written exam papers then the scanned image be improved into an editable text using OCR tool and the evaluation will perform by matching the key terms which is maintained in the template. It is entirely integrated approach upon dissimilar level of knowledge by the method of examination, evaluation, result and formulation of subject papers. In a field of education though teaching, evaluation, and the performance method many organizations initiated with the use of advanced technologies. The approach of evaluating the examination papers is evolved although use of computer wherever the utilization of computer is obligatory in every varied technique for analysis. This paper structured in different patterns such as offline, online and the manual exam paper evaluation with different analysis and techniques [2].

The study of Bhavna Galhotra and Devesh Lowe investigated under the topic of AI Based Examination System: A Paradigm Shift in Education Sector [14]. By reading this paper, we can see how the transition to online teaching and assessment was seen by many educators as a paradigm change in pedagogy in the field of education. Many found it unsatisfying to teach classes online, but it was acknowledged that this was the only practical option at the time. The identified issue that will be covered in this paper relates to online options for proctored exams using artificial intelligence, online objective-type assessment forms, and pre-assigned rubrics for grades and evaluation. They have mentioned about how the questions that were created by the work's online evaluation are like a great window of light. More specifically, administering online tests using proctored live tests or submitting evaluation forms has evolved into the standard. Due to a situation, the proctor was required to be present and watch the student's conduct, their motions, and the application to see if the student's face could be recognized using the data already submitted, if there was any noise, and whether anyone else was in the room. We have identified by reading this paper, all these elements have improved the efficiency and effectiveness of the online examination system, enabling universities to administer exams conveniently and via remote access. Finally, we can say this paper's primary goal is to give research of several online assessment technologies made available by technology businesses for efficient exam administration. This study also offers a comparison of the features of these tools and how they are used in various fields.

The study of Ganga Sanuvala and Syeda Sameen Fatima investigated under the topic of A study of automated Evaluation of Student's Examination paper using Machine Learning Techniques [12]. This research introduces the design and implementation of Handwritten Answer Evaluation (HAES) system for student exam papers [3]. The HAES is an automatic response assessment system that makes it possible to recognize text in answer sheets and can determine the grade of each response using the model's prior knowledge. In this study, Optical Character Recognition (OCR) tool is used to extract the text from human written scanned answer script and natural language processing (NLP) techniques are used for grading the answer sheets [3]. They have proposed, students' textual work is divided into two main sections: high grade documents and low-grade documents and the widely used models like Naïve Bayes (NB) [4], Logistic Regression (LR) [5], Support Vector Machine (SVM) [6], and Gradient Boost Decision Tree(GBDT) [7] are considered to conduct experiment.

3. APPROACH

Initial stage researcher collected data through online survey and physical meetings. In this occasion researcher obtained valuable details from few of educators in tertiary level and local school teachers. This was focus to obtaining ideas about quality, structure and use standards of exam question papers. It was more valuable ideas from tertiary level academics. Researcher have identified top level benefits of NLP which would help for many sectors' performance. Automate processes in real-time, perform large-scale analysis, and Tailor NLP tools for industry are revealed benefits in this study. According to the summary of the survey researcher propose a system for the task of question paper pre-evaluation. It has two types of evaluation functions with friendly template. First task is obtained BRT level for each Question and another task is matching Course content with exam paper questions and present similarity percentage. Both tasks will generate output values for the user to take decision about their exam question paper. Both models will generate reports with related values for the user to take decision about their exam questions. Below diagram (Fig.1) shows each step of the proposed system.

The proposed framework is derived from data collected through an online survey. This was focus of obtaining a practical overview of online exams. According to the summary of the survey researcher propose a framework for evaluate the exam papers automatically. Initially, shows Below diagram as the architecture of the system.

The main problem of manually evaluation process is very time consuming one and boring method. Sometime missing factors are there in the course unit. Other thing is the paper must be flow with low cognitive level to high cognitive level using some standard model.

The aim of this research is to automatically analyze exam paper question quality. Here researcher use two models as main tasks. One is use Bloom's Revised Taxonomy for evaluate each Question and another one is matching Course Module outline with exam paper questions. Both models will generate reports for the user to take decision about their exam questions.
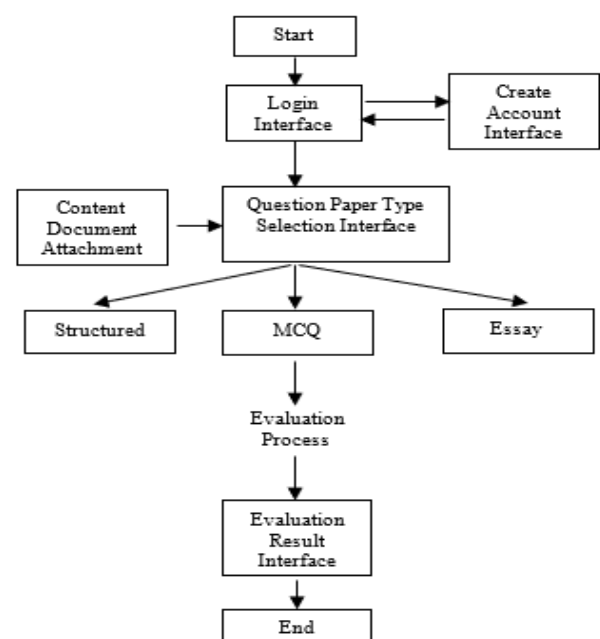

Fig. 1. Initial System Flow

The proposed framework has sign-in and sign-up interface, which is use for entering the system. Then user can select their option as MCQ or Structured or Essay type. After that system will appear user friendly interface which has text boxes to enter questions and marks. Then user need to click Evaluation button for generate evaluation reports. User can upload course module outline document for reach a higher evaluation task. The system has main four interfaces for the process which are Login (user login and create account), User Dashboard, Paper creation and Evaluation and Report and suggestion.
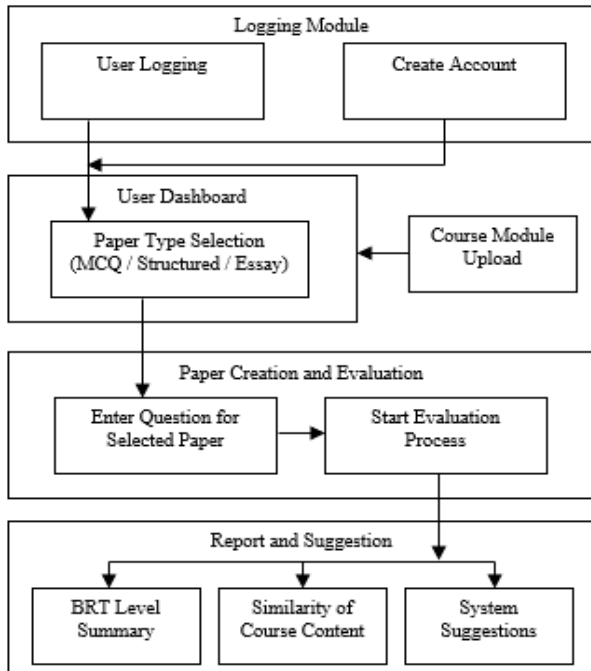

Fig. 2. User Interfaces

At the initial stage researcher had found the necessity of the research topic by collecting data from educators in various universities using virtual meetings and google form. Here researcher had various comments and instructions. From those researchers have taken brief idea for the project. Basically, created interfaces and validation process using HTML and CSS. At the user dashboard user can start exam paper creation and evaluation process. Here researcher apply Bloom's Revised Taxonomy model as standard question quality measuring method. Researcher have found that the specific levels are important for students of each year of the degree. Propose framework will present those things as system suggestions at the end of evaluation process. Researcher matches the module outline document with each question and reveal covered percentage of the course module by the proposedframework. This task uses Natural Language Processing for successful outcome.

3.1 Bloom's Revised Taxonomy – BRT

To create engaging quizzes, the lecturer frequently uses Bloom's Taxonomy to identify their question. To enhance the correlation between cognitive levels and learning activities in the twenty-first century, Bloom's taxonomy was changed in 2001 and released under the name Revised Bloom's Taxonomy. As seen in Fig. 3, the Revised Bloom's Taxonomy hierarchy likewise has six levels, with the difference being that the noun's category changes to become a verb.

Remember - Recall or recognize information, and ideas from related course unit. Lecturer should provide verbal or written texts about the subject that can be answered by recalling the information the student has learned.

Understand - Understand the main idea of material heard, viewed, or read. Interpret or summarize the ideas in own words. Ask questions that the student can answer in his/her own words by stating facts or by identifying the main idea.

Apply - Use a practical situation to put an abstract concept to use to solve an issue or connect it to prior knowledge. The lecturer must give the student the chance to apply concepts, theories, or problem-solving strategies to fresh circumstances. Check the student's work to make sure they are independently applying problem solving strategies. Provide questions that call on the student to identify and resolve issues.

Analyze - Divide a thought or idea into its component elements and demonstrate the connections between them. The lecturer should give the class time to break down concepts and ideas into their component elements. Asking students to defend their decision to use a certain approach to an issue and the reason their solution works.

Evaluate- Make intelligent decisions on the worth of concepts or resources. To boost beliefs and ideas, use standards and criteria. The lecturer should give students the chance to evaluate information using relevant standards. Ask students to give examples of how they have used standards and criteria to evaluate, analyze, or interpret processes, materials, procedures, etc.

Create- Assemble knowledge's components into a whole and create connections for novel circumstances. The lecturer should provide the students the chance to put pieces of knowledge together through creative problem solving and thinking outside the box. Students should be asked to show that they can combine ideas to create new ones for different circumstances [10].
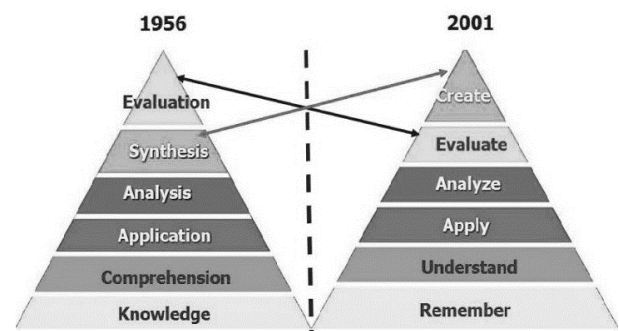

Fig. 3. Bloom's Taxonomy and Revised

3.2 Natural Language Processing - NLP

Natural Language Processing (NLP) is a subject of AI, and it works with combining human and machine languages [11]. This is the computational study of linguistic data, most frequently in the form of textual data like papers or publications [8]. By utilizing linguistic knowledge, natural language processing aims to create a representation of the text that adds structure to the unstructured natural language [8]. NLP enables to program computers to process and analyze large amounts of natural language data. Speech interpretation, natural language understanding, and natural language generation are the most complicated tasks of Natural

Language Processing. Computer translation, question-answering, automated text summarization, answer script evaluation, and other applications use NLP extensively [13].

According to the Fig.4 NLP techniques has four major steps which are dependent on the applications like translation or a grammatical correction, or a generated response based on rules or training data. For tackling NLP tasks, a variety of tools and libraries are available in the Python programming language. The Natural Language Tool Kit (NLTK), an opensource collection of libraries, tools, and educational resources for developing NLP programs, contains several of them. Libraries for many of the mentioned NLP tasks, as well as libraries for subtasks like sentence parsing, word segmentation, stemming, and lemmatization, are all part of the NLTK. specifically for tokenizing words, phrases, sentences, paragraphs, and chapters to improve the computer's understanding of the text. Additionally, it has libraries for implementing functions like semantic reasoning, which allows users to generate logical inferences from text-based evidence.
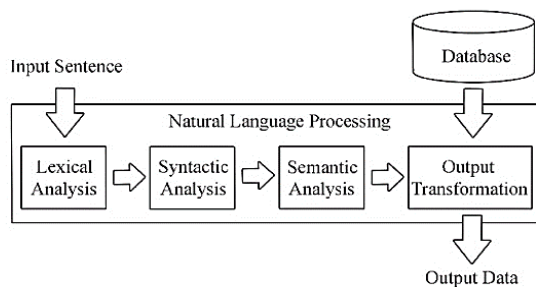

Fig. 4. NLP working steps

Researcher have identified top level benefits of NLP which would help for many sectors' performance. Automate processes in real-time, perform large-scale analysis, and Tailor NLP tools for industry are revealed benefits in this study.

Building a Machine Learning model depends on how effectively the data has been preprocessed. This is a method of convert data into machine understandable way. Natural Language Tool Kit (NLTK) is a common framework for text preprocessing in NLP. NLTK has more sub libraries, which are tokenization, classification, Stop-words, grouping, lemmatizing, parsing, tagging and semantic. Tokenize will split sentence into words, Stop-word will delete Stop-words (an, the, is, a, are, in, . . . etc.) and reduces the words to a word existing in the language by lemmatization and remove redundant word and so on are occur in text preprocessing.

3.3 Experiment - First Evaluation
3.3.1 Input Questions:
User inputs questions in the HTML page and start first Evaluation Process. Python variables collect these questions one by one and send them to Pre-process task.

3.3.2 Pre-Processing of Text:
Split the questions in to sentences and further split them into words using NLTK word tokenize method. Example python code;

```
# word tokenization
from nltk.tokenize import word_tokenize
sentence = "what methods could be used to
shorten the project duration"
words = word_tokenize(sentence)
print(words)
```

Output: ['What', 'methods', 'could', 'be', 'used', 'to', 'shorten', 'the', 'project', 'duration?'] Then remove stop words by corpus and Stop-word. Those are very commonly used words, which do not signify any importance for the task.

```
#Stop Word Remove
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
sentence = "what methods could be used to
shorten the project duration"
stop_words = set(stopwords.words('english'))
word_tokens = word_tokenize(sentence)
filtered_sentence = [w for w in word_tokens if not
w instop_words]
print(filtered_sentence)
```

Output: ['What', 'methods', 'could', 'used', 'shorten', 'project', 'duration?'] In the same sentence the words 'be','to' and 'the' have deducted by above algorithm. Finally transform the words to its root format by Stemming Process.

```
# Stemming Process
import nltk
from nltk.stem import PorterStemmer
ps = PorterStemmer()
sentence = "what methods could used shorten
project duration?"
for word in sentence.split():
print(ps.stem(word))
```

Output (Fig.7): ['What', 'method', 'can', 'use', 'short', 'project', 'duration']

3.3.3 Jaccard Similarity:
Jaccard coefficient has been introduced in and has been extensively used to measurethe similarity of sets [9]. The Jaccard coefficient is frequently used in many different applications. The similarity of text, which when treated as a bag of words, equates to quantitative, multidimensional data, is sometimes measured using the coefficient.

$$J(A, B) = \frac{|S_A \cap S_B|}{|S_A \cup S_B|}$$

SA and SB are sets with binary representations A and B, Jaccard's formula is equals: Since there are no common items, the Jaccard index for disjoint sets is equal to 0. The result of this measure's calculation for two sets with the identical members, however, would be 1.

Any two sets have a Jaccard coefficient that falls between [0, 1]. A higher score suggests that there are more common items than there are total elements in sets (TABLE 1). High Jaccard coefficient scores thus indicate a high level of similarity. However, depending on the problem domain, thresholds may need to be specified.Researcher reveals the most suitable Bloom's Revised Taxonomy Level for each question using above NLP techniques. Here use Jaccard similarity function for reveal similarity score, then present applicable Bloom's Revised Taxonomy Level to report. This report should be shown as below. Here have taken four questions related to paper.

Table 1. Example of Report 1

Question Number	Bloom's Revised Taxonomy Level with Marks					
	Remember	Understand	Apply	Analyze	Evaluate	Create
01	0.5	0.3	0.0	0.0	0.0	0.0
02	0.0	0.2	0.1	0.2	0.0	0.0
03	0.1	0.1	0.1	0.0	0.0	0.0
04	0.0	0.1	0.1	0.0	0.0	0.0
Average Marks for each Level	**0.15**	**0.17**	**0.07**	**0.05**	**0.0**	**0.0**
Predicted Marks	0.17					
Predicted Level	Remember					

3.4 Experiment - Second Evaluation

3.4.1 Gensim

This evaluation process uses Gensim package relatively evaluation one. It can create dictionary objects. Dictionary object can map each word with unique id. Sentences in the documents can convert to the list of word and pass them to the corpora.Dictionary() object. The advantage of the Gensim is, it's lets user to read the text and update dictionary line by line without loading entire document. After this the document's word will contained with a bag of word. This is a object known as Corpus and contains the word id and its frequency in each document. Sentences or paragraphs can be considered "documents," whereas a "corpus" is a collection of documents that take the form of a bag of words. It will now create a corpus of words send them to Dictionary.doc2bow().

3.4.2 Similarity Object:

The important one is creating the similarity object. First, similarity class create a index of set of documents and splits the index into several smaller sub-indexes (Fig. 8). After the index is created, we will determine how similar each index document is to the exam paper questions on by one. Similarities are shown as float number between -1 and +1.Then calculate average similarity of the whole questions that user Provide. The library Numpy is help for calculate average.
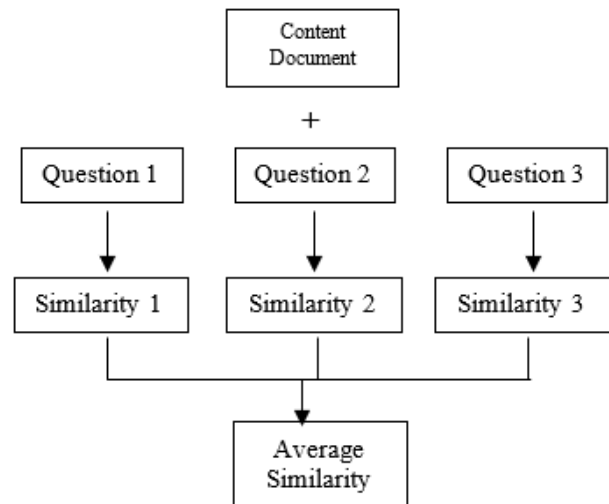

Fig. 9. Similarity finding Architecture

Finally, researcher reveal the covered percentage of the course content by the online exam paper. Then the report shows in the page as below table (TABLE 2). Here have taken four questions related to paper.

Table 2. Table Example Report 2

Question Number	Similarity	Average Similarity	Covered Percentage (%)
01	0.34		
02	0.15	0.235	23
03	0.19		
04	0.26		

There are few challenges acquired by this research study. Those are mentioned below,

- Selecting Suitable Techniques.
- Implementing of working machine learning Algorithms.
- Selecting right features of the proposed system.
- Accuracy more than manual process.

4. CONCLUSIONS

In this research, we propose a system for the automatic evaluation of exam papers that uses Natural Language Processing (NLP) algorithms to calculate the results. Also,this system use Bloom's Revised Taxonomy Model as a standard. We developed algorithms to measure the similarity of exam questions with course content and find the level of difficulty of each question and the same of overall exam paper using the Bloom's Taxonomy. This application helps to reduce time spent in manual exam paper moderation process and provide a faster evaluation of exam papers. It allows the educator to create an online exam paper in proposed system and start evaluationprocess. The system checks the exam paper against the topics in the module outline and the action verbs associated with the Bloom's Taxonomy Level using NLP techniques and represent the reports to the users. Researcher have used the NLTK library of NLP for text preprocessing task and then tokenization, gensim package, tf-idf and numpy as the resources for evaluating Processes. In future work, the system will be improved to check against Learning Outcomes of the module using deep learning approaches for increased accuracy. Finally propose Bloom's Revised Taxonomy Key

word modifications and pre-trained question types to add each level.

5. REFERENCES

[1] D. Dayananda, K. Chathumini and S. Vasanthapriyan, "A Novel Framework for Online Exams during the Pandemic of COVID-19: Evaluation Methods, Students' Priorities and Academic Dishonesty in Online Exams," 2021 IEEE 1st International Conference on Advanced Learning Technologies on Education Research (ICALTER), 2021, pp. 1-4, doi: 10.1109/ICALTER54105.2021.9675092.

[2] R. Rajesh. and R. Kanimozhi., "Digitized Exam Paper Evaluation,"2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), 2019, pp. 1-5, doi: 10.1109/ICSCAN. 2019.8878791.

[3] G. Sanuvala and S. S. Fatima, "A Study of Automated Evaluation of Student's Examination Paper using Machine Learning Techniques," 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 2021, pp. 1049-1054, doi: 10.1109/ICCCIS51004.2021.9397227.

[4] Rish, I., 2001, August. An empirical study of the naive Bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence (Vol. 3, No. 22, pp. 41-46).

[5] Couronne, R., Probst, P. and Boulesteix, A.L., 2018. Random forest versus logistic regression: a large-scale benchmark experiment. BMC bioinformatics, 19(1), p.270.

[6] Huang, S., Cai, N., Pacheco, P.P., Narrandes, S., Wang, Y. and Xu, W., 2018. Applications of support vector machine (SVM) learning in cancer genomics. Cancer Genomics-Proteomics, 15(1), pp.41-51.

[7] Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Annals of statistics, pp.1189-1232.

[8] Verspoor, Karin Cohen, Kevin. (2013). Natural Language Processing. 10.1007/978-1-4419-9863-7-158.

[9] Kabasakal, InancSoyuer, Haluk. (2021). A Jaccard Similarity-Based Model to Match Stakeholders for Collaboration in an Industry - Driven Portal. Proceedings. 74. 15. 10.3390/proceedings2021074015.

[10] Krathwohl, David. "https://www.tandfonline.com/doi/abs/10.1207/s15430421tip4104_2?journalCode=htip20." *A Revision of Bloom's Taxonomy: An Overview*, 24 June 2010, www.tandfonline.com/doi/abs/10.1207/s15430421tip4104_2?journalCode=htip20.

[11] M. Agarwal, R. Kalia, V. Bahel and A. Thomas, "AutoEval: A NLP Approach for Automatic Test Evaluation System," 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), 2021, pp. 1-6, doi: 10.1109/GUCON50781.2021.9573769.

[12] Vij, S., Tayal, D.K., & Jain, A. (2020). A Machine Learning Approach for Automated Evaluation of Short Answers Using Text Similarity Based on WordNet Graphs. Wireless Personal Communications, 111, 1271-1282.

[13] M. Agarwal, R. Kalia, V. Bahel and A. Thomas, "AutoEval: A NLP Approach for Automatic Test Evaluation System," 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), 2021, pp. 1-6, doi: 10.1109/GUCON50781.2021.9573769.

[14] B. Galhotra and D. Lowe, "AI Based Examination System: A Paradigm Shift in Education Sector," 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), 2022, pp. 386-392, doi: 10.1109/COM-IT-CON54601.2022.9850452.

[15] Family roomcafe, "Latest Android Projects - Digitized Exam Paper Evaluation - Family Room Cafe," *Family Room Cafe*, Jan. 06, 2020. http://familyroomcafe.com/digitized-exam-paper-evaluation/ (accessed Nov. 20, 2022).