



# Subjective Answer Evaluation Using Machine Learning And Natural Language Processing

Karthika. K<sup>1\*</sup>, Akshaya. G<sup>2</sup>, Manoshree. T<sup>3</sup>

<sup>1</sup>Assistant Professor, Department of Computer Science and Engineering, Anjalai Ammal Mahalingam Engineering College, India.

<sup>2,3</sup>Student, Department of Computer Science and Engineering, Anjalai Ammal Mahalingam Engineering College, India.

\*Corresponding author

DoI: <https://doi.org/10.5281/zenodo.7883201>

## Abstract

The manual evaluation of subjective papers is a difficult and time-consuming task. In education and other professions, evaluating subjective responses is a crucial activity, but it is frequently subjective and time-consuming. In this study, we suggest a machine-learning and natural language processing-based automated method for assessing subjective responses.

The system will employ NLP approaches to extract variables like word frequency, phrase length, and sentiment analysis after being trained on a dataset of graded subjective replies. The grades of fresh subjective answers will subsequently be predicted using machine learning methods based on these extracted features. We will compare our system's performance to other methods already in use for evaluating subjective answers using measures like accuracy, precision, and recall. Additionally, we will conduct cross-validation to make sure that our model applies well to fresh data. The project has the ability to provide objective and consistent evaluations while also drastically reducing the time and effort needed for subjective answer evaluation. It can be used in a variety of ways in education and other industries, like e-learning tools and online education platforms. By using these technologies to address a real-world issue, it can also advance the fields of machine learning and NLP. Using cutting-edge machine

---

learning and NLP approaches, this project seeks to automate and enhance the subjective answer evaluation process in order to increase its effectiveness, accuracy, and dependability.

**Keywords:** Subjective answer evaluation, natural language process, machine learning, grades of the answer.

---

## 1. Introduction

In the current digital era, technology has fundamentally changed a number of industries, including education. Subjective answer grading is still a manual, time-consuming job that usually involves subjectivity and error. To address this issue, we published the project "Subjective Answer Evaluation using Machine Learning and NLP." The objective of the research is to develop an automated system that can effectively evaluate subjective answers by combining machine learning and natural language processing. In order to score a subjective response, the checker must carefully read each word in the response. The total outcome is greatly influenced by the checker's mental state, level of exhaustion, and objectivity. Therefore, allowing a machine to undertake this time-consuming and fairly important duty of

Additionally, it can help in providing objective and consistent judgements, reducing the likelihood of subjectivity and errors. Academic institutions, e-learning platforms, and online learning systems may all be applications for the project. The application of these methods to a real-world problem can enhance the study of machine learning and natural language processing. Overall, this endeavor has the potential to transform subjective answer evaluation and raise its efficacy, reliability, and correctness.

## 2. Literature Review

1. **Kissan G. Gauns Dessai; Venkatesh V. Kamat; Ramrao S. Wagh et al Proposed” Effective Use of Rubrics in Computer Assisted Subjective Answer-Script Evaluation” IEEE-202**

Compiling the results and evaluating student exam answer sheets are two time-consuming but essential post-exam duties. Errors in marking, tabulating, recording, and computation are frequently present in the manual assessment and result compilation process, which also frequently includes the influence of the Examiner's subjectivity or variability. Most of the approaches now in use work to only partially resolve these issues by adding more time and labor. Adopting a consistent strategy would promote uniformity across examiners and would allow for a more effective response to problems related to evaluation and result compilation. This work offers a practical and empirically developed rubric-based computer assisted evaluation of subjective answer scripts. The proposed strategy is focused on enhancing the assessment and result compilation duties by minimizing/eliminating errors and examiner

2. **Ragasudha Ragasudha; M. Saravanan et al Proposed” Secure Automatic Question Paper Generation with the Subjective Answer Evaluation System”IEEE-2022**

The necessity for automation in the educational system is now crucial since individuals are leaning towards automation in the modern world to save time and make work easier. Both career progress and personal development depend heavily on education. Teachers currently work very hard to create a question paper based on the syllabus and evaluate the answer sheet. More time and labor must be expended manually for the question preparation and answer evaluation. In addition to describing safe automatic question paper production, this study assesses computerized answers in a subjective manner. The administrator must construct a database that is the form of questions with Bloom's taxonomy in order for the automatic question paper production to work. through the

generate.

### 3. Chhanda Roy; Chitrita Chaudhuri et al Proposed “Case Based Modeling of Answer Points to Expedite Semi-Automated Evaluation of Subjective Papers”IEEE-2019

Page | 127

The automation of examination systems has been the subject of past and present research. However, the majority of them aim for online tests that, at best, only allow for extremely brief descriptive or choice-based replies. In order to simplify a semi-automated evaluation process, this paper's main objective is to present a system in which textual papers with subjective question types are augmented with model response points. The suggested framework includes possibilities for incentive and penalty programmes as well. The examinees would receive bonus marks as prizes under the reward system for any additional valid points they offered. The examiner can add automated fairness to the checking process by gradually upgrading the question case-base with these additional answer-points. Unfair methods were used in the penalty system.

### 3. Methodology

There are six methodologies Namely they are Data Collection, Data Preprocessing, Feature Extraction, Model training, Model Evaluation, Model Deployment.

#### 3.1. Data Collection

Data gathering, often known as data collection, is the process of compiling information or data from multiple sources for analysis, decision-making, or research reasons. The collection of data can be done in a number of ways, including surveys, questionnaires,

interviews, observations, experiments, and more. Understanding the reason for collecting the data as well as the kind of data required is crucial for efficient data collection. As a result, the best techniques for gathering the data will be determined, and the accuracy and dependability of the data will be guaranteed. Assuring the privacy and confidentiality of the data, getting participants' informed consent, employing the right sampling strategies to assure representative samples, and making sure that the data collection process is transparent are some important factors to take into account.

### **3.2. Data Preprocessing**

In this module, the raw text data is cleaned and converted into a numerical representation that may be used as input in a machine learning model. This could entail eliminating stop words, stemming, representing the text as a bag of words, or utilizing other preprocessing methods. Pre-processing refers to the modifications done to our data before we give it to the algorithm. A technique for turning filthy data into clean data sets is data preparation. In other words, if data are gathered from various sources, they are gathered in an unprocessed way that prevents analysis.

### **3.3. Feature Extraction**

The automated feature engineering technique of feature extraction creates new variables by taking existing ones from the raw data. This step's major goal is to decrease the amount of data so that it can be used and managed for data modeling more simply. The process of extracting characteristics from data based on subjective assessments or judgements made by users or experts on humans entails employing subjective answer evaluation. In disciplines like natural language processing, sentiment analysis, and picture or video analysis, this method is frequently applied. For instance, features can be derived from text data in natural language processing based on subjective assessments

---

of its sentiment, tone, or other qualities. Asking human subject-matter experts or crowdsourcing laborers to offer

### 3.4. Model Training

When training machine learning models, subjective response evaluation entails incorporating subjective ratings or judgements offered by human experts or users. This method is frequently applied in fields including user experience design, image or video analysis, and natural language processing. For instance, in natural language processing, language models for sentiment analysis models can be trained using subjective assessments or judgements of text data. To do this, either use the ratings as features in unsupervised learning algorithms or use them as labels for supervised learning methods. Similar to text analysis, machine learning models for tasks like object recognition or image or video categorization can be trained using subjective assessments or judgements of visual content.

### 3.5. Model Evaluation

The performance of a model is evaluated using a process known as subjective answer evaluation, which depends on human judgment or opinion. It entails inviting human reviewers to offer their individualized evaluations or comments regarding the model's results or projections. In situations when the work being done by the model is subjective in nature or where there is no objective ground truth available for comparison, subjective answer evaluation might be helpful. It can also be used to evaluate how well the model performs in terms of elements like readability, fluency, and naturalness. In order to undertake subjective answer evaluation, you would normally start by choosing a group of human evaluators who are knowledgeable with the task and the relevant topic. Then, inquiries would be made of these assessors.

### 3.6. Model Deployment

This module involves deploying the trained model to predict the scores for new, unseen answers to questions. The output of this module is a score indicating the quality or relevance of each new answer.

## 4. System Architecture

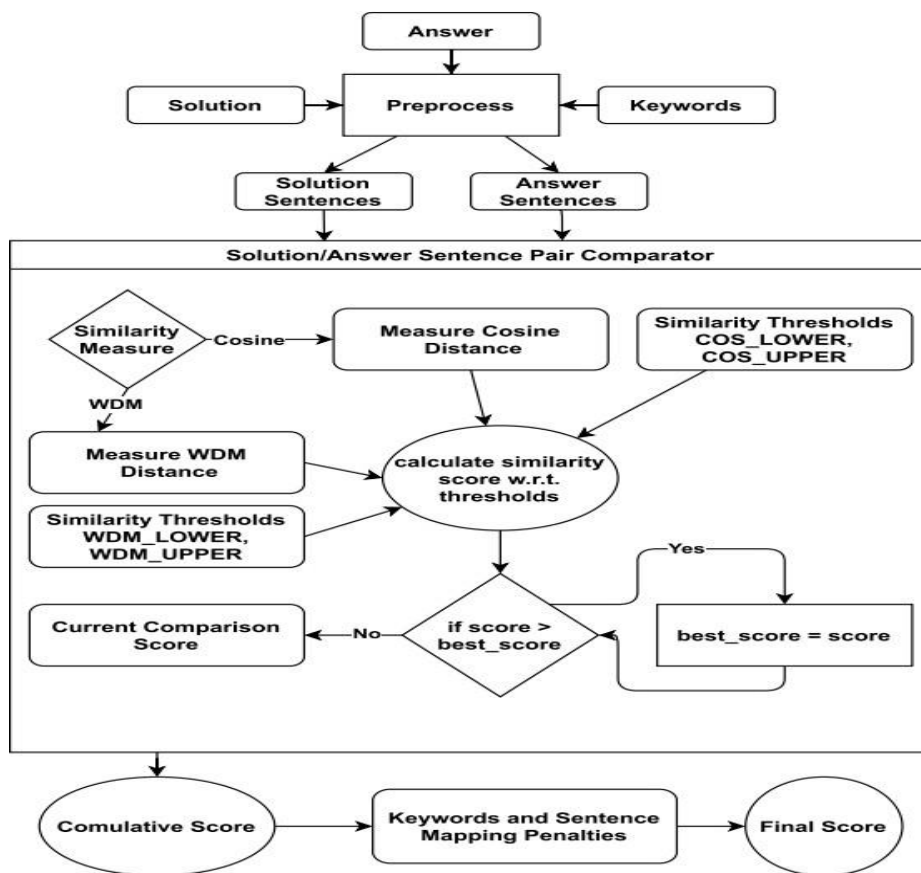


Figure 1. System Archicecture

## 5. Algorithm

### 5.1. Data Collection Algorithm

- Determine the source of data.
- Identify the type of data.
- Obtain consent from participants.
- Collect the data.
- Clean the data.
- Label the data.
- Store the data.

### 5.2. Data Preprocessing Algorithm

- Tokenization: Split the text into words or tokens.
- Stopword Removal: Remove common words that do not add much value to the analysis.
- Stemming/Lemmatization: Reduce words to their root form to handle variations in language.
- Part-of-speech tagging: Label each word with its part of speech, such as noun or verb.

### 5.3. Feature Extraction

Extract pertinent textual elements like word frequency, sentence length, or sentiment from the preprocessed text.



#### 5.4. Model Training Algorithm

- Split the dataset into training and testing sets.
- Train a machine learning model using the training set and the extracted features.
- Optimize the model by tuning hyperparameters and evaluating performance metrics.
- Evaluate the model on the testing set to measure its accuracy and generalizability.

#### 5.5. Model Evaluation Algorithm

- Determine performance parameters such F1-score, recall, and precision.
- Benchmarks and other existing models should be used to compare the model's performance.
- Interpret the results and draw conclusions about the model's efficacy.

#### 5.6. Model Deployment Algorithm

- Select a suitable deployment platform.
- Prepare the data for inference.
- Train the final model on the entire dataset or a large portion of it.
- Save the trained model in a suitable format.
- Deploy the saved model on the selected platform.
- Test the deployment thoroughly to ensure that it is functioning correctly.
- Monitor and maintain the deployment regularly.

### 6. Result and Discussion

In research paper outcome and discussion parts, subjective answer evaluation can be especially

crucial. It is common practice for authors to present their findings in these parts together with an analysis, interpretation, discussion, and conclusion. A subjective answer evaluation is required to assess the caliber of the author's investigation and conclusions because there is frequently more than one "correct" way to interpret the findings. Data must be presented in a logical and organized manner in the results section by writers. The quality of the data presentation, including the use of relevant statistical analyses and the accuracy of the labeling and organization of the data, may be evaluated subjectively using the answers. Authors must evaluate and explain their findings in the discussion section as well as talk about the implications and constraints of their research.

```

Python 3.6.5 Shell
File Edit Shell Debug Options Window Help
Detected= direct
Detected= manufacturer
count= 13

Marks for length = 5/10
Length = 455
Marks obtained for keyword:30/30
Marks obtained after parsing= 10 /10
Marks obtained out of 50 is= 45 /50
{'sales promotion activities commonly': 16.0, 'coke announced 2 liter bottles': 16.0, 'rebate refers': 4.0, 'selling prod
uct': 4.0, 'limited period': 4.0, 'excessive inventory': 4.0, 'discount refers': 4.0, 'special price': 3.75, 'original pr
ice': 3.75, 'list price': 3.75, 'price': 1.75, 'marketer': 1.0, 'increase': 1.0, 'sale': 1.0, 'time': 1.0, 'offer': 1.0,
'clear': 1.0, 'stock': 1.0, 'rs 35': 1.0, 'reduction': 1.0, 'percentage': 1.0}
data/docs/mp2.txt
Explain in detail different sales promotion techniques adopted by the marketers.
keyword in original file= ['Rebate', 'Inventory', 'Discounts', 'Refunds', 'Premiums', 'Gifts', 'Product', 'Combination',
'Quantity', 'Deals', 'Samples', 'Contests', 'Instant', 'Draws', 'Lucky', 'Draws', 'Usable', 'Benefits', 'Full ', 'Finance
', 'Packaged', 'Premium', 'Container', 'Premium']
No of keywords in original file= 24
Detected= rebate refers
Detected= selling product
Detected= excessive inventory
count= 3

Marks for length = 5/10
Length = 447
Marks obtained for keyword:0/30
Marks obtained after parsing= 10 /10
Marks obtained out of 50 is= 15 /50

```

## 7. Conclusion

For online university, school, and college level exams, the Subjective Answer Checker System Using NLP and Machine Learning (Pariksha Software) would be useful. The majority of educational institutions are still offering online exams during this COVID-19 pandemic, however these tests only include multiple-choice questions. Our software's Subjective Answer

Evaluation awards marks to subjective questions based on the length of the answer, keyword matching, grammar check, cosine similarity, and contextual resemblance to the faculty-provided Model answer and student response. Additionally, we created an algorithm to identify claims from student responses that disagree with model responses. Even while student responses do not exactly match model responses given by professors, our system is still able to evaluate responses based on context.

#### REFERENCES

- [1]. Kissan G. Gauns Dessai; Venkatesh V. Kamat; Ramrao S. Wagh et al Proposed "Effective Use of Rubrics in Computer Assisted Subjective Answer-Script Evaluation" IEEE-2020.
- [2]. Ragasudha Ragasudha; M. Saravanan et al Proposed "Secure Automatic Question Paper Generation with the Subjective Answer Evaluation System" IEEE-2022.
- [3]. Sarika Singh; Yash Shah; Yug Vajani; Surekha Dholay et al Proposed "Automated Paper Evaluation System for Subjective Handwritten Answers" IEEE-2021.
- [4]. Chhanda Roy; Chitrita Chaudhuri et al Proposed "Case Based Modeling of Answer Points to Expedite Semi-Automated Evaluation of Subjective Papers" IEEE-2019.
- [5]. Shreya Singh; Omkar Manchekar; Ambar Patwardhan; Uday Rote; Sheetal Jagtap; Hariram Chavan et al Proposed "Tool for Evaluating Subjective Answers using AI (TESA)" IEEE-2021.
- [6]. Birpal Singh J. Kapoor; Shubham M. Nagpure; Sushil S. Kolhatkar; Prajwal G. Chanore et al Proposed "An Analysis of Automated Answer Evaluation Systems based on Machine Learning" IEEE-2020.
- [7]. J. Wang and Y. Dong, "Measurement of text similarity: A survey," *Information*, vol. 11, no. 9, p. 421, Aug. 2020.
- [8]. M. Han, X. Zhang, X. Yuan, J. Jiang, W. Yun, and C. Gao, "A survey on the techniques, applications, and performance of short text semantic similarity," *Concurrency Comput., Pract. Exper.*, vol. 33, no. 5, Mar. 2021.
- [9]. M. S. M. Patil and M. S. Patil, "Evaluating Student descriptive answers using natural language processing," *Int. J. Eng. Res. Technol.*, vol. 3, no. 3, pp. 1716–1718, 2014.
- [10]. P. Patil, S. Patil, V. Miniyaar, and A. Bandal, "Subjective answer evaluation using machine learning," *Int. J. Pure Appl. Math.*, vol. 118, no. 24, pp. 1–13, 2018.
- [11]. J. Muangprathub, S. Kajornkasirat, and A. Wanichsombat, "Document plagiarism detection using a new concept similarity in formal concept analysis," *J. Appl. Math.*, vol. 2021, pp. 1–10, Mar. 2021.
- [12]. X. Hu and H. Xia, "Automated assessment system for subjective questions based on LSI," in *Proc. 3rd Int. Symp. Intell. Inf. Technol. Secur. Informat.*, Apr. 2010, pp. 250–254.
- [13]. M. Kusner, Y. Sun, N. Kolkin, and K. Weinberger, "From word embeddings to document distances," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 957–966.
- [14]. C. Xia, T. He, W. Li, Z. Qin, and Z. Zou, "Similarity analysis of law documents based on Word2vec," in *Proc. IEEE 19th Int. Conf. Softw. Qual., Rel. Secur. Companion (QRS-C)*, Jul. 2019, pp. 354–357.
- [15]. H. Mittal and M. S. Devi, "Subjective evaluation: A comparison of several statistical techniques," *Appl. Artif. Intell.*, vol. 32, no. 1, pp. 85–95, Jan. 2018.
- [16]. L. A. Cutrone and M. Chang, "Automarking: Automatic assessment of open questions," in *Proc. 10th IEEE Int. Conf. Adv. Learn. Technol.*, Sousse, Tunisia, Jul. 2010, pp. 143–147.
- [17]. G. Srivastava, P. K. R. Maddikunta, and T. R. Gadekallu, "A two-stage text feature selection algorithm for improving text classification," *Tech. Rep.*, 2021.

- 
- [18]. H. Mangassarian and H. Artail, "A general framework for subjective information extraction from unstructured English text," *Data Knowl. Eng.*, vol. 62, no. 2, pp. 352–367, Aug. 2007.
- [19]. B. Oral, E. Emekligil, S. Arslan, and G. Eryigit, "Information extraction from text intensive and visually rich banking documents," *Inf. Process. Manage.*, vol. 57, no. 6, Nov. 2020, Art. no. 102361.
- [20]. H. Khan, M. U. Asghar, M. Z. Asghar, G. Srivastava, P. K. R. Maddikunta, and T. R. Gadekallu, "Fake review classification using supervised machine learning," in *Proc. Pattern Recognit. Int. Workshops Challenges (ICPR)*. New York, NY, USA: Springer, 2021, pp. 269–288.
- [21]. S. Afzal, M. Asim, A. R. Javed, M. O. Beg, and T. Baker, "URLdeepDetect: A deep learning approach for detecting malicious URLs using semantic vector models," *J. Netw. Syst. Manage.*, vol. 29, no. 3, pp. 1–27, Mar. 2021.
- [22]. N. Madnani and A. Cahill, "Automated scoring: Beyond natural language processing," in *Proc. 27th Int. Conf. Comput. Linguistics (COLING)*, E. M. Bender, L. Derczynski, and P. Isabelle, Eds. Santa Fe, NM, USA: Association for Computational Linguistics, Aug. 2018, pp. 1099–1109.
- [23]. Z. Lin, H. Wang, and S. I. McClean, "Measuring tree similarity for natural language processing based information retrieval," in *Proc. Int. Conf. Appl. Natural Lang. Inf. Syst. (NLDB) (Lecture Notes in Computer Science)*, vol. 6177, C. J. Hopfe, Y. Rezgui, E. Métais, A. D. Preece, and H. Li, Eds. Cardiff, U.K.: Springer, 2010, pp. 13–23.