Evaluating descriptive answer sheets using AI

Aadhithya sunil

(25-06-2024)

ABSTRACT

The rise in AI tools and automation is largely due to the need to resources on tasks consuming huge amount of time and workforce. One such area in need of automation is the evaluation of descriptive answer in the response sheet of students.

India being the most populus country in the world with the largest student community it has always been a herculean task to manually evaluate each response given by all students in different examinations. While AI struggles with high-level professional course assessments, it can effectively evaluate theoretical answers. This software can reduce the workload on educators and ease the evaluation process, which can provide more accurate and adequate results compared to human evaluation.

Problem Statement:

The current manual evaluation process for answer sheets in educational institutions can be improved with several significant challenges. Evaluating a large number of answer sheets is highly time-consuming, which can lead to delays in delivering results to students. Moreover, the consistency and fairness of grading are compromised by human factors. Even experienced evaluators can suffer from fatigue and overload when faced with large volumes of work, leading to unintentional errors and variability in grading.

Additionally, different educators often have varying evaluation standards and subjective biases, which can result in inconsistent grading practices. This can create unfairness, where students do not receive grades that accurately reflect their performance. As these grades can influence academic progression and other future opportunities, we should take effective measure to change it.

In the context of the advancements in Artificial Intelligence (AI), there is a pressing need to explore and implement AI-driven solutions to address these challenges. AI has the potential to revolutionize the evaluation process by providing faster, more accurate, and consistent grading. Automated grading systems, powered by AI, can handle both objective and subjective questions, ensuring a fair grading to all. This can significantly reduce the workload on educators and minimize human error.

Exploring the current capabilities and possibilities of such AI-driven evaluation software in the market, why and where should we use this? how can we implement this?

1. Introduction

India's educational sector is one of the largest in the world, with over 10 million students graduating each year from colleges and many more from schools. Providing quality education to such a vast population is never a formidable task, which is why the lot of students in current generation opt to pursue their education abroad instead of staying in India.

While India has numerous prestigious institutions that offer high-quality education, due to high competition for admission many deserving candidates are left out. However, the focus of this discussion is not the quality of education but the quality of evaluation. After various exams, terms like 'scam' and 'undeserving' frequently surface in news. There are instances where human error leading to some students receiving much lower grades than expected, only to have their scores corrected upon re-checking as per his/her request. Although AI is not prone to make mistakes, the error rate of AI evaluation is significantly lower compared to manual error in evaluation.

1.1 Methods of current manual evaluation

Let us start with the school level, for school level unless it is boards examinations most of the evaluation process is straight-forward and is done by the respective faculties teaching the subjects. Unlike board exams there is no anonymity in the correction of answer sheets which may lead to some favouring but other than that there is not much role of improvement for AI evaluation. While these exams are mainly for self-assessment and carry less weight compared to board exams, AI evaluation can enhance the process by providing feedback to teachers on areas where students commonly underperform, allowing for improving in the future.

Coming to board examinations, properly evaluating answer sheets is a significant challenge. Board exams play a crucial role in students' futures, and schools aim to maximize their success rates. Although various boards in India have different evaluation approaches, most government and private boards set up evaluation camps in schools. This process can be exhausting for educators, leading to errors due to pressure and heavy workloads. Students can request re-evaluation if they believe their grades received are lower than deserved. Incorporating AI at this stage can not only improve accuracy, this will eliminate the lengthy process such as double-evaluation and re-evaluation, and ensure that each student receives the grade they actually deserve.

Now focussing in the undergraduate and postgraduate the education there are approach done by various colleges may be different the root problems in evaluation remains the same. The rate of students applying for re-evaluation at this level is also very high despite the high amount of money is charged for it by the institutes as mentioned above this can be avoided using an AI evaluation tool. The grades a student get in his/her graduate degree has significant role in their future by letting different type of evaluation method to evaluate each student during their degree has always caused inequality among students graduating from different universities

Other than the sections mentioned above this technology can be also used in any descriptive response analysis like the students are being evaluated of their written ability in exams like CAT, UPSC and similar exams the evaluation of these long essays may be very time consuming and tiring this can be avoided using out evaluation method and the candidate can also self-analyse them using this.

1.2 **Necessity of the change**

There are discussions going on about AI taking over the world often overlooks the fact that new technologies are not meant to replace jobs or automate everything completely. Instead, AI tools should be used effectively to enhance accuracy and save time. At this pace, we might soon see software accomplishing tasks we never thought possible. This evolution simplifies our lives by reducing workload and providing efficient outputs.

Since this software is not widely used globally, questions will arise when it is introduced in India, given its not tested successfully in other developed countries. However, the need for AI evaluation is particularly crucial in India due to its vast student population. This AI evaluation software can have a more significant impact in India than in any other country. Despite being well-developed in the scientific field, India often implements only those technologies that have been successful elsewhere, rather than experimenting with innovations that could be highly effective domestically.

1.3 **Challenges in execution**

In the era of ‘Chat GPT’ where the answers for any descriptive exams can be found there it is not a big challenge to execute this software and even if there is chance of error at the start eventually AI will learn as it is trained by this several data and we can also make our model way better after each trial. Research suggests that AI can evaluate essay and long answers with 91% accuracy compared to the 72% which is the accuracy of manual evaluation.

We can use various tools which are already available like handwritten note to text translator to take the input of our customers and through natural language processing algorithms we can train the AI to understand the response by which AI can understand what the user was implementing by their response.

2. Market research

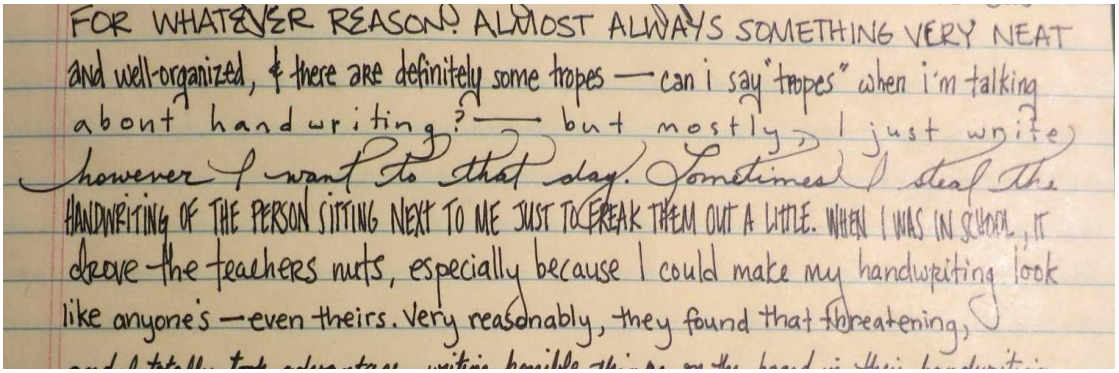
This product has been on the market for some time. While some businesses have already established a model comparable to this one, they are only offering only limited services to particular educational institutions. This product's market extends way beyond its current reach. This programme can be used to grade any test that requests descriptive responses. Because it is a technological innovation that is not extensively used worldwide, the public perception of using such product will benefit not only the service providers but also our customers. Consider a private school that defies convention by evaluating answer sheets using technology. This will generate a lot of buzz in our society and enhance the institution's reputation, same with the coaching institutions. We can also provide a detailed overview of how the student performed in different areas, which will make the institutes who use or product stand out from their competitors.

The product can be gradually introduced to private schools by testing the response of the product in lower grades we can slowly increase the scale of our product. We can expand the range of our product service to even professional degrees like the medical and economic science field. But this might not be practical at the current stage but at the rate technology is changing the market forecast for this product will only exponentially grow in the coming years. Few years back the pen and paper method were the only way to conduct tests now we can see a smooth transition from this traditional pen and paper method to computer protected tests in certain fields. These changes can give a huge boost in market for our product and the potential market reach of the product in a futuristic examination scenario is way bigger than any AI model can forecast.

3. External Search

Handwritten statements to text converter can be created by training neural network with the text images such data sets are available in open source at Iam handwriting which consist of:

657 writers contributed samples of their handwriting
1'539 pages of scanned text
5'685 isolated and labeled sentences
13'353 isolated and labeled text lines
115'320 isolated and labeled words



By using Optical Character Recognition (OCR) we can scan each character one by one and convert it into the desired word.

Text can be converted unless they are written in a confusing way which is unrecognisable to improve the accuracy of our model we can train the handwriting of our students.

Prototype of the model can be found in GitHub.

https://github.com/im-minion/Subjective-Answer-Evaluation

https://github.com/KashyapBhat/Paper-Evaluation-using-AI-ML

4. Bench Marking

Although the product is not widely used in the industry, we can find some similar service provided by some platforms over the internet. Some mini prototype model of the which some researches and students have done as a part of their project.

Since these products have not reached our target audience they could not be labelled as the competitor benchmark itself in the tech industry.

Some similar products:

https://www.eklavvya.com/auto-descriptive-answer-evaluation-demo/

https://evaluateai.vercel.app/

5. Business Model

The services can be provided to an educational institution or a single individual the monetization can be done through measuring the amount of answer scripts to be evaluated we can use a charge a certain amount per answer script and we can reduce that amount per answer scripts as we get high amount of answer scripts to evaluate. The rate per evaluation can be fixed by considering the amount of money and time the institutes usually used to evaluate the answer scripts. So that we can propose a rate beneficial to our customers too.

Coming to revenue from individuals, there is not much difference in the manual and our evaluation process. But our product still can give valuable insights in the areas the candidate should focus to improve themselves. We can use credit-based system for individuals in our online site for this, so charging certain credits or token for evaluating number of answer sheets will give us a good revenue, free trial up to few scanned copies as a demo will also attract the audience more.

Additionally we can provide paid test series with popular written ability test examinations to candidates and giving them a quick result and overall analysis of their response.

6. Execution of the concept

Execution part of the model can be highly complex since expect it to give us highly accurate output with minimal error. To achieve this, training our model with the appropriate data is highly important. With respect to the evaluation policy of customers they can train the model with the necessary keywords required. The model will now summarize the student's response and check the similarity between the given response and the keywords and will give a desired mark as output.

The methods used in this model are:

1.**Cosine similarity :** A metric used to measure how similar two vectors are by calculating the cosine of the angle between them. In text analysis, it is often used to measure document similarity by representing documents as vectors of word frequencies or TF-IDF scores. Give the degree of similarity between both texts.

2.FuzzyWuzzy : Python library for string matching that uses the Levenshtein Distance to calculate the differences between sequences. It provides tools for fuzzy matching, scoring how similar two strings are, and identifying approximate matches, which is useful for tasks like record linkage, deduplication, and data cleaning.

3.Jaccard Similarity : A statistic used for gauging the similarity and diversity of sample sets. It is defined as the size of the intersection divided by the size of the union of two sets. In text analysis, it can be used to compare the similarity between two documents based on the words they contain. Use mathematical methods to find the union of both sentences.

4.BERT : A state-of-the-art machine learning model for natural language understanding. Developed by Google, BERT is designed to understand the context of a word in search queries by considering the words that come before and after it, thus providing more accurate results in tasks such as question answering, text classification, and language inference. Understands the oversll structure of the sentence.

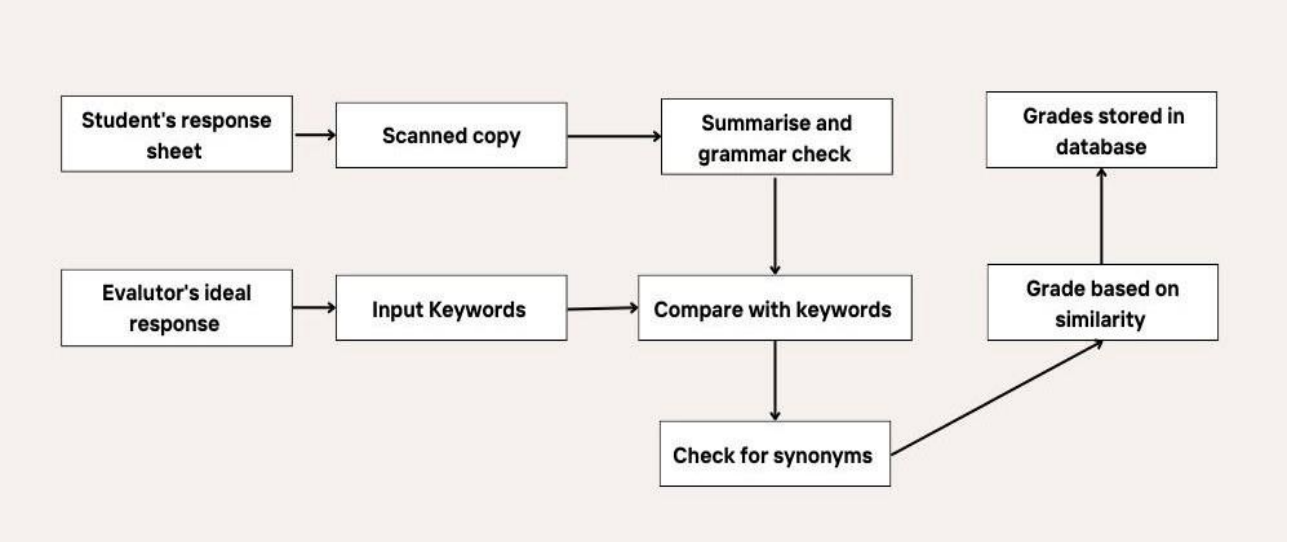
5.Grammar check : A feature or tool in natural language processing that evaluates the grammatical correctness of text. It identifies and suggests corrections for grammatical errors, such as punctuation mistakes, verb tense inconsistencies, and improper syntax, enhancing the readability and correctness of written language. Appropriate marks should be reduced for weak grammar.

6.PyDictionary : A Python library that serves as a wrapper for various dictionary APIs to fetch definitions, synonyms, antonyms, and translations of words. It allows users to easily integrate dictionary functionality into their Python applications. Helps in finding similar words used other than the keywords.

7. **Final product prototype**

Input : The keywords given by the evaluators and response submitted by the students are two inputs required for the model. Model finds similarity in the both and give appropriate marks for each response.

Output : Output will be the total marks obtained by a student it can be stored in a data base with other relevant data like marks obtained for each questions, areas at which he/she performed weak etc.



****made using canva*

Here input keyword can also be an evaluator who can check foe example 10% of the answer sheets, so that AI can recognise the evaluation pattern of the evaluator and then check the remaining answer sheets with the same pattern.

8. Conclusion

One of the key objectives of the new age tech industry is to automate repetitive and time-consuming tasks. Artificial intelligence (AI) is being used extensively in many industries, including manufacturing, finance, medical, and education. But despite the recent surge of AI, education remains largely unchanged. We still follow the established methods for teaching and evaluating.

The inclusion of this model into the school level may indulge high curiosity about the working of AI in the young minds as they are directly watching and being affected by the use of this AI evaluator.

A few years ago, the functioning of the model was highly doubtful. However, we are nearly done to developing and implementing an artificial intelligence tool to replace manual evaluation without any error. Although there will initially be resistance to its use, we cannot advance in technology if we are unwilling to let go of some of our traditional habits in order to make room for emerging technologies. I find it impossible to think that in five years the answer sheets will be evaluated using the same procedure as it is now. **This is the future.**

References

- https://joirem.com/wp-content/uploads/journal/published_paper/volume-10/issue-4/J_0r3x8Lg0.pdf
- https://www.eklavvya.com/blog/descriptive-answers-evaluation/
- https://github.com/KashyapBhat/Paper-Evaluation-using-AI-ML
- https://link.springer.com/article/10.1007/s11277-019-06913-x