

# Enhancement of the Automatic Evaluation Systems for Descriptive Type Answers Using Neural Networks

Sushila Choudhari <sup>#1</sup>, Bindu Garg <sup>\*2</sup>

1 M. Tech Student, Department of Computer Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, sschoudharipg20-comp@bvucoep.edu.in

2 Professor, Department of Computer Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, brgarg@bvucoep.edu.in

## Article Info

**Page Number:** 2203-2219

**Publication Issue:**

**Vol. 71 No. 4 (2022)**

## Article History

**Article Received:** 25 March 2022

**Revised:** 30 April 2022

**Accepted:** 15 June 2022

**Publication:** 19 August 2022

## Abstract

Constantly informative foundations lead various appraisals, which join institutional and non-institutional driving forward tests. Right now online tests and appraisals are turning out to be noticeable to lessen the generosity of the assessment cycle. The electronic tests set either helpful or choice based demands. In this affiliation, unique based demands and answers are not locked in with light of the perplexing nature and limit of the assessment cycle. In like manner, the programming applications used to check dynamic reactions may be more useful for designating etchings to the client soon after really looking at the strategies with an online assessment. This kind of instrument/application/system should have a lot of resource instructive record, including questions, relating answers. Also, it ought to comparably have the etchings given out to the separating answers. Simultaneously structure correspondingly needs to check the reactions given by the clients by truly looking at the game plan of replies. As a conceded consequence of this man-made speculation based response verifier, the evaluator's critical venture can be created, with additional work capacity.

**Keywords:** Answer verifier; Artificial Intelligence; Similarity check; Subjective answer verifier.

## Introduction

In the current reality, Online Examinations have arisen as a huge contraption to survey the reactions of clients. For web set together tests depend with concession on genuine demands and tests are getting digitized beginning with one side of the planet then onto the following. In the ongoing circumstance, test questions could be set up on energized reactions. The standard tests ordinarily contained speculative reactions, which were not the best technique for investigating the understudy's impression of the subject. Since every once in a while, overseers get depleted by truly taking a gander at various response sheets, and there may be expansion in the phony assessment.

Thusly, the Artificial Intelligence-put together Answer Verifier is depended with respect to grade the understudy after he/she has attempted the request paper. Notwithstanding, the development decreases the responsibility of the inspector through robotizing the manual really looking at process. A changed energized solicitation's wandering is a principal progress in the connection test system. To deal with this issue, Artificial Intelligence (AI) based Answer

Verifier (AV) is depended upon to look at like an assessing instructor and think while checking questions whose answers can be laid out on vibes of the understudy.

In various foundations, the results are accounted for later time since educators consume a gigantic piece of the day to survey the energized papers. As a store of answers, booklets ought to be evaluated and each booklet could contain the reaction with a particular goal in mind, which requires a more extended length. Hence, by and by, an AI-based Answer Verifier can come into work [4], taking into account the closeness speculation of cushioned ascertaining [2]. It does the adjusted scoring of extraordinary sales through unequivocal reference regards. In the assessment affiliation, a decision based scoring estimation is seen as a productive one [6]. The development subject to the AI will save time and effort of journalist.

The electronic assessment makes a charming work space by giving speedier access. The ongoing plan concludes the score subject to a couple of limits like enunciations, or reciprocals. In the "Fake Insight based Answer Verifier", it enrolls the score of the understudy by combining different limits, like enunciations, question unequivocal things close by the real language, which in communicating gives a more precise score. The man-made thinking set up technique will continue concerning making.

Manual reaction assessment is a staggeringly somber endeavor. The manual checking is unquestionably dull correspondence, and it requires bundle of manual work. Furthermore, the paper checker can't give stamps.

Our development will correspondingly survey answer subject to some enunciation and hence work will be saved. One essential to evaluate the paper then the structure will part the reaction using OCR [1]. Right when watchword is found in the reaction then the development will give the etchings to the sales as demonstrated by the dataset present [2]. There is a requirement for such application which will offer key evaluation of reaction and can give qualified engravings. In like manner, this application will help various schools, school, and arranging relationship to concentrate on the reaction amazingly speedy with less work.

Checking answers requires high obsession for the titanic degree of time which reliably prompts messes up. The computerization of this task will foster the adequacy of answer evaluation for a colossal extension. It was seen that answer sheet is inspected recalling express watchwords that middle individuals look at the response while assessing a response. Our proposed calculation will require watchwords as data sources. Our proposed calculation will facilitate these watchwords with perceived words that are killed from the reaction sheet utilizing oversight learning assessment. Learning season of the model will require truly made dataset [3] for English language letters generally together. These datasets are accessible web-based in different plans to be utilized to set up the model.

Online Assessment is phenomenally important to clients. The characteristic of this attempt is to give fast, brief and essential method for seeming, by all accounts, to be the test. It can give extraordinary benefits to the students/contenders that can't be found elsewhere through relationship with working environments/sheets that are planning the unmistakable decision sort appraisal. Students' licenses selecting for the test and instructor awards enlisting for

organizing the test. This will keep on making at last giving a total breadth of associations for supportive to the students. Tests can be made on an irregular justification for each student. The online evaluation construction can hence add the etchings appointed in each inquiry to pick the out and out etching for the test. The electronic examination framework limits the events a student can shape a test. Students can be obliged to go through each solicitation once, prior to leaving the test. Students can be permitted to leave the test right after finishing all of the requests. The objective is to check, and stamp shaped responses like an individual. This programming application is used to research energetic responses on the web assessment and allot inscriptions to the client coming about to certifying the response. The design guesses that you ought to store the key response for the system. Whenever the client enters his/her responses the design then, looks at this solution for exceptional response written in information base and regulates checks besides. Both the responses should not be same word to word. Framework diminishes the commitment of Inspectors through robotizing the manual cycle.

The need for online assessment blended predominantly to persevere through the downsides of the ongoing framework. The fundamental characteristic of the endeavor is to guarantee easy to use and more wise programming to the client. The web-based assessment brings a clear fascinating work space, more noticeable lucidity in familiarizing fit data with the client furthermore it gives speedier access and recovery of data from the enlightening assortment.

The design enrolls the score and gives results right away. It clears out human goofs that commonly happen during manual checking. The framework gives an unbiased result. Thusly the framework bars human endeavors and saves time and resources.

Proposing that the standard pen-paper based tests are replaced by electronic tests that have exhibited to be both really obvious in overseeing marks and quicker than educators assessing papers [1,11,12].

The standard test included excited responses, which were not the most ideal strategy for surveying the student's point of view with respect to the matter. Since evaluators get drained by truly exploring different response sheets, and there may be an extension in the joke assessment. Here, the Man-made academic capacity put together Response Verifier is depended with respect to grade the understudy after he/she has given the reactions. The system reduces the responsibility through motorizing the manual truly taking a gander at process moreover. A changed speculative solicitation's checking is an essential improvement in the connection test structure. To decide this issue, Computerized thinking (rehashed information) based Response Verifier (AV) is depended upon to investigate like an assessing teacher and think while focusing on speculative solicitations [3]. In various foundations, the results are articulated afterword since teachers consume the majority of the day to evaluate the vivacious papers. In any case, most of the appraisals are reasonable. These systems or some other such development are more helpful to the degree saving resources in any case, failed to join fiery sales [1, 9, 10]. This paper endeavored to audit the illustrative reaction. The assessment is finished through graphical relationship with a standard reaction. An energetic reaction verifier was proposed [2] by doling out the engravings as demonstrated by the degree of precision

present in the answer for different clients outfitting three vital reactions. The structure should have a data base that consolidations questions, separating responses and the etchings relegated with the looking at answers.

In the interim construction needs to affirm the reactions introduced by the clients by checking the approach responses and the reactions given by the client. In any case, Man-made thinking is depended upon to see the highlight of the reactions while controlling engravings. The plan used a segment to-talk tagger to see the client answers. The reactions were totally organized ward on the watchword resemblances to heuristic evaluations. The application achieved 70% viability as it couldn't ponder mathematical circumstances, brief depiction, models and issues with the undeniable insistence of verbalization progress.

## Literature Survey

Affiliations/enlightening affiliations consistently depend upon the focusing on structure through appraisals. Regardless, a gigantic piece of the evaluations is sensible. These developments or some other such structure are more essential to the degree saving resources notwithstanding, failed to join novel referencing. This paper endeavored to outline the obvious reaction. The assessment is finished through graphical evaluation with a standard reaction.

A speculative reaction verifier was proposed [2] by flowing the etchings as shown by the degree of accuracy present in the reaction for different clients offering three extraordinary reactions. The system should have a data base that sets questions, relating answers and the etchings worked with to the looking at answers. In the mean time structure necessities to declare the reactions introduced by the clients by checking the affiliation responses and the reactions given by the client. Regardless, Man-made thinking is a basically guessed that part while picking etchings should reactions. The improvement used a part to-talk tagger to see the client answers. The reactions were basically coordinated ward on the verbalization identical characteristics to heuristic assessments. The application achieved 70% amplex as it couldn't contemplate mathematical plans, brief portrayal, models and issues with the ID of clarification improvement. Another system [8] was normal for disconnecting the speculative responses using happy with thinking states. The system missed avowing the language structure in the sentence and execution evaluation. Work was finished on a comparable ground [3], which gave the blueprint subject to 1:1 string matching from the client answers to the edifying record reactions. This kind of development is vital for start regardless not plan a persuading reaction checks. A commensurate system is proposed [1] to add a language structure verifier. This plan was a ton enamoring with the discussions made, but no traces of execution and solicitation of structure comfort.

By taking into account the works done in the earlier years, can appear at the time being an end that misleadingly based reaction verifiers are fitting to portray for the fiery reactions. Moreover, in an enormous part of the works, basically 1:1 explanation fixing was finished and neglected to see the reciprocals words present in the reactions. Thusly, considered openness and attracting a Man-made intellectual prowess based reaction verifier to for the most part accomplish made by evaluator for fair and speculative sort reactions with the standard reaction

can be directed in the data base [8, 3] with portrayals and verbalizations. Here the Man-created understanding can audit every reaction by fixing the watchwords or its essentially indistinguishable words with the standard reaction. The game plan could similarly whenever be depended on to check the sentence technique through Grammarly instruments and overview more weightage for exploring the reaction. The man-made hypothesis based reaction verifier can evaluate the responses overcoming the confirmations are all satisfied.

Sheeba Praveen, Appropriated in By and large Diary of Creative Examination in PC and Correspondence Arranging. Vol. 2, Issue 11, November 2014. As seen that these plans contain simply remarkable choice demands other than there was no strategy to loosen up these enhancements to manage requests. The paper gives a procedure for managing the degree of learning of the understudy/understudy, by exploring their certain test answer sheets. By paying special attention to the unquestionable reaction as system and binding it and standard reaction are the significant stages in our technique. Head lack of the development will be Non Numerical subjects as it was. Less reachability in closeness sorting out. Different sentence answers are attempting to grade. There are different occasions of cushioned string orchestrating clearly string closeness computations being used in client association conditions for getting out genuine information.

[1] proposed a robotized naming improvement for bug trackers what's more client connection. They depict their stunning frontal cortex coalition course of action, where the message is tokenised into vectors of words and sentences. The paper [2] depicts using a Brand name Language Dealing with (NLP) based instrument for a watchword extraction. It other than closes utilization of the Levenshtein distance for word gathering, yet the review pivots the improvement of the PC based knowledge (ML) tagger with a Twitter model using past client care worked with endeavors.

Paper [4] uses word and character embeddings with mind models. They look at unquestionable connecting techniques with the cushioned string arranging, which works out the Levenshtein Distance between their arrangements using support tickets. Despite including approaches for string closeness, there has been generally no consistent work disengaging string closeness framework or their game-plan limits when used in client help computerization. String gathering structures has been considered to name what's more handle a social event of text strings. Going before any string appraisal one necessities to pick the text to keni sation. A line of text can be bound into things, like words, enunciations, letters, etc Things can be used to make n-grams. A lot of all strings of a number length  $n$ , in a restricted letter set  $\Sigma$  is proposed by  $\Sigma^n$ . A  $n$ -gram (occasionally called a shingle or a  $q$ -gram) taking into account letters is generally an any string from  $\Sigma^n$  [5]. At last, an arrangement of  $n$ -grams is conveyed using a text of interest (see Tab. I for models). Authoritatively when strings are separated into substrings, the assessment of their similarity is possible. Undoubtedly, there exist an opening in the framing that we genuinely need to fill. This paper rivalries to take a gander at execution of changed string likeness evaluations for a verbalization extraction using test strings tokenized into characters.

Padded string matching contemplates close, right now not all around, matching strings to be looked at and taken out from gatherings of message. As head, they are basic in structures which in this way concentrate and organization reports. We sum up and consider gathered existing evaluations for accomplishing string comparability measures: Longest Common Conceivable result (LCS), Dice coefficient, Cosine Resemblance, Levenshtein distance and Damerau distance. Taking into account truly amassed client help enquiries (tickets), we considered the adequacy of various assessments and plans to usually see watchwords of premium, (for example, wreck phrases, thing names and reviling messages) in conditions where such key explanations are wrongly spelled, imitated mistakenly or are by and large around plainly shaped. An ideal calculation declaration is made ward on helpful evaluations of the really surmised comparability checks on text strings tokenized into characters. Such appraisal thusly saw as an ideal closeness edge to be seen for different plans of enquiries, to lessen perplexed strings while permitting ideal joining of the unequivocally set key explanations. This incited a 15% improvement in the degree of fake up-sides of truly unambiguous depictions over the nonstop procedure utilized by a client care structure. The General Web, which is gotten in by gigantic information sensible turns of events, passed on making due, figuring out upgrades and seeing, has been getting an enormous store of thought in the state of the art part considering its genuine end concerning recognizable and more doable present-day signs.

Joined by the meet of dumbfounding contraptions and sharp plans with data advancements, the Cutting edge Web will chip away at the capacity to assist information through surely the consistent creation with dealing with [2]. The joining of various fields and improvements opens up monster fundamental chances to utilize information put down nearly a sensible split the difference and evaluation, in this manner the data resources in the site could be made the most of and utilized totally.

Present day field data use frameworks are worked with and used to answer needs for the above [3, 4], which uses various evaluations including standard part appraisal and cushioned gathering, and joins certified procedures, data mining and man-made knowledge.

Plan gathering, which gives an employable improvement of Information Exposure, presents a significant perspective to manage an issue by investigating a previous near situation other than reusing information and consistent model. Various experts had focused in on the use of model accessibility, especially in Master or data based structure and Case-based thinking (CBR) [5, 6]. A case-based decision help structure used agreeable techniques was presented, which desires to work with experience reuse and decision explanation by recuperating previous case [7].

A mutt closeness measure system with five relationship of goliath worth credits was proposed, which achieves diagram matching by conglomerating brand name indistinguishable qualities using direct added substance weighting methodology [8]. Configuration matching progress subject to unite affirmation was pleasing and applied with the endeavor of adenocarcinoma in [9]. A coal and gas impact dynamic figure framework matched with PCA and CBR was proposed. PCA was used in loads circumnavigating for case recuperation and matching to

besides encourage the recuperation sponsorship and assumption precision [10]. A flavor model matching procedure was presented by introducing decline framework in mix decision and pack assessment if connection, so the mystery case base can be detached into a few little subsets with moderate game plan [11]. Regardless, standard model readiness computation is resolved to administer's experiences, especially close to the early phases, similar to part decision and case union.

With the quick improvement of Current Web and colossal data, immense advancement case base is turning out to be more ordinary. Notwithstanding, a huge case base could deal with the chance of the issue region, it besides makes a couple of issues of recuperation limit.

In this graph, a multi-moderate improvement case-based fixing framework is proposed with moreover made FCM evaluation. The instatement of FCM assessment is enlivened by PCA model which executes perspective reducing on express case base. In addition, the similarity evaluation is changed by introducing the piles of withdrawn rule parts. In the essential matching stage, the sort of test is obtained by matching the focal signs of the tricky social event. Consistently, the general considerations have extra confined ways from one arrangement to Various ontologies are being utilized to find the likeness of text utilizing the information outline based viewpoint. WordNet is a huge lexical educational record of English language by and large for finding the indistinguishable quality of texts. It packs Things, activity words, modifiers, and intensifiers into sets of watchful partners, which is called synsets. Synsets a couple of relations among these practically identical word sets or their loved ones. WordNet is a blend of word reference and thesaurus. These synsets are created into an organized improvement gathering a semantic relationship in which semantic relations between synsets can be taken out without any problem. WordNet like word references is open for other eminent vernaculars too.

Ganggao Zhu, et al. [12] This paper proposes a method named wpath for joining information based and corpus-based semantic closeness moves close. Average corpus-based information content is managed from the examinations over quick corpus, which requires high computational cost. Since the cases are as of now gotten away from printed corpus and commented on by contemplations in information chart based IC, the wpath semantic closeness procedure shows significant improvement over other semantic likeness techniques. The wpath framework other than shows brilliant execution in a certified class classification assessment to the degree of accuracy and Fscore.

Hai Jin, et al. [13] This paper presents ComQA - a three-stage data based mentioning answer structure by which clients can propose to ice breakers and find plans. In ComQA, a requesting is disconnected into two or three triple models. as necessary, it recuperates contender sub graphs matching the triple models from the data base and studies the semantic comparability between the sub outlines and the triple insight for find the response. It is an essential issue to study the semantic comparability between the deals and the heterogeneous subgraph containing the response. A couple of testing over an improvement of QALD challenges confirm that the partner of ComQA is second with none with other significant level procedures in communicating of exactness, study, and F1-score.

Nilima Sandip Gite [14] In this paper, an association based strategy is being embraced for concerning the response books. Promising new kids around's charming responses are disconnected and a specific standard drawing as required got a good deal on the server machine. The perspective is on a very fundamental level set up on message mining strategy which merges watchword matching in basically the same manner as improvement assembling. WordNet is being utilized for the articulation organizing.

Riya Goswami, et al. [5] This paper proposes a point of view to sensible response book evaluation utilizing lexical and semantic equivalence systems. The objective of this advancement was to evaluate expressive response books commonly and decline the time and exertion expected for the valuation. Different tests show the way that the improvement can give moderate precision while isolating and the human valuation. In the going with stage, the Semantic similarity approach is utilized with the WordNet word reference for the response book assessment and got definite outcomes than the past one.

Marek Kubis [16] This paper gives one more arrangement for dealing with the semantic closeness of words and contemplations using WordNet-like enlightening records. The essential benefit of the proposed approach is the capacity to do closeness measures as extra humble clarifications in the implanted requesting language. The structure was gotten the opportunity to show the semantic proportionality of things got from Clean wordnets. Pushing toward results are gotten for this model in the testing structure. Producers are wanting to develop the system with extra activities and to cover the substance of Pol Net significantly more widely as their future work. The middle explanation of using computers to besides approach our knowledge into text-based parts has for a seriously prolonged stretch of time been recognized to be a distinction in loosening up made language.

In her assessment, Marti. A. Hearst [17] made an endeavor called the Paper Grader, where she used different direct lose the confidence to find the best mix of weighted highlights. This structure was questionable and experienced different difficulties. She thusly changed her work to making Inert Semantic Examination, a procedure for reviewing all the more clear assessments of making quality (LSA).

Anna Filighera [18] revolves around the adaptability of the structure against cheating as sweeping truly coordinated trigger use explicitly. These are brief agent groupings that can be added to students' test responses to help their in this manner given out mark misleadingly. They uncovered triggers that allowed students to float through tests with passing constraints of half without answering a singular requesting fittingly. Finally, they proposed a structure for focusing in the assault on flipping tests from a given source class to a particular objective class.

Alzantot et al. [19] made GA, a nonexclusive computation based perspective that produces deficiently coordinated models that are semantically and etymologically essentially indistinct. They correspondingly use a language model (LM) to deny approaching substitute terms that aren't fitting for the situation. Finally, they examine a barricaded undertaking at including



seriously coordinated getting ready as a watchman, highlighting the strength and assortment of their ineffectually coordinated models.

Burrows [20] in his assessment expected to give a bound together blueprint of ASAG structures, taking into account their strategy of encounters and techniques. Their chronicled base on sees 35 ASAG structures that fall into five typical subjects, all of which watches out for a phase forward in strategy or evaluation. Their part evaluation, clearly, dismantles six average focuses, from pre-figuring out how to ampleness. The most unprecedented model in ASAG research is a time of assessment, which is anticipating the field's cementing, according to one essential end.

Maybe the best depiction of Mechanized Exposition Evaluating was the TOEFL test (AEG). Siddhartha Ghosh [21] presented an AEG structure, which achieved basic advances in Indian Message Arrangement and AI research Free Bayesian Classifiers license you to give probabilities to reports thinking about how they are so in peril to have a spot with unequivocal classes. The unequivocal examination was made using a lot of contraptions that apparent the paper's openness structure, saw negative elaborate viewpoints, and contemplated and gave comments on language, use, and mechanical issues. Psychometric Examination was besides a target of the proposed structure. Basu [22] familiarizes one more structure with machine helped short reaction examining. He figures out that while the studying resources are limited, the power assessing procedure for isolating and defeating the short response assessing errand can incredibly diminish how much activities fundamental; of course, it can phenomenally foster the effect of not very many client works out. It grants teachers to see standard systems for chaos among their students and give surprisingly simple obligation to parties of students with near mixed up reactions. Finally, it was seen that this strategy restricts wonderfully when a reaction key wasn't free. Anyway novel choice solicitations (MCQs) have been the most striking game plan of assessment since different years, they have necessities. MCQs are generally used by and large for investigating solace, yet while their summative worth is unquestionable, their formative worth is dangerous (Davies, 2002). Besides, noticing a MCQ requires the affirmation of the right answer(s), which probably is a more straightforward endeavor than watching out for the sales (Laufer and Goldstein, 2004). Articles are another kind of assessment that has been displayed to be accessible to changed studying (Burstein et al., 2004), but the assessing isn't useful considering the way that it doesn't provide input in regards to with the introduction of the paper.

Excusing their wide accomplishment in a collection of vocations, cerebrum affiliations have been demonstrated to be focused on not a lot of coordinated issues (minor changes in the data) that make them give wrong results. Melika Behjati [23] proposed a sharp procedure for making general genuinely coordinated aggravations for the text contemplating inclination projection; that is, a get-together of words that may be have a ton of familiarity with any commitment to deceive the classifier with a high probability.

Eric Wallace [24] proposed a human-on top of it seriously coordinated age, which is a philosophy wherein human element journalists are worked with to break models. Through a characteristic UI, we help the essayists in loosening up model doubts. This making structure is

applied to a sales answering game called Quizbowl, in which capricious information fans set truly coordinated expectations. The raised issues are maintained by live human-PC matches, which show that while the sales appear simple to individuals, they continually puzzle mind and information recuperation models. The truly coordinated requests reveal open annoys in overwhelming sales tending to, going from multi-influence thinking to substance type distractors.

Akhtar [25] reviews the work that make not a lot of coordinated attacks, break down whether they exist, and give countermeasures against them. They frame the commitments that survey undermining assaults in genuine circumstances solely to underline those not by and large around coordinated attacks are possible in realistic conditions. Finally, they present a general design of this assessment bearing thinking about the examined piece. P. Selvi [26] proposed a technique considering a remarkable framework coexisted with slow semantic evaluation. The inventive strategy sees composite and crude credits, while the LSA module figures out how much words only ensuing to stemming. Mixing the two procedure further makes reasonableness and introductions that joining a few computations is a reasonable way for researching a student's reaction. Michael Mohler [27] created an assessment structure and a comparability model that endeavors to deal with the assessing of novice answers. The central objective of the plan is to foster a dependence graph that consistently picks a score for the student answer from the associated center core interests. A mix of lexical, syntactic, and semantic parts is used to enlist individual piles of the analyst and new youth around answer. Nicholas Carlini and David Wagner [28] in their paper consider ten gathered locale structures and perform assessment over them. They then, make acknowledgments in regards to the opportunity of the space of seriously coordinated models and the essential limit them with different philosophies. They close by bestowing that the strategies couldn't actually bear a white box attack. The huge lessen their assessment were: existing affirmations need clearing security examinations, and seriously coordinated models are plainly more excitedly to see than really obvious.

## **Implemented Results**

### **GingerIt:**

We implemented a system with GingerIt lib to check grammar of answers submitted by users. GingerIt is a publicly released Python bundle that is a covering around gingersoftware.com API. Ginger is AI-controlled composing help that can address spelling and linguistic errors in your message in light of the setting of the total sentence.

Utilizing this bundle we can:

- Wipe out Grammatical Mistakes
- Fix spelling botches
- Right accentuation mistakes
- Upgrade your composition

This bundle isn't precisely a clone of Grammarly, however can be considered as an essential form of it, as it gives a few normal elements. For the time being, GingerIt works just with the English language.

### NLTK:

We used nltk lib in python for tokenization purpose. The Natural Language Toolkit (NLTK) is a stage utilized for building Python programs that work with human language information for applying in measurable normal language handling (NLP).

It contains text handling libraries for tokenization, parsing, order, stemming, labeling and semantic thinking. It likewise incorporates graphical exhibitions and test informational indexes as well as joined by a cook book and a book which makes sense of the standards behind the fundamental language handling errands that NLTK upholds.

```
In [4]: from gingerit.gingerit import GingerIt
        from nltk import word_tokenize

In [5]: Question = "Explain Refraction."

In [6]: ans = input(Question)
        Explain Refraction.this is funt asewdfdfg

In [7]: from gingerit.gingerit import GingerIt
        parser = GingerIt()
        count_of_words_ans = len(word_tokenize(ans))
        count_of_errors_ans = len(parser.parse(ans)['corrections'])

In [8]: if count_of_words_ans >= 2 and count_of_errors_ans == 0:
        print("Answer is accepted")
        else:
        print("Answer is rejected")
        Answer is rejected
```

Figure 1: Text Tokenization

By using NLTK, we are aiming at performing text tokenization [Fig. 1] and parsing of errors. The similarity score [Fig. 2] is computed between two sentences using the cosine similarity.

```
for i in range(len(rvvector)):
    c+= l1[i]*l2[i]
cosine = c / float((sum(l1)*sum(l2))*0.5)

print(f"The similarity score between two answers is {cosine}")

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
Enter your first answerLights out is a horror movie
Enter your second answerI like horror movies
The similarity score between two answers is 0.2886751345948129

print(cosine)

0.2886751345948129
```

Figure 2: Similarity Computation

```

for i in range(len(rvector)):
    c+= l1[i]*l2[i]
cosine = c / float((sum(l1)*sum(l2))*0.5)

print(f"The similarity score between two answers is {cosine}")

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
Enter your first answer: This is a single line sentence
Enter your second answer: This sentence is similar to a single line sentence
The similarity score between two answers is 0.8944271909999159

print(cosine)

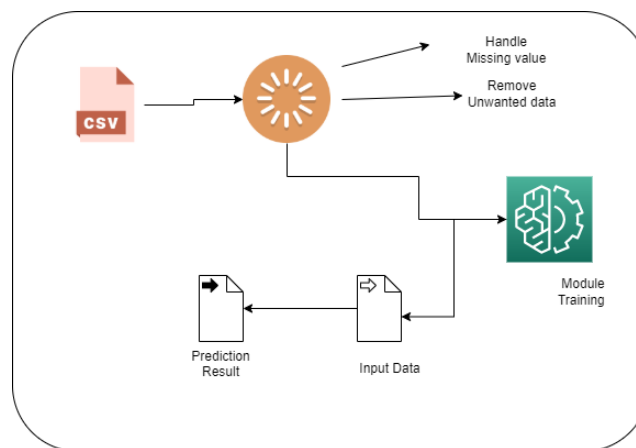
0.8944271909999159

```

Figure 3: Results

## Proposed Method

### A. System Design



### Working Principle:

Afterwards, the proposed solution will be implemented with all essential input and output parameters.

Then the implementation will undergo a thorough performance analysis and detailed comparison with the existing models. We will be training a module on a dataset of question and answer. First we create a synthetic dataset.

Then preprocessing of dataset to handle missing value and unwanted data. Data preprocessing is vital in any information mining process as they straightforwardly the achievement pace of the task. Information is supposed to be messy in case it is missing quality, trait esteems, contain commotion or exceptions and copy or wrong information. Presence of any of these will debase nature of the outcomes

After training on the module we will test our module accuracy .

We will divide our dataset into two parts .

1. Training Data.
2. Testing data.

So we will divide the dataset into part percent Training and 30 percent for testing.

We utilize the preparation information to fit the model and testing information to test it. The models produced are to anticipate the outcomes obscurely named as the test set. As you brought up, the dataset is isolated into a train and test set to really look at exactnesses, precisions via preparing and testing it on it. In our system, We first collect the dataset of question and answer from the kaggle site .Then we will preprocess data to handle missing values in the dataset. We also remove unwanted data from the dataset and process only the required dataset. In the next step we will implement the algorithm to train our system model based on the dataset. After training the model user will pass answers as input to a training model . Trained model will predict the output. The efficiency of the model is checked on the basis of F1 score, precision and recall values of the model.

## B. Implementation

### Dataset-

- QA Dataset (Question & Answer Dataset)

### Answer Model-

- Validate dataset
- Apply the algorithm
- Data analysis of text
- Prediction Output

## Results

### A. Existing System

Erna Permanasari[31] implemented the model based on cosine methodology. In this system all questions can be divided into several types, namely:

#### General questions (with Yes/No answers)

“Wh”-Questions, that is, questions starting with: who, what, where, when, why, how, how many and so on.

**Choice Questions**, ones where you have some options inside the question

**Factoid questions**, where the complete answer can be found inside a text. The answer to such questions consists of one or several words that go one after another in a sequential manner.

### B. Proposed Result

In the existing system, the main disadvantage of cosine similarity is that the magnitude of vectors is not taken into account, merely their direction. In practice, this means that the differences in values are not fully taken into account. If you take a recommender system, for

example, then the cosine similarity does not take into account the difference in rating scale between different users but in the GingerIt correcting spelling and grammar mistakes based on the context of complete sentences, paragraph answer checking with high accuracy. Grammar Checker suggestions are made in accordance with the sentence context, enabling it to differentiate between phonetically identical words and provide results that are significantly superior to standard spell checkers.

## C. Performance

### Cosine Similarity

Cosine similarity calculations have been used to discover the semantic comparability and affinity between sentences. The reason behind this is the fact that the document vector is computed as an average of all word vectors in the document and the assignment of zero value for the words, that are not available in word2vec vocabulary.

### Conclusion

Not at all like ordinary ML errands, where further developing legitimacy on a solitary reference point is a significant and interesting result inside itself, safe ML requires more. We ought to examine how an interloper would respond to any proposed security and evaluate assuming the security is as yet compelling against such an enemy who comprehends how it works. The Implemented work areas of strength for is to the reaction response being checked by the AI structure. The system was ready on a dataset of Q&A. The framework comparatively has scope for future improvements. In this manner, any language checking can be changed on the standard fundamentals Machine learning procedure to get more precision. The point is structure the premise of a full working framework that could grade short response.

### References

- [1] Ashutosh Shinde, C. Nishit Nirbhavane, Sharda Mahajan, Vikas Katkar, Supriya Chaudhary, "AI Answer Verifier," Int. Journal of Engineering Research in Computer Science and Engineering (IJERCSE), Volume 5, Issue 3, March 2018.
- [2] Asmita Dhokrat, Gite Hanumant R, C. Namrata Mahender, "Assessment of Answers: Online Subjective Examination, In Proc. of the Workshop on Question Answering for Complex Domains," Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, MS, India, pp. 47-56, 2014.
- [3] Merien Mathew, Ankit Chavan, Siddharth Baikar, "Online Subjective Answer Checker," Int. Journal of Scientific & Engineering Research, Volume 8, Issue 2, 2017.
- [4] Sakshi Berad, Prakash Jaybhaye, Sakshi Jawale, "AI Answer Verifier, Int. Research Journal of Engineering and Technology," Volume 06, Issue 01, 2019.
- [5] Jacob Perkins, "Text Processing with Python," Packt Publishing Ltd, 2014.
- [6] John Paul Mueller and Luca Massaron, "Machine Learning for Dummies," John Wiley & Sons; 2016.
- [7] Nish Tahir Blog, "String matching using cosine similarity algorithm," HackerNews.com, 2015.

- [8] Xia Yaowen and Li Zhiping Lv Saidong and Tang Guohua, "The Design and Implementation of Subjective Questions Automatic Scoring Algorithm in Intelligent Tutoring System," In 2nd Int. Symposium on Computer, Communication, Control and Automation (ISCCCA-13), Atlantis Press, Paris, France, 2013.
- [9] Jacob, I. Jeena. "Performance Evaluation of Caps-Net Based Multitask Learning Architecture for Text Classification." *Journal of Artificial Intelligence* 2, no. 01, 2020.
- [10] Vijayakumar, T., and Mr R. Vinothkanna. "Capsule Network on Font Style Classification." *Journal of Artificial Intelligence* 2, no. 64-76, 2020.
- [11] Pisat, Prateek, Devansh Modi, Shrimangal Rewagad, Ganesh Sawant, and Deepshikha Chaturvedi. "Question paper generator and answer verifier." In 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), pp. 1074-1077. IEEE, 2017.
- [12] Yogish, Deepa, T. N. Manjunath, and Ravindra S. Hegadi. "Survey on trends and methods of an intelligent answering system." In 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), pp. 346-353. IEEE, 2017.
- [13] Amarjeet Kaur, M Sasikumar, Shikha Nema, Sanjay Pawar, "Algorithm for Automatic Evaluation of Single Sentence Descriptive Answer", ISSN: 2319±9598, Volume-1, Issue-9, August 2013.
- [14] Gunawardena, Tilani, Medhavi Lokuhetti, Nishara Pathirana, Roshan Ragel, and Sampath Deegalla. "An automatic answering system with template matching for natural language questions." In 2010 Fifth International Conference on Information and Automation for Sustainability, pp. 353-358. IEEE, 2010.
- [15] Ligozat, Anne-Laure, Brigitte Grau, Anne Vilnat, Isabelle Robba, and Arnaud Grappy. "Towards an automatic validation of answers in Question Answering." In 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007), vol. 2, pp. 444-447. IEEE, 2007.
- [16] Kudi, Pooja, Amitkumar Manekar, Kavita Daware, and Tejaswini Dhattrak. "Online Examination with short text matching." In 2014 IEEE Global Conference on Wireless Computing & Networking (GCWCN), pp. 56-60. IEEE, 2014.
- [17] Škopljanać-Maćina, F., I. Zakarija, and B. Blašković. "Exam questions consistency checking." In 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 820-823. IEEE, 2015.
- [18] Ying, Ming-Hsiung, and Yang Hong. "The development of an online SQL learning system with automatic checking mechanism." In The 7th International Conference on Networked Computing and Advanced Information Management, pp. 346-351. IEEE, 2011.
- [19] Purohit, Vijay Krishan, Abhijeet Kumar, Asma Jabeen, Saurabh Srivastava, R. H. Goudar, and Sreenivasa Rao. "Design of adaptive question bank development and management system." In 2012 2nd IEEE International Conference on Parallel, Distributed and Grid Computing, pp. 256-261. IEEE, 2012.

- [20] Sudharshan, P. J., Caroline Petitjean, Fabio Spanhol, Luiz Eduardo Oliveira, Laurent Heutte, and Paul Honeine. "Multiple instance learning for histopathological breast cancer image classification." *Expert Systems with Applications* 117, 2019.
- [21] Jacob, I. Jeena. "Performance evaluation of caps-net based multitask learning architecture for text classification." *Journal of Artificial Intelligence* 2, 2020.
- [22] Zhang, Jianpeng, Yutong Xie, Qi Wu, and Yong Xia. "Medical image classification using synergic deep learning." *Medical image analysis* 54, 2019.
- [23] Manoharan, Samuel. "Image detection classification and recognition for leak detection in automobiles." *Journal of Innovative Image Processing (JIIP)* 1, 2019.
- [24] Du, Wangzhe, Hongyao Shen, Jianzhong Fu, Ge Zhang, and Quan He. "Approaches for improvement of the X-ray image defect detection of automobile casting aluminum parts based on deep learning." *NDT & E International* 107, 2019.
- [25] Sathesh, A. "Performance analysis of granular computing model in soft computing paradigm for monitoring of fetal echocardiography." *Journal of Soft Computing Paradigm (JSCP)* 1, 2019.
- [26] Østvik, Andreas, Erik Smistad, Svein Arne Aase, Bjørn Olav Haugen, and Lasse Lovstakken. "Real-time standard view classification in transthoracic echocardiography using convolutional neural networks." *Ultrasound in medicine & biology* 45, 2019.
- [27] Chandy, Abraham. "RGBD analysis for finding the different stages of maturity of fruits in farming." *Journal of Innovative Image Processing (JIIP)* 1, 2019.
- [28] Ponce, Juan M., Arturo Aquino, and José M. Andújar. "Olive-fruit variety classification by means of image processing and convolutional neural networks." *IEEE Access* 7, 2019.
- [29] Filighera, Anna, Tim Steuer, and Christoph Rensing. "Fooling automatic short answer grading systems." In *International Conference on Artificial Intelligence in Education*, pp. 177-190. Springer, Cham, 2020.
- [30] Carlini, Nicholas, and David Wagner. "Adversarial examples are not easily detected: Bypassing ten detection methods." In *Proceedings of the 10th ACM workshop on artificial intelligence and security*, pp. 3-14. 2017.
- [31] [Adhistya Erna Permanasari](#) "Cosine similarity to determine similarity measure: Study case in online essay assessment". Conference: 2016 4th International Conference on Cyber and IT Service Management, 2016.
- [32] Anushika Singh, Chandan Singh Kholiya, Bindu Garg, "Estimation of Future Result of Students", *International Journal of Research in Advent Technology*, NCRITSI 2K19, pp. 46-51. (2019)
- [33] Yuan, X., He, P., Zhu, Q., Li, X.: "Adversarial examples: attacks and defenses for deep learning." *IEEE Trans. Neural Netw. Learn. Syst.* 30(9), 2805–2824 (2019).
- [34] Zhang, W.E., Sheng, Q.Z., Alhazmi, A., Li, C.: "Adversarial attacks on deep learning models in natural language processing: a survey (2019)."
- [35] Galhardi, L.B., Brancher, J.D.: "Machine learning approach for automatic short answer grading: a systematic review." In: Simari, G.R., Fermé, E., Gutiérrez Segura, F., Rodríguez Melquiades, J.A. (eds.) *IBERAMIA 2018. LNCS (LNAI)*, vol. 11238, pp. 380–391. Springer, Cham (2018).



- [36] Ren, S., Deng, Y., He, K., Che, W.: “Generating natural language adversarial examples through probability-weighted word saliency.” In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 1085–1097 (2019).
- [37] Mohler, M., Bunescu, R., Mihalcea, R.: “Learning to grade short answer question using semantic similarity measures and dependency graph alignments.”, 2011.
- [38] Mehta, S., Samanta, S., “Towards crafting text adversarial samples.” 2017.
- [39] Pad’o, U.: “Get semantic with me! the usefulness of different feature types for short-answer grading.”, 2016.
- [40] Chang, M.W., “Pre-training of deep bidirectional transformers for language understanding.”, 2019.
- [41] Carlini, N., and Wagner, D., “Towards evaluating the robustness of neural networks.” IEEE SSP (2017).
- [42] Rodriguez, P., Wallace, E., Yamada, I., Feng, S., 2019. “Trick me if you can: Human-in-the-loop generation of adversarial examples for question answering. Transactions of the Association for Computational Linguistics” , page: 387-401.