Evaluating Descriptive Answer Using Machine Learning And Natural Language Processing

T. Sai Kumari¹, Hemant Kumar², Shaik Ahmed³, Janvee Sree⁴, and Josna Sravani⁵

¹Assistant Professor, Malla Reddy Institute of Technology and Science, Hyderabad-500100
²Student, Malla Reddy Institute of Technology and Science, Hyderabad-500100
³Student, Malla Reddy Institute of Technology and Science, Hyderabad-500100
⁴Student, Malla Reddy Institute of Technology and Science, Hyderabad-500100
⁵Department of Computer Science & Engineering

Abstract
Assessment of the descriptive paper is a tough challenge. It requires lot of man power. In educational institutions a primary hassle is evaluation of descriptive type questions & answers. Using NLP this descriptive answer assessment proper answers and the pupil answer scripts. More marks is probably awarded, if the similarity with inside the answers is more. There are several advantages of using this automated assessment device as it permits in decreasing the time concerned about the resource of the usage of faculty to correct the descriptive papers. The examinations can be accomplished online, the answers can be evaluated right now and it would be beneficial for universities, schools and schools for academic purpose with the resource of the usage of presenting ease to colleges and the examination evaluation cell.

Keyword: Descriptive answer evaluation, machine learning, natural language processing, word2vec.

I. INTRODUCTION

Students’ academic usual overall performance is assessed at the concept of the examination, each as a descriptive or intention pattern. In the present scenario, manual evaluation of descriptive answers is a hectic challenge. There are multiple systems that may have a look at intention type or MCQ short. The pre-defined correct answers are furnished to the machine.

This only permits only in competitive exams evaluation. The descriptive examinations are the backbone of all the schools and board diploma examinations. Strong tests are seen as truly confusing and disturbing with the resource of the utilization of every single school students and teachers considering their one central brand name, setting. A dazzling reaction needs the checker test every statement of the reaction for scoring, truly, and the checker’s highbrow achievement, exhaustion, and objectivity play an immense brand name within the standard quit result. Hence, it is a thoroughly examined plan extra time and resource standard to permit a contraption address this long and truly central starter of contemplating spellbinding reactions. Examining point type answers with machines is clear test. A thing can be overseen questions and single word answers that could short guide school students’ responses. Regardless, clear reactions are a striking plan more completion to make due. They are different long and set a titanic degree of language. Also, people commonly by and large will regularly use similar words and open shortenings, which makes the machine that an astounding blueprint captivating.

Large type of pupil attend descriptive type exam. Bases upon the descriptive answer, the evaluator will recognize the amount of information the pupil has received throughout his/her academic, on which evaluator assign the marks to the pupil. Manual Assessment of descriptive answers is a completely tiresome, time-consuming challenge and requires masses of manpower. Answer evaluation varies from evaluator to evaluator consistent with how evaluator performs the evaluation, mood of evaluator at time of evaluation of the answer sheet and behaviour/rela-tion amongst pupil and evaluator. This also can moreover influences the quit end result of the pupil.

Much craftsmanship has been finished concerning expressive reactions evaluation in an unmarried new improvement obviously if nothing else a couple, which joins, finding the setting aft of the message and arranging it with the framework’s fundamental circumstance, counting the thing verbalization within the reports, exploring closeness
among momentous messages, terms, or maybe records, matching key terms withinside the reactions, and so on. Taking into account everything, issues which concretes tf-idf loosening semantic setting, nonattendance of hyper-limits tuning, rich readiness, and need for better datasets notwithstanding exist.

II. In this paper, our huge stone depends all over on standard tongues coordinating procedures which concretes tokenization, lemmatization, message truly zeroing in on structures which works with tf-idf, heap of terms, word2vec, closeness focusing in on approaches which joins cosine similarity, and word mover’s distance, importance framework which sets multinomial credible bayes. We use grand examination checks which sets f1-score, accuracy, and consider to advance toward the overall in standard showcase of several models being suggested every specific. We almost talk several structures used withinside the past for drawing in reactions assessment or text closeness appraisal in regarded

II. LITERATURE SURVEY

Kusner et al. [7] gave a completely striking thought of using word mover’s distance (weapon of mass destruction) to track down the partition among texts. The contraption used no hyperparameters and used an astounding weapon of mass pounding structure to loosen up the vector place limits. Dataset included eight confirmed everything pondered units, similar to twitter feeling pieces of data and bbc sports practices sports articles. Word2vec model from google pieces of data come to be used, and clear custom models were told. Knn importance method come to be used to work with the sorting out experiments. As a quit conceivable outcome, fulfilling weapon of mass destruction decreased the mistake charges and prompted 2 to different times speedier significance.

Kim et al. [32] proposed a structure for investigating short verifiable reactions lexico-semantic model (lsp) taking into account its guaranteed standard by and large with morphologically complex korean language. A muddled once-over come to be moreover finished to help with extra uplifting the enormous thing terms basically so they in shape several answer styles. Dataset come to be gotten from 88 school students and changed over totally to lsp, which come to be later in evaluation with the outline lsp to acquire the reaction.

Oghbaie and zanjireh [33] proposed a couple-outstanding closeness explanation to affirmation the proportionality among reports based unequivocally at the essential thing terms which appear in something like one of the records. The proposed similarity clarification come to be finished to message mining applications which sets reports license, awesome nearest neighbors (knn) for unmarried-name significance, and kmeans gathering. An assessment interest of precision come to be used, and as a quit result, the pdsm thinking conveyed managing results than express assessments close as the jaccard coefficient with the resource of the use of 0.08 consider.

Orkphol and yang [34] used the word2vec improvement to address terms on a fix-sized vector put model and as such turned the comparability of sentences using a cosine closeness interest. Word2vec from google come to be used, and the sentence vector come to be gotten considering a standard of terms withinside the sentence. The score come to be regarded in case it passed a specific end for regard results, among 0 and 1. Evaluation interest of consider and exactness come to be used, and as a quit result, the contraption’s common overall to be 50.9% with and 48.7% without the opportunity of revel being made.

Xia et al. [8] mixed the word2vec structure with the undermining report corpus to see the worth in similarities among stand-separated rule records. As a quit outcome, word2vec managed the accuracy with the asset of the use of 0.2 in evaluation to the sack of terms structure, that might in basically a near way whenever at some point at long last be associated with the asset of the utilization of 0.05-0.10 through setting up the word2vec model on rule stories.

Wagh and anand [35] proposed a multi-necessities decisionmaking mentality to find the similarity among reprobate records. Dataset come to be gotten from indian confounding court plan case decisions from years starting from 1950 to 1993.As a quit result, a thought based distinctly closeness system which solidifies the best proposed withinside the craftsmanship accomplished better veered from express methodology which coordinates tf-idf, getting a f1-score of as a ton as 0.8.

Alilan and awajan [36] focused in on various parts affecting sentence closeness and summing up express authentication using confounding word introducing models, gathering examinations, and weighting remembering to track down the setting of sentences. Pre-shown embeddings included aravec and fasttex, each educated for the arabic language. The arabic orchestrating dataset included cycle 77,600,000 tweets. As a quit result, pre-competent giving named evaluations from experts worked with better consider and exactness of 0.87 and 0.782 for good
enough-thinking and agglomerative party.

III. TECHNICAL BACKGROUND

Natural Language Processing
Natural language processing (nlp) refers to the branch of computer science—and more specifically, the branch of artificial intelligence or ai—involves giving pc structures the capability to apprehend text and spoken terms in a exceptional way human beings can. Nlp permits developers to put together and form information to perform obligations like translation, summarization, named entity recognition, relationship extraction, speech recognition, and situation remember segmentation.

i) Stemming
Stemming is the machine of decreasing inflection (prefix, suffix) in terms to their root paperwork which incorporates mapping a hard and fast of terms to the same stem no matter the truth that the stem itself isn’t a valid word within the language. Stemming will take away suffix and prefix from the word.

ii) Lemmatization
This accumulates normally twisted kinds of a word to break them genuine all of how directly down to their root word reference improvement or lemma. Words are isolated into a semantic part (such word types) with the resource of the utilization of system for the considerations of supplement. The way to lemmatization is semantics.

iii) Stopwords
Stopwords are the most now no longer unusualplace terms in any natural language. They are the terms in any language which does now not add a exceptional deal this means that to a sentence. They can well be omitted without sacrificing the means that of the sentence. For some are seeking engines, the ones are some of the most now no longer unusualplace, short feature terms, which incorporates the, is, at, which, and on.

Machine Learning Algorithms
Naïve Bayes
Naïve bayes set of guidelines is a supervised reading set of guidelines, that is based mostly on bayes theorem and used for solving magnificence issues. Naïve bayes are at the complete applied in natural language processing (nlp) issues. Naïve bayes are anticipating the tag of a text. They calculate the possibility of each tag for a given text and then output the tag with the high-quality one. The machine for bayes theorem is given as

\[
P(A|B) = \frac{P(B|A)P(A)}{P(B)}
\]

Posterior probability: Probability of hypothesis on the observed event B.

P(B|A) is Likelihood probability: Probability of the evidence for the purpose that the possibility of a hypothesis is true. P (A) is Prior Probability: Probability of hypothesis in advance than observing the evidence.

P(B) is Marginal Probability: Probability of Evidence. It is a exceptional and at the complete used magnificence Algorithm. However it gives a lot much less accuracy for this model.

Decision Tree Classification
Decision tree is an arranged exploring structure that can be used for every significance and break conviction issues, yet at the full scale it is leaned toward for settling quality issues. It is a tree-presented classifier, wherein inside centers address the gifts of a dataset, branches address the propensity evaluations and each leaf locale truly bases on the outcome. It is a tree-presented classifier, wherein inside centers address the gifts of a dataset, branches address the decision assessments and each leaf put genuinely bases on the outcome. In a penchant tree, there are centers, which can be the statement place point and leaf center point. Decision center circumstances are used to make any proclivity and brand name different branches, in the meantime as leaf centers are the possible surrendered outcome of those different choices and do now not join any covering branches. Different choices or the test are accomplished at gifts of the given dataset. It is a graphical event for getting each of the sensible responses for an issue/tendency ward generally talking upon given conditions. It is proposed as an inclination tree considering the way that, similar to a tree, it starts off evolved with the fantasy locale, which becomes other than branches and makes a tree-like turn of events.
IV. PROPOSED SYSTEM

The proposed device encompass the following things: it have 2 modules, 1 for faculty and the alternative 1 for
the pupil. every the pupil and the faculty has a login page, if they may be now not registered then they might get
registered with the resource of the usage of growing their account.

The 1st module is the faculty section, they might set the question paper for the exam .faculty need to create a test
with the resource of the usage of the usage of of together with test name, date and time found with the resource of the usage of
of together with the question alongside aspect marks and actual answer for each question.

The 2nd module is the pupil section .the pupil need to login and the exam is probably visible. There is probably a
timer consistent with which the exam will start on and quit . After attempting the exam the test want to be
submitted. After attempting the exam with the resource of the usage of the pupil those answer is probably checked
with the ML model. These answers is probably checked on basis of the answer duration, key-phrase test, grammar
test, context & semantic similarity and cosine similarity in competition to model answer.

After obtaining all values for the above parameter, machine reading model is used to get the consequences
.each question is evaluated with the resource of the usage of of ML model and marks are also allotted definitely
upon the answer given with the resource of the usage of the pupil. If the instructor found that the answer isn't
accurate as in line with the model answer that come to be assigned with the resource of the usage of set of
guidelines then instructor can exchange the marks manually.

![Proposed System diagram]

Fig. . Proposed System diagram

Modules for Evaluation of Answer

Answer length
In the existing device, we found that the pupil gets a entire mark on writing only key terms. And so, answer duration
is also an critical aspect. As pupil also can moreover write all key terms and grammatically correct short sentences.
Such answer may get entire marks for key terms and grammar but a lot much less marks for short duration.

Keyword Matching:
Keywords is a few different crucial component which we don’t forget to check whether or not or now no longer
student cover all crucial idea (keypoint) in his answer or now no longer. If the important thing-phrase fits the
maximum marks is probably allotted as in step with the important thing-phrase considering its correctness. But if few
the important thing terms variety with the model answer key terms then the marks is probably deducted accordingly.

Contextual Similarity
If the set of rules reveals the similarities among the 2 solutions. if solutions aren’t contextual comparable then it’ll
deliver 0 else it deliver the similarity percentage.

Semantic Similarity
If solution are contextual comparable then we’ve taken into consideration similarity among solutions with recognize
to contextual which means through the use of API.
Contradiction

Contradiction detection is a mission to understand contradiction family members among a couple of sentences. Despite the effectiveness of conventional context-based totally phrase embedding getting to know algorithms in lots of herbal language processing tasks, such algorithms aren’t effective sufficient for contradiction detection. Contrasting phrases such as “overfull” and “empty” are frequently mapped into near vectors in such embedding space. we use pos tagger to locate contradicting sentences from pupil solution towards version solution.

Grammar Check

To test and come across grammatical errors and spelling mistakes Grammar, Textgear API is utilized in our device. When the textual content is despatched to Grammar Textgear’s API, it returns some of grammatical blunders found in textual content

Antonyms and Synonyms:

NLTK wordnet is a lexical database of English language. Nouns, verbs, adjectives and adverbs which are grouped into units of cognitive synonyms (synsets), every expressing a wonderful concept. The lemmas() feature of the synset. It returns synonyms in addition to antonyms of that unique synset. Synonyms are taken into consideration to locate semantic similarity among distinct solution having identical contextual which means. Antonyms are taken into consideration to locate conflicting or contradicting sentences from pupil solution and version solution.

V. CONCLUSION

This paper proposed a descriptive solution checker device the use of nlp and device getting to know Descriptive solution assessment software program assign mark to descriptive query primarily based totally on solution length, key-word matching, grammar test, cosine similarity and contextual similarity towards version solution furnished through college and pupil solution. We additionally advanced set of rules to locate contradicting statements from pupil solution with version solution. Even solution written through pupil does now no longer suit phrase to phrase with version solution furnished through teacher, however nevertheless our device succesful to assess solution primarily based totally on context

VI. ACKNOWLEDGMENT

With deep feel of gratitude we renowned the steering of our venture manual Asst. Professor Mrs. T. Sai Kumari the time-to-time help and encouragement through her has performed an essential position within side the improvement of our venture and made our adventure successfull.

VII. REFERENCES

32. Dr. V. Senthil kumar, Mr. P. Jeevanantham, Dr. A. Viswanathan, Dr. Vignesh Janarthanan, Dr. M. Umamaheswari, Dr. S. Sivaprabhakar. Emperor Journal of Applied Scientific Research “Improve Design and Analysis of Friend-to-Friend Content Dissemination System “Volume - 3 Issue - 3 2021
33. S.B. Jayabharathi, S. Manjula, E. Tamil Selvan, P. Vengatesh and V. Senthil Kumar” Semantic Risk Analysis Model Cancer Data Prediction” Journal on Science Engineering and Technology Volume 5, No. 02, April 2018