International Journal of Technology Engineering Arts Mathematics Science
Vol. 1, Iss. 2, January 2022, pp. 1~7
ISSN: 2583-1224

Voice based Answer Detection and Evaluation System using NLP and ML

Mansi Kuwadia[1], Sarang Chambhare[2], Abhishek Zade[3], Sourav Shinde[4], Dr. Meenakshi Thalor[5]

[1,2,3,4,5]Department of Information Technology, AISSMS IOIT, Pune, Maharashtra, India
Corresponding Author: mansikuwadia12@gmail.com

Article Information

Article history:

Received December 2, 2021
Revised January 15, 2022
Accepted January 15, 2022

ABSTRACT

As the globe evolves toward automation, there is a need for automation in answer evaluation systems in our modern age. Because online answer evaluation is now only available for mcq-based questions, the checker's job is made more difficult when evaluating theory answers. The teacher carefully checks the answer and assigns the appropriate grade. The existing system necessitates additional staff and time in order to assess the response. An application based on the evaluation of answers using machine learning is presented in this publication. The goal of this paper is to reduce labor and time usage in particular because manual answer evaluation requires significantly more people and time. Also, with the manual approach, it's possible that two identical responses will receive different marks. This application system enables an automatic evaluation of answers based on the keyword provided to the application in the form of an input by the moderator, ensuring that marks are distributed evenly and saving time and personnel.

KEYWORDS: Natural Language Processing, Machine Learning, Speech to Text, Text to Speech, Naive Bayes, Word2vec, Cosine Similarity.

1. INTRODUCTION

In general, students' academic success is evaluated based on their examination results, which might be subjective or objective. There are a number of systems that can swiftly evaluate objective or multiple-choice questions. After providing pre-defined accurate responses, these strategies are evaluated in machines. However, it is only useful for evaluating competitive or objective exams. Subjective examinations provide the foundation of all university and board-level exams. The moderator will know how much knowledge the student has obtained during his academic career based on the descriptive answer, on which the moderator will allocate marks. Manually evaluating subjective responses is a time-consuming and labor-intensive operation that requires a large number of people. According to their method of evaluation, the mood at the moment of evaluation, and interrelationship between student and moderator, answer evaluation varies from moderator to moderator. This has an impact on the student's grade. The goal of the research is to use machine learning and natural language processing to automate the evaluation process for subjective answers.

The mapping, succession, linear sequence matching, quantitative identification, and semantic research methods are utilized to evaluate student answers utilizing natural language processing and machine learning. The major goals of this research are to review student descriptive type responses using NLP and ANN algorithms, as well as to build a tool for evaluating student descriptive type replies using NLP for Grammatical checking, keywords and evaluation of marks, and ANN for normal answer comparison and producing marks. Also, the proposed system aims to provide voice assistant for visually impaired students. This project aims for the exam evaluation for students who can't write but are able to speak by providing speech-to-text and text-to-speech technology.

2. LITERATURE SURVEY

The existing method of evaluating subjective papers is ineffective. The challenge of evaluating subjective answers is crucial. When a human being evaluates anything, the quality of the evaluation can vary depending on the person's emotions. As a result, a plethora of automated subjective answer evaluation systems has emerged. We have briefly discussed some of the existing research articles that are connected to our work in this part. For details on related research publications refer Table I.

Table I. Literature Review

Sr. No.	Title	Conclusion	Limitations
1.	Subjective Answer Evaluation using Natural Language Processing and Machine Learning [1]	Subjective Answer Evaluation software assigns a grade to a subjective question based on the length of the answer, keyword matching, grammar check, cosine similarity, and contextual resemblance to the faculty's model answer and the student's answer.	The length of the answer determines the grade, although the length varies from person to person.
2.	Answer Evaluation Using Machine Learning [2]	When compared to a manual system, the proposed technique is around 75-87.5 percent accurate. The proposed approach eliminates all human effort and time required to analyze a response.	Only printed text can be recognized by the proposed technology, not handwritten text.
3.	Evaluating Student Descriptive Answers Using Natural Language Processing [3]	The main goal of this newly proposed method is to determine the semantic meaning of student responses, taking into account that students can respond to questions in a variety of ways.	The proposed system takes into account the meaning of collective utterances, which may conflict with the meaning of students' responses.
4.	Automatic Answer Evaluation Using Machine Learning [4]	The suggested system will use OCR with a backpropagation method and an artificial neural network. The algorithm will assess the response based on the scanned answer sheet, the moderator's keywords, and the length provided. The marks are given out based on the following factors: a. the number of keywords that were matched; and b. the length of the answer.	Exact keyword is required; synonyms or similar words are not acceptable.
5.	Design Engineering Automated Explanatory Answer Evaluation Using Machine Learning Approach [5]	Using Natural Language Processing, the system assesses the responses by extracting keywords from the students' replies and the tutor answer key. The cosine similarity metrics are used to verify for similarity between the student's answer and the answer key.	The technique primarily relies on cosine similarity for grading purposes.
6.	Automatic Online Subjective Text Evaluation using Text Mining [6]	The project is driven by similar data that a human would consider when analyzing, such as answer length, keyword presence, and keyword context. Natural Language Processing, in combination with categorization algorithms, is used to look for keywords and answer specific questions.	For model training, greater computing power is required.
7.	Evaluating Students Descriptive Answers Using Natural Language Processing and Artificial Neural Networks [7]	After the text mining process is completed, the student answer is compared to the correct answer using the ANN algorithm, and the student answer is checked for spelling and grammatical errors using the NLP algorithm.	No feedback option for students to suggest any improvements.
8.	Computerized Evaluation of Subjective Answers Using Hybrid Technique [8]	The methods of evaluation utilized the combination of LSA and BLEU is complimentary. The usage of WordNet reduces the number of keywords required because it finds synonyms for the specified keywords. This assures that the student can use any language he wants.	Models (LSA & BLEU) are required for evaluation. It is impossible to work with a single model.

Sr. No.	Title	Conclusion	Limitations
9.	Subjective Answer Grader System Based on Machine Learning.[9]	By comparing the computed scores of the replies with the human evaluation scores, the accuracy of LSA and IG algorithms may be determined. The accuracy of LSA and IG has grown from less than 40% to more than 75% once they were enhanced using WordNet.	Formulas and diagrams are not supported by the system.
10.	A Keyword Based Technique to Evaluate Broad Question Answer Script [10]	The system assesses the student's response using the keywords. The student will be assigned marks based on the sample answer and the student's answer.	Teachers' errors in submitting answer keys will result in erroneous evaluations.
11.	Text Similarity Analysis for Evaluation of Descriptive Answers [11]	The system employs natural language processing (NLP) and data mining, as well as an LSTM (recurrent neural network). Marks are distributed in a highly regular manner.	Answers must be typed into the system by the students. After the evaluation, no feedback is given.
12.	ASSESS-Automated Subjective Answer Evaluation Using Semantic Learning [12]	To generate system embedding, the proposed system employs NLP, semantic learning, and Google's USE algorithm. It eliminates the need for the user to write further answers and gives appropriate feedback.	Marks are assigned based on two factors: similarity and keywords. There is no synonym module included.
13.	Computer Application for Assessing Subjective Answers Using Artificial Intelligence [13]	The system is based on machine learning, natural language processing, and artificial intelligence, as well as methods such as HMM, RNN, STS, and CFG. Sequencing, watchword planning, and summarization are the basic concepts of evaluation.	After the evaluation, no feedback is given. It is impossible to evaluate answers that include equations, graphs, or formulas.
14.	Subjective Answer Grader System Based on machine learning [14]	In the system, algorithms such as LSA and IG are used, and WordNet is used to improve their evaluation. It has an accuracy of up to 83 percent.	For equations and graphs, there is no evaluation. For improved results, algorithms require WordNet.
15.	Speech To a Text Translation Enabling Multilingualism [15]	NLP with pre-existing data This method makes use of Google's Speech Recognition software. It can be used by people who have no prior computer experience. It is not necessary to have access to the internet. Speech can be translated into a variety of languages.	This model requires a pre-installed Speech Recognition program. Converts a small number of words at a time.
16.	Subjective Answer Evaluation Using Machine Learning [16]	The proposed approach employs machine learning and natural language processing, utilizing Naive Bayes as the underlying classifier. It evaluates quickly and efficiently. The system has a 90 percent accuracy rate.	The system does not provide feedback. Teachers must provide required keywords manually.
17.	Speech to text conversion using GRU And one hot vector encoding [17]	This system employs an RNN-based model with a gated recurrent unit. It is capable of achieving an accuracy of 87 percent.	It is only available in a single language. Converts a small number of words at a time.
18.	A focus on code mixing and code switching in Tamil speech to text [18]	Google's cloud conversion API and Google's Speech to Text model are used by the system. The technology allows for two or more languages to be spoken simultaneously, as well as the usage of two or more languages to compose a word.	Requires a quiet setting. Few accents are supported. For improper pronunciation, no output is produced.

3. METHODOLGY

The proposed model uses natural language processing and machine learning to evaluate voice-based answers. All of the student replies, as well as one standard answer each question, are used as inputs. The output is the pupils' final grades. To begin, input is pre-processed to prepare it for usage in the review process. Tokenization, synonym search, stop words, and stemming of student and standard replies have been completed. The proposed technique is designed to evaluate exam results for students who are unable to write but can talk. The proposed technique is designed to evaluate exam results for students who are unable to write but can talk. **Fig. 1** shows how the assessment technique is executed in a sequence of phases.

There are two aspects to the proposed system:

- Conversion of speech to text and text to speech
- Evaluation of the responses.

3.1 Text Extraction

Text extraction, often known as audio extraction, is a method of automatically scanning text and extracting

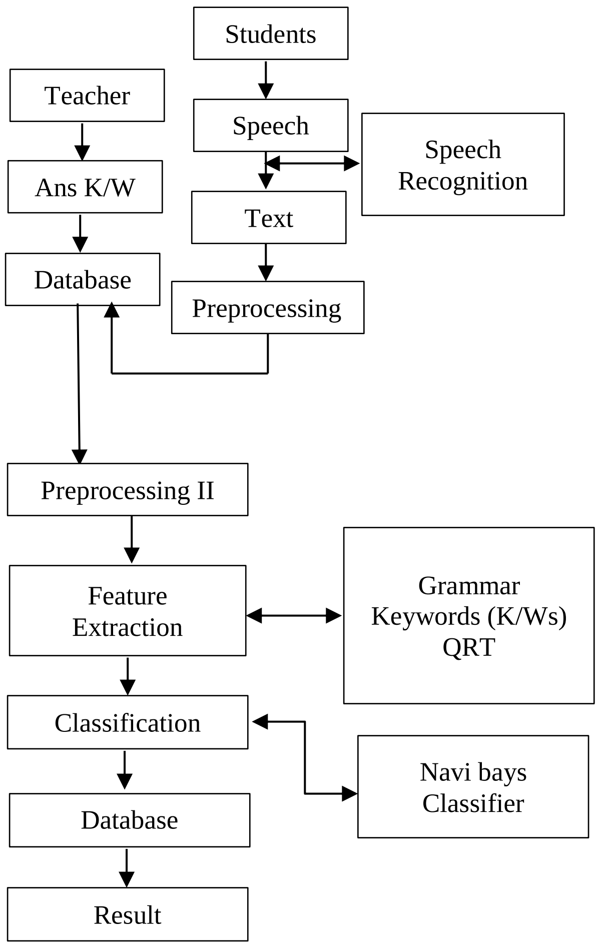

Fig. 1. System Architecture

relevant or core words and phrases from unstructured data using machine learning. This section focuses primarily on speech-to-text conversion. As we all know, speech-to-text conversion is part of natural language processing. We record the sound, process it, and then output it as text. We can utilize two alternative natural language processing modules to convert speech to text:
1. Deep Speech: Deep Speech is an open-source speech-to-text library that converts text to speech in real time. However, based on the examined literature, we may conclude that Deep Speech outcomes are good, albeit at the cost of the module's enormous file size.
2. Python Speech recognizer: The speech recognizer package in Python is used to do voice recognition with a variety of engines and APIs. The Recognizer class in Python speech recognizer utilizes seven different methods for recognizing speech for audio sources.

3.2 Answer Evaluation

The following step is the answer evaluation, which comes after the text extraction procedure. There are three elements to the answer evaluation:

- The pre-processing stage
- Extraction of features
- Classification of Scores.

3.2.1The Pre-Processing Stage

Text pre-processing is a common stage in Natural Language Processing (NLP). It converts material into a more digestible format, allowing machine learning algorithms to perform the rest of the work more efficiently. As a result, we must pre-process the instructor and student data before evaluating them. We don't need some features in data because we're going to compare them, for example, we don't require punctuation or stop words. Pre-processing begins with the removal of punctuation, followed by word tokenization to convert the entire work into tokens, and finally the removal of stop words. The stop words are essentially "the," "and," and so on. Following the removal of stop words, stemming is used to eliminate all distinct versions of the same word, such as move and moving. All other phases in feature extraction, with the exception of grammar assessment, are pre-processed.

3.2.2. Feature Extraction

Because document data cannot be computed, it must be converted to numerical data, such as a vector space model. Feature extraction of document data is the common term for this transformation activity. Feature extraction is divided into three stages. These three components are an important feature of the suggested assessment method since they serve as the foundation for our evaluation.

a) Checking for similarity

In this exercise, we will compare the similarity of teacher and student responses and calculate the percentage of similarity. We are going to calculate Tf-IDF (Term frequency and inverse document frequency) of terms that are included in teacher and student answers, and then we will calculate percentage similarity using cosine similarity.

b) Extraction of Keywords

As we receive essential keywords connected to teacher responses. To extract relevant terms from the answer, we use the Rake-NLTK module. This is done to cut down on the length of the answer and to allow for quick and efficient comparison. After extracting essential terms from student responses, we'll compare them to keywords used by professors. To compare keywords, we'll use a sequence matcher from difflib, which compares two strings and returns a ratio, such as the ratio of similarity between the words "abc" and "abc123." We've detected the keywords if the similarity ratio is 0.8 or higher. Following the comparison, the proportion of keywords for each category is found.

c) Grammar Evaluation

We need to evaluate the grammar because we want the response to have meaning. If we don't assess grammar, some pupils may rely solely on keywords and pass the exam. This is done to ensure a fair assessment. We don't need to perform pre-processing for grammar evaluation because punctuation is required to complete the sentence's meaning. We're sentencing, which involves tokenizing paragraphs, then words tokenizing sentences, chunking and chinking, and lastly calculating sentence mistakes. If there is no problem in the sentence, it is grammatically correct; if there is, it is grammatically incorrect. Finally, we will calculate the average accurate percentage in this manner.

3.2.3. Classification of Scores

Now that we have all of the percentages, we must assess them. We'll categories using a Naïve Bayes classification machine learning technique to figure out which group our answer belongs to.

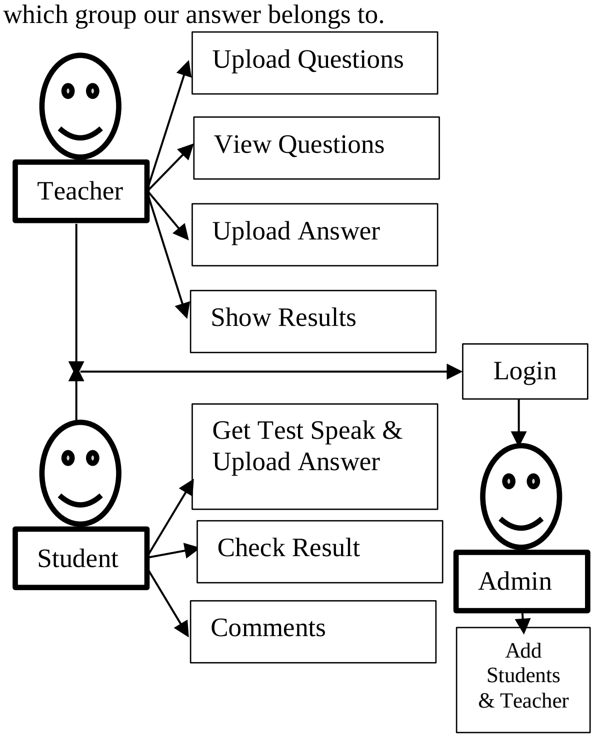

Fig. 2. Use Case Diagram

We'll train some example datasets and put our findings to the test. As a result, we'll come to a conclusion about the end result. Refer **Fig. 3** for use case diagram.

4 ALGORITHMS

4.1 Naïve Bayes

The Naive Bayes method is a supervised learning algorithm for addressing classification issues that is based on the Bayes theorem. It is mostly utilized in text classification tasks that require a large training dataset. The Naive Bayes Classifier is a simple and effective classification method that aids in the development of fast machine learning models capable of making quick predictions. It's a probabilistic classifier, which means it makes predictions based on an object's probability. Bayes Theorem is given in Equation (1).

$$P(A/B) = \frac{P(A/B)\, P(A)}{P(B)} \tag{1}$$

Where,

- Posterior probability (P(A|B)) is the probability of a hypothesis on the observed event B.
- P(B|A) stands for Likelihood, which is the probability of the evidence provided that a hypothesis' probability is true.
- Prior Probability (P(A)) is the probability of a hypothesis before looking at the evidence.
- P(B) stands for Probability of Evidence Marginal Probability.

It's a popular and effective classification algorithm.

4.2 Natural Language Processing

Natural language processing (NLP) is a branch of linguistics, computer science, and artificial intelligence that studies how computers interact with human language, particularly how to design computers to process and analyze massive amounts of natural language data. The goal is to create a computer that can "understand" the contents of papers, including the nuances of language in context. The system can then extract accurate information and insights from the papers, as well as categories and organize them.

Natural language processing techniques and approaches:

Natural language processing employs two basic techniques: syntax and semantic analysis. The arranging of words in a phrase to make grammatical sense is known as syntax. NLP analyses a language's meaning using syntax and grammatical rules. Syntax techniques include the following:

Parsing: This is a sentence's grammatical analysis. The text "The dog barked" is supplied to a natural language processing system as an example. Parsing is the process of breaking down a sentence into its constituent components of speech, such as dog = noun and barked

= verb. This is beneficial for activities that require more complex downstream processing.

Segmentation of words: This is the process of deriving word formations from a string of text. A person scans a handwritten paper into a computer, for example. The program would be able to examine the page and detect the presence of white spaces between the text.

Sentence Breaking: In lengthy texts, this creates sentence borders. The text is fed through a natural language processing system, for example, "The dog let out a bark. I became aware of my surroundings." The algorithm is able to distinguish the period that is used to break up the phrases.

Segmentation morphologically: This breaks down words into smaller units known as morphemes. The word untestable, for example, would be broken down into [[un[[test]able]]ly, where the algorithm recognizes the morphemes "un," "test," "able," and "ly." This is very important in speech recognition and machine translation.

Stemming: This separates words that have inflections into root forms. For example, the algorithm would be able to determine that the root of the word "barked" is "bark" in the sentence "The dog barked." If a user was looking for all instances of the word bark, as well as all of its conjugations, this would be handy. Even if the letters are different, the algorithm recognizes that they are basically the same term.

Semantics is the study of how words are used and what they mean. Algorithms are used in natural language processing to understand the meaning and structure of sentences. Figure 4 shows semantic similarity-based technique. Techniques used in semantics include:

Disambiguation of word meanings: This method uses context to determine the meaning of a word. Consider the following sentence: "The pig is in the pen." The term "pen" has a variety of meanings. This approach allows an algorithm to grasp that the term "pen" refers to a fenced-in region rather than a writing implement.

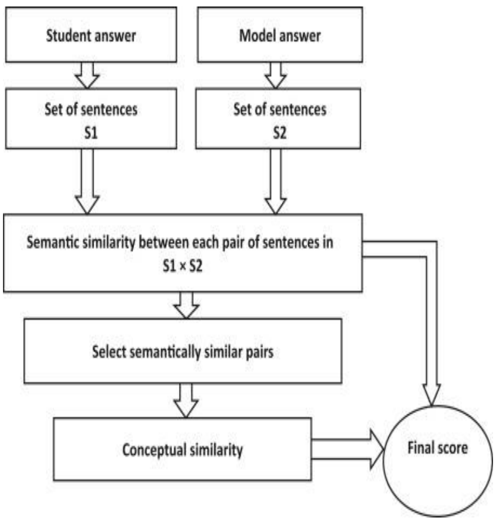

Fig. 4. Semantic Similarity based Subjective Answer Evaluation Model

Recognition of named entities: This defines which words can be grouped together. For example, using this strategy, an algorithm could evaluate a news story and find all mentions of a specific firm or product. It would be able to distinguish between visually similar entities using the semantics of the text. The system might detect the two instances of "McDonald's" in the line "Daniel McDonald's son went to McDonald's and ordered a Happy Meal," for example, as two unique entities one a restaurant and the other a person.

Generating natural language: The semantics behind words is determined using a database, and fresh text is generated. For instance, an algorithm may generate a summary of findings from a business intelligence platform automatically, connecting particular terms and phrases to elements of the BI platform's data. Another example would be creating news stories or tweets automatically based on a body of content used for training.

5 CONCLUSION

The project is driven by similar data that a human would consider when analyzing, such as answer length, keyword presence, and keyword context. Also, the proposed system uses Python speech recognizer for speech to text and text to speech conversion. Natural Language Processing, in combination with categorization algorithms, is used to look for keywords and answer specific questions. Because the system analyses for the occurrence of keywords, synonyms, correct word context, and coverage of all subjects, students will have a lot of leeway while crafting the answer. As a result of the robust evaluation system, it can be determined that applying ML approaches yields satisfactory results. The accuracy of the evaluation can be improved by providing it a big and accurate training dataset. We aim to create a Voice-based answer evaluation system using natural language processing and machine learning for physically disabled students. From the literature survey, for our proposed system speech recognizer is the best option available. And finally, we can classify results using the naive Bayes classifiers algorithm.

REFERENCES

[1] Abhishek Girkar, Mohit khambayat, Ajay Waghmare, Supriya Chaudhary, “Subjective Answer Evaluation using Natural Language Processing and Machine Learning.” *International Research Journal of Engineering and Technology (IRJET)* e-ISSN: 2395-0056 Vol. 08 Issue: 04, Apr 2021.

[2] Prince Sinha, Sharad Bharadia, Ayush Kaul, Dr. Sheetal Rathi, “Answer Evaluation Using Machine Learning.” *Conference Paper · March 2018, researchgate.net*.

[3] Shweta M. Patil, Sonal Patil. Evaluating Student Descriptive Answers Using Natural Language Processing. *International Journal of Engineering Research & Technology (IJERT)*, Vol. 3 Issue 3, March - 2014.

[4] Nandita Bharambe, Pooja Barhate, Prachi Dhannawat., “Automatic Answer Evaluation Using Machine Learning”, *International Journal of Information Technology (IJIT)* – Vol. 7 Issue 2, Mar - Apr 2021.

[5]A.Mercy Rani., "Automated Explanatory Answer Evaluation Using Machine Learning Approach", *Article in Design Engineering (Toronto)* · July 2021.

[6] Shweta Deotare, Rubina A. Khan, "Automatic Online Subjective Text Evaluation using Text Mining", *International Journal of Recent Technology and Engineering (IJRTE),* ISSN: 2277-3878, Vol.-8 Issue-2, July 2019.

[7] V. Lakshmi, Dr. V. Ramesh, "Evaluating Students' Descriptive Answers Using Natural Language Processing and Artificial Neural Networks*", International* Journal of Creative Research Thoughts, Vol. 5, Issue 4 December 2017.

[8] Himani Mittal, M. Syamala Devi, "Computerized Evaluation of Subjective Answers Using Hybrid Technique", *Springer Science Business Media Singapore 2016.*

[9] Avani Sakhapara, Dipti Pawade, Bhakti Chaudhari, Rishabh Gada, Aakash Mishra and Shweta Bhanushali, "Subjective Answer Grader System Based on Machine Learning", *Springer Nature Singapore Pte Ltd. 2019 J. Wang et al. (eds.), Soft Computing and Signal Processing, Advances in Intelligent Systems and Computing 898*.

[10] Tamim Al Mahmud, Md Gulzar Hussain, Sumaiya Kabir, Hasnain Ahmad, Mahmudus Sobhan, "A Keyword Based Technique to Evaluate Broad Question Answer Script", *ICSCA '20,* February 18–21, 2020, Langkawi, Malaysia.

[11] Bahel, Vedant, and Achamma Thomas. "Text similarity analysis for evaluation of descriptive answers." *arXiv preprint arXiv:*2105.02935 (2021).

[12] Johri, Era and Dedhia, Nidhi and Bohra, Kunal and Chandak, Prem and Adhikari, Hunain, "ASSESS - Automated Subjective Answer Evaluation Using Semantic Learning" (May 7, 2021).

[13] Singh, Shreya and Manchekar, Omkar and Patwardhan, Ambar and Rote, Uday and Jagtap, Sheetal and Chavan, Dr. Hariram, "Computer Application for Assessing Subjective Answers using AI" (May 7, 2021).

[14] Piyush Patil, Sachin Patil, Vaibhav Miniyar, Amol Bandal Subjective, "Answer Evaluation Using Machine Learning", *International Journal of Pure and Applied Mathematics*, Vol. 118, Iss. . 24 (2018).

[15] PavuluriJithendra, ShahanaBano, Gorsa Lakshmi Niharika, YalavarthiSikhi, "Speech to Text Translation enabling Multilingualism", *IEEE International Conference for Innovation in Technology (INOCON)* Bengaluru, India. Nov 6-8, 2020.

[16] AvaniSakhpara, Dipti Pawade, Bhakti Chaudhari, Rishabh Gada,Aakash Mishra and Shweta Bhanushali., "Subjective Answer Evaluation Using Machine Learning. Springer Nature Singapore Pte Ltd. 2019 J. Wang et al. (eds.)", *Soft Computing and Signal Processing, Advances in Intelligent Systems and Computing* 898.

[17] Manpreet Singh Bhatia, Alok Aggarwal, Narendra Kumar, "Speech-to-text conversion using GRU and one hot vector encodings", *Palarch's Journal of Archaeology of Egypt/Egyptology*, Vol. 17, Iss. 9. 2020.