Context-Aware Scoring: A Hybrid Approach to Evaluating Descriptive Responses

***1 Lt.Dr.Babu S Associate Professor, Department of ISM, Jha Agarsen College, Chennai -600060, Tamil Nadu, India. babusanjana@gmail.com**

*** 2 Dr.T.S.Umamaheswari, Dean of Academics, Jha Agarsen College, Chennai -600060.**

***3 Dr. P. Parameswari , Principal, Palanisamy College of Arts, Perundurai, Erode, Tamil Nadu, India.**

Abstract

Assessment of academic performance is a dynamic aspect of education, and test styles can be either objective or subjective. While assessing objective responses is straightforward, assessing descriptive responses can be challenging. Evaluating objective questions is made simpler when they have a predefined set of responses. However, because they portray themselves differently, it gets difficult when students give detailed answers. Apart from the time-consuming, nature of grammar, word choice, and response presentation, the student's grades may be significantly impacted by the mood swings and stress levels of the assessors. Additionally, the automatic assessment methods for evaluating objective answers are straightforward because the evaluation is based on predetermined answers that have been saved. Descriptive responses are challenging since there are several correct solutions for the same question using various phrases. In order to address these issues with the descriptive responses, this research creates a model that uses natural language processing and machine learning approaches to automate the process, improving accuracy and efficiency. There are several processes in this automatic evaluation method for descriptive responses. The human-written text is first converted into machine-readable text using optical character recognition, followed by the text is preprocessed to make it consistent. The BERT model, which is utilized to comprehend the text's context in both left and right directions, is then implemented. This content is then subjected to the XGBoost model, which provides classification and scoring. After contrasting this assessment with the human assessment, the model's correctness is established. When compared to human review, this model's 99% accuracy rate will streamline the grading process in less time.

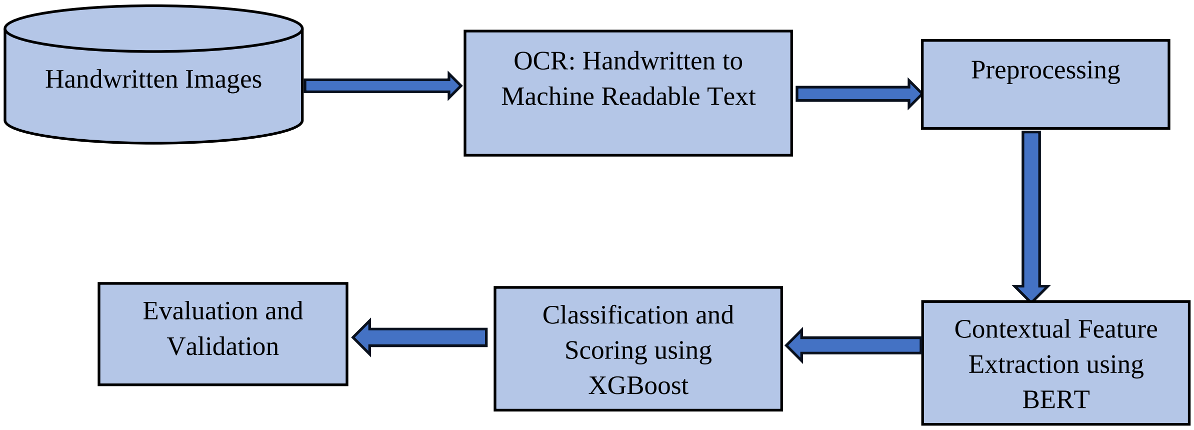
Introduction

The process of assessing students' understanding, proficiency, and knowledge in a subject or field is known as examination. It is also used to analyses student’s comprehension, critical thinking abilities, and overall learning progress. The conventional approach to analysis might be either subjective or objective. True/false questions, selecting the correct responses, and filling in the blanks with a single word and a pre-established list of solutions are the objectives. Because these predefined replies are more accurate and take less time to evaluate, both human evaluators and automatic evaluation systems find it easy to employ these kinds of answers. However, evaluating these response papers requires more time and work for descriptive responses, and the results may not be accurate. Because when the process is designed by people, the result is heavily influenced by the mood swings of each individual. When evaluating, the person may be preoccupied with personal matters, focused with another task, or not completely involved in the process; for example, they may offer different grades to different people for the same response. Consequently, the final outcome may be erroneous. Therefore, in order to overcome these limitations of the human evaluation system, we need some alternative approaches. We have developed a model for this research that will evaluate the answer scripts automatically and deliver the paper's score. The challenges highlighted by the human evaluation system are addressed by the automatic evaluation system, which is based on natural language processing (NLP).

The specified responses have led to the implementation of the automated evaluation system for objective answers; however, there are still some issues with the automated system for descriptive answers. The automated assessment system designed for descriptive responses needs to be able to comprehend and analyses the responses that students write. Additionally, this model is able to comprehend the contextual subtleties and semantic meaning of the student response, which consists of multiple words that convey the same idea. This research work develops a model that combines advanced machine learning techniques with Natural Language Processing (NLP) to overcome these difficulties in the automatic evaluation of descriptive responses. This model will improve the descriptive type assessments' correctness and efficiency. This model will also cut decrease the amount of time needed for grading and prevent human mistake in the evaluation process. There will be multiple steps involved in this automatic assessment of descriptive response models: First, the human-written text is transformed into a machine-readable format using optical character recognition. This stage is crucial for digitization because it will aid in the processing of NLP. In order to standardize the data, the digital text is then sent for preparation. Preprocessing entails a number of procedures to eliminate special characters, fix typos, eliminate white spaces, and more.

The implementation of the Bidirectional Encoder Representations from Transformers (BERT) is the subsequent stage of this automatic evaluation method. This technique uses both forward and backward directions to capture the contextual meaning of each word in the sentence. This methodology is used to determine the semantic meaning and contextual relationships of the words in the text in order to gain a deeper understanding of the students' descriptive responses. The Extreme Gradient Boosting Algorithm uses this feature as its input. This algorithm is incredibly powerful and provides increased efficiency and accuracy. The descriptive responses will be categorized by this system, which also allocates scores. When the scores produced by this automatic evaluation model are finally compared to the scores from human evaluations, the model's accuracy is 99%. The key characteristic of this model is that it avoids human errors and inaccuracies caused by the mood fluctuations of human evaluators while saving more time and being more accurate.

Methodology

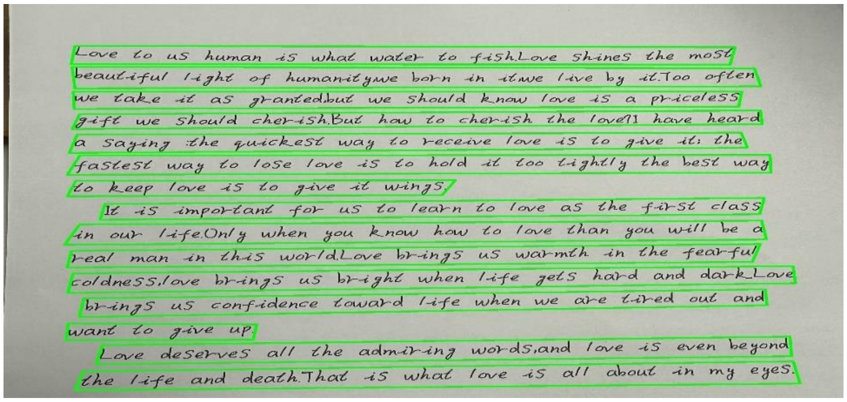

Handwritten Images
OCR: Handwritten to Machine Readable Text
Preprocessing
Contextual Feature Extraction using BERT
Classification and Scoring using XGBoost
Evaluation and Validation

The primary goal of this research work is to develop a model for the automated assessment of descriptive responses. Several steps are involved in creating the model for this purpose: Students' handwritten responses are first provided as input, and then they are converted into machine-readable text using optical character recognition (OCR). The BERT approach, which is utilized to comprehend the bidirectional contextual of every word in the sentence, is then employed. The XGBoost algorithms are then used to classify and score this vector. The resulting grading is then compared to the human evaluation grading, and the approach provides a 99% accuracy rate, replacing the human evaluation in less time. Figure 1 shows the methodology diagram for this research work

Figure 1: Methodology

Dataset Description

The dataset, which consists of handwritten images of different people, originates from Kaggle for the purpose for training this model for the automatic evaluation of descriptive answers. This dataset is primarily used to train optical character recognition software, which transforms handwritten text into machine-readable text. Various handwritten English phrase and word structures may be present in the dataset. Certain images are clearer than others, depending on their resolution, while other images contain text that overlaps. Each image has a unique writing style since it was created by a distinct person. Data must be preprocessed to transform it into a structured format if this dataset is to be sent for additional processing. Figure 2 is an example of one of these images.



Love to us human is what water to fish.Love shines the most
beautiful light of humanity.we born in it.we live by it.Too often
we take it as granted.but we should know love is a priceless
gift we should cherish.But how to cherish the love?I have heard
a saying :the quickest way to receive love is to give it; the
fastest way to lose love is to hold it too tightly the best way
to keep love is to give it wings.

It is important for us to learn to love as the first class
in our life.Only when you know how to love than you will be a
real man in this world.Love brings us warmth in the fearful
coldness.love brings us bright when life gets hard and dark.Love
brings us confidence toward life when we are tired out and
want to give up.

Love deserves all the admiring words,and love is even beyond
the life and death.That is what love is all about in my eyes.

Figure 2. Sample Dataset

Preprocessing

The image collection contains low-quality text, and if it is processed without preprocessing, the desired results will not be obtained. Therefore, various preprocessing processes must be performed on this data in order to improve the quality and accuracy of the text that is present in the dataset. Some of the preprocessing steps used in this research work to enhance the dataset's quality are listed below.

.

1) **Noise Removal**: The OCR blocks the characters because of several noises in the image documents, including stains, spots, and other symbols. Consequently, this noise must be removed from this dataset. The median filtering method is used for this purpose in this research. The median value of the neighbouring pixels is used by this median filter to substitute these values.
2) **Thresholding/Binarization:** When the images are coloured, OCR becomes extremely difficult to convert the text into a machine-readable format. Therefore, this greyscale color images must be converted into black and white binary images. This makes the text more contrast than the backgrounds which is easy for the OCR to identify contrasted text and transform them.
3) **DE skewing**: When the text of the document is aligned horizontally, the OCR engines are able to accurately recognize and process the characters. This necessitates the right alignments, which call for the papers to be oriented correctly. As a result, the document is verified for correct alignment in this stage; if not, it will be rotated. This will enable OCR to perform at its highest level and deliver more precise outcomes.

These crucial preprocessing processes are part of this research work before the document is sent to optical character recognition (OCR), which improves OCR's efficiency and accuracy in turning handwritten text from humans into machine-readable text.

Optical Character Recognition

A technique called optical character recognition is used to extract data from image files and transform them into machine-readable files. This file could contain an image, printed text, handwritten text, bills of materials, and receipts. This OCR software will transform everything into text that can be read by machines. The image file will be provided as input first, and once it has been scanned by the OCR system, the text can be extracted into machine-readable text. The students' descriptive answer sheets are provided as an image file for this research project's OCR input, which extracts machine-readable text from the image. The OCR itself contains several steps to make this process. Figure 3 shows the steps involved in OCR.

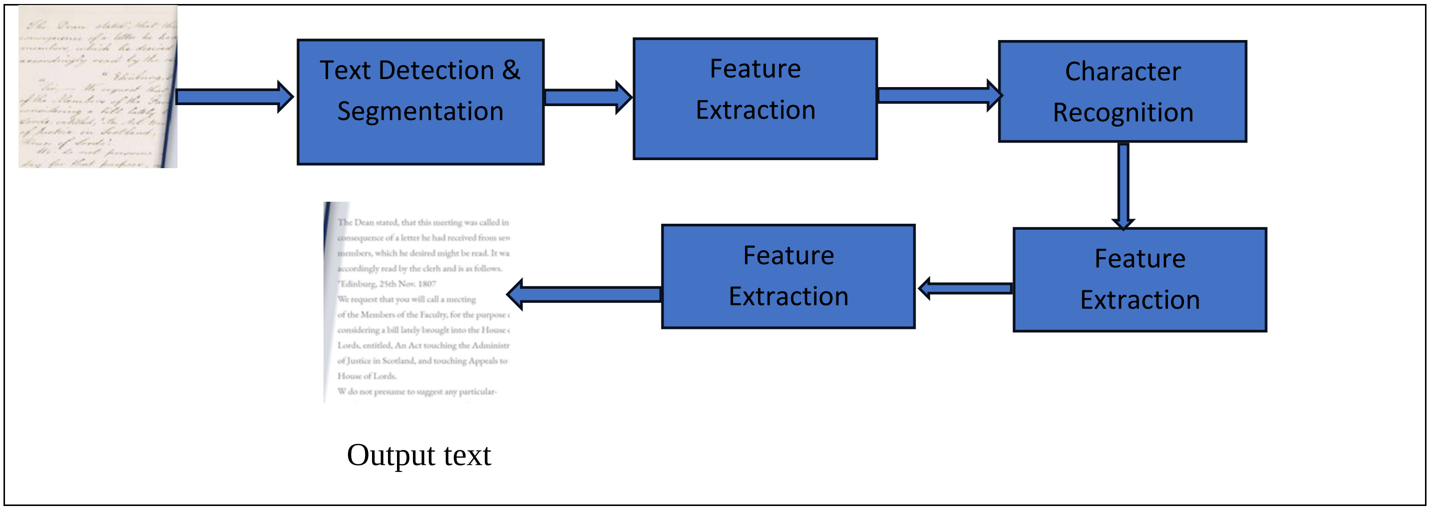


Figure 3. OCR

Text Detection & Segmentation: The OCR first segments the text, or identifies the sections of the text that are displayed in the document, after receiving the preprocessed handwriting images. This will make the conversion process easier by classifying the text sections into line, word, and character segments. Finding the space between the vertical lines is necessary for line segmentation in order to divide the text line. The method of determining the distance between each word in a text is known as word segmentation. Lastly, the process of determining the distance between each character in the text is known as character segmentation. The crucial step in the process of identifying each textual component is segmentation.

Feature Extraction: An image of text will appear to the computer as a collection of pixels. Important features such as strokes, forms, sizes, and patterns are extracted from a scanned image of text using this OCR stage. Every character in the dataset is analyzed and categorized using these attributes.

Character Recognition: This essential OCR stage compares the retrieved features with a predetermined set of patterns to transform each character into machine-readable text. In order to identify the characters and match them with specified features that provide more accuracy, template matching or certain machine learning techniques are utilized in this stage.

Post Processing: Spelling errors and character misinterpretations must be present in the image text once it has been transformed into machine-readable text. Consequently, the postprocessing stage entails fixing these mistakes through the use of spell checking, which compares the identified characters with dictionary words and fixes any inaccuracies utilizing this OCR step.

These procedures will result in machine-readable text via optical character recognition, which is helpful for this descriptive answer evaluation process. This machine-readable text will be sent to BERT, which will determine each word's contextual meaning.

Proposed Method

In the present research, the BERT model and XGBoost algorithm are used to automatically evaluate descriptive responses. One deep learning model used for natural language processing is Transformer's Bidirectional Encoders Representations. This approach analyses in both directions to provide a good understanding of the contextual meaning of each word in the text. The best machine learning algorithm for classification and regression tasks is called XGBoost. This algorithm is used in this research to score and classify the descriptive responses. The BERT model for determining the semantic meaning of sentences involves the following steps.

Preprocessing: The first stage in this BERT model is to preprocess and segment each word so that they can be divided into various parts. The tokenizer creates this process by breaking each word down into subworlds and adding unique tokens to indicate the beginning and finish of sentences.

Encoding: To produce the encoded representation, the input text is routed through several transformer layers following preparation. Each transformer in this model has a feed-forward neural network and a self-attention mechanism.

Fine-Tuning: This layer fine-tunes the pretrained BERT model for the particular task in this research, which is the description answer correction. This is accomplished by overlaying the encoded representation with the task-specific layer.

Training: To modify the parameters of the neural network and the self-attention process, the refined model is trained on a labelled dataset.

Evaluation: To determine the model's performance, the provided dataset is analyzed again with the original evaluations.

The output of this BERT model is a series of hidden state vectors, each of which represents the contextual understanding of a corresponding token with the dataset. The XGBoost classification algorithm uses these vectors as input for scoring and classification. Regression and classification are two applications for the potent machine learning technique known as eXtreme Gradient Boosting. This approach is widely used since it is more accurate, efficient, and capable of handling big datasets. This algorithm, which is an ensemble approach, combines more weak learners to produce a strong learning, namely the decision tree collection. In this case, each model iteratively fixes the mistakes of earlier models to generate outcomes with a high degree of accuracy. The following are the steps in this algorithm.

Input Feature: This XGBoost algorithm uses the contextual vectors produced by the BERT model as its input for scoring and assessment.

Model Initialization: The model is initiated by a base prediction for all the data points.

Residual Calculations: The error rate will be the difference between the original values, or ratings provided by the human evaluators, and the anticipated values.

Regularization and pruning: To avoid the overfitting issue, the superfluous branches are cut depending on the lowest information gain threshold value.

Weighted score aggregation: The prediction made by each tree is used to determine its contribution to the learning rate. The ultimate score is the total of all the trees' predictions.

When compared to the assessment of human evaluators, the automatic scoring is assigned with high accuracy to the student's descriptive responses once this model has been applied.

Result and Discussion

The primary objective of this research work is to develop a model for the automated assessment of descriptive responses. This is accomplished by extracting the handwritten images from the Kaggle dataset, which is then used to train the model. There are some sounds in the dataset that was extracted from Kaggle. The output contains some error if the dataset is provided to the model for the process. Thus, the image dataset is preprocessed to provide formatted structured data after it has been extracted. A number of procedures, including noise reduction, binarization, and de-skewing, are used in the preprocessing stage. The OCR then uses this prepared data as input to turn the handwritten text into a machine-readable format for additional processing. To turn this into machine-readable text, OCR itself goes through a number of processes. The embedding, a collection of vectors that represent each word in the text, is the result of OCR. The BERT model is then applied to determine the semantic meaning of each sentence, which facilitates the evaluation process. The descriptive responses are classified and scored using the Boost classification algorithm, which assigns a score between 1 and 10. The model's performance is then assessed by comparing the scores provided by the human assessors.

Both the automatic system and the human evaluators receive the same student's descriptive responses during the evaluation process. Each question will be given a score between 1 and 10 by the automated system and human assessors. The model provided by the human assessors and the marks determined by them are displayed in table 1.

Table 1. comparison of Scores

S.No	**Marks Given by manual evaluation**	**Marks given by the system**
1	9	9
2	6	6
3	8	8
4	9	9
5	6	6
6	7	6
7	8	8
8	8	8
9	9	9
10	7	7

The automatic evaluation model provides identical scores for every question and has 99% accuracy when compared to human evaluation. Figure 1 displays the comparison's graphical depiction.

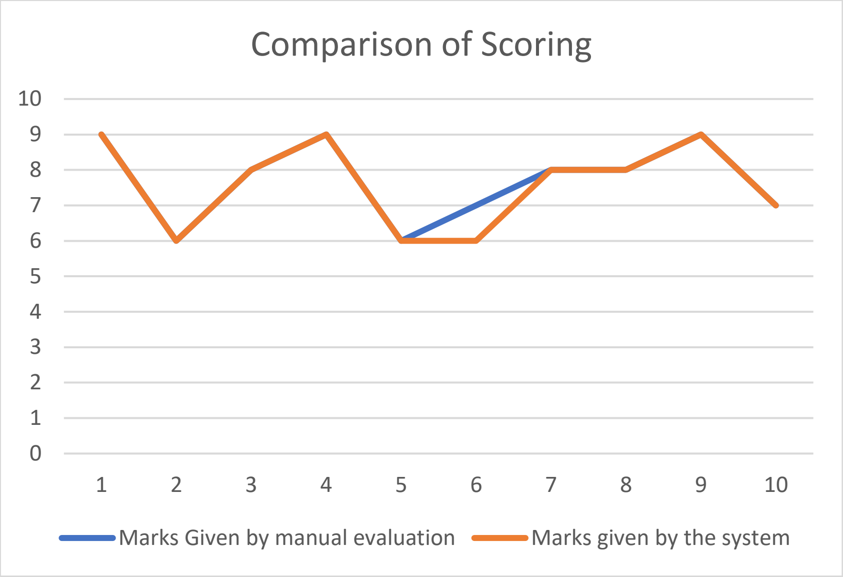


Figure 4. Comparison of scores

Conclusion

The model was developed for the high-accuracy, time-saving automatic evaluation of descriptive responses. Additionally, this program will steer clear of errors that arise from human evaluators' fluctuating moods and stress levels. The implementation of the BERT model strengthens this model by clearly understanding the semantic meaning of each word, making it simple to grade. The XGBoost algorithm, a very powerful classifier, is then used for the categorization and scoring of the responses. Using decision trees, this classifier will fix the mistakes in every iteration. The final product is the sum of the results from all of the weak learners, which produces a strong learning that yields more precise results. The marks are assigned from 1 to 10 using this categorization system. This scoring is then compared to the scoring from the human evaluation. It is evident from the comparison that the automatic model will have an accuracy of 99%. This methodology was primarily developed to save time and prevent human error. More sophisticated pretrained models will be used in the future and applied to real-time data for the benefit of students.

References

[1] Xipung et.al., "A feature learning approach based on XGBoost for driving assessment and risk prediction," *Accid. Anal. Prev.*, vol. 129, pp. 170-179, 2019.

[2] T. Kasahara and D. Kawahara, "Exploring automatic evaluation methods based on a decoder-based large language model for text generation," *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics: Student Research Workshop*, pp. 24–31, Nov. 1–4, 2023.

[3] V. Kumari, P. Godbole, and Y. Sharma, "Automatic subjective answer evaluation," *12th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2023)*, 2023.

[4] S. Rajini, C. Chamarathi Sai Chethan, R. Balaji, and A. Sujith Kumar, "Automatic answer evaluation for descriptive answers," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 3, no. 5, May 2021.

[5] R. Manjunath and S. K. Guruswamy, "Automation of answer scripts evaluation - A review," *ResearchGate*, 2021.

[6] S. Joshi, A. Bachkar, O. Awaje, R. Bhoir, and K. Urane, "Automated answer sheet evaluation using BERT," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 10, no. 3, May-June 2024.

[7] S. J. Shaikh, P. S. Patil, J. A. Pardhe, S. V. Marathe, and S. P. Patil, "Automated descriptive answer evaluation system using machine learning," *International Journal of Advanced Research in Innovative Ideas in Engineering*, vol. 7, no. 3, 2021

[8] S. Santhanavijayan, "Automatic generation of multiple-choice questions for e-assessment," *International Journal of Signal and Imaging Systems Engineering*, vol. 10, nos. 1/2, 2017.

[9] S. Mahmud and A. Alam, "Automatic multiple choice question evaluation using Tesseract OCR and YOLOv8," *2024 IEEE Conference on Artificial Intelligence (CAI)*, 2024.

[10] S. S. Kare, P. Karanjikar, S. Shabir, G. Deshmukh, and S. Lokhande, "Subjective answer evaluation using BERT," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 6, no. 4, Apr. 2024.