SUBJECTIVE ANSWER EVALUATION SYSTEM

Presented by Prashu Vimal

Problem Statement

Subjective paper evaluation is a tricky and tiresome task to do by manual labor. Insufficient understanding and acceptance of data are crucial challenges while analyzing subjective papers using Artificial Intelligence (AI). Several attempts have been made to score students' answers using computer science. However, most of the work uses traditional counts or specific words to achieve this task. Furthermore, there is a lack of curated data sets as well.

Solution Proposed

The solution proposed will utilizes various machine learning, natural language processing techniques, and tools such as Keyword/TF IDF,Semantic Embeddings, Concept Mapping ,Concept Weighting , Transformer Fine-tuning to evaluate descriptive answers. We use different evaluation measures such as TF-IDF Cosine similarity+Keyword Coverage , BERTScore + Concept Overlap, Graph Similarity +,MAE+RMSE etc. to evaluate the performance of various models performance.

TECH STACK

Frontend Technologies

- React.js
- HTML
- CSS
- Axios

OCR Technologies

- Tesseract OCR
- pytesseract
- pdf2image
- Pillow (PIL)

Backend Technologies

- Python
- Flask
- Flask-CORS

NLP & AI Technologies

- SBERT (Sentence-BERT)
- Transformer Architecture
- all-MiniLM-L6-v2 model

Similarity & Scoring Techniques

- Techniques Used
- Cosine Similarity
- TF-IDF (supportive role)

Model Features

- Semantic evaluation using Transformer-based SBERT.
- Human-like subjective answer assessment.
- Paraphrase-aware and context-sensitive grading.
- OCR-enabled input for scanned PDFs and images.
- Knowledge-graph assisted concept validation.
- Hybrid evaluation combining SBERT + TF-IDF + KG.
- Fast, lightweight, and real-time performance.
- Automated scoring with grade and feedback generation.
- Modular and scalable system architecture.

Dataset Collected

03


```
"contexts": [
  {
    "Context_id": "recursion_1",
    "Basic_Level": {
      "Questions": [
        {
          "id": "1",
          "question": "What is recursion?",
          "max-marks": 2
        },
        {
          "id": "1_1",
          "question": "Define recursion in programming.",
          "max-marks": 3
        },
        {
          "id": "1_2",
          "question": "What does it mean when a function is called \"recursive\"?",
          "max-marks": 2
        },
        {
          "id": "2",
          "question": "What is the base case in a recursive function?",
          "max-marks": 2
        },
        {
          "id": "2_1",
          "question": "Why is a base case important in recursion?",
```

04


```
{
  "answer_id": "4_1_1",
  "context_id": "sorting_1",
  "question_id": "4_1",
  "correct_answer": "Recursive sorting algorithms use function calls such as Quick Sort, while non-recursive ones include Selection Sort and
  "correct_score": 2,
  "partially_correct_answer": "Some sorting algorithms use recursion.",
  "partially_correct_score": 1,
  "weak_answer": "Recursion means looping.",
  "weak_score": 0
},
{
  "answer_id": "5_1",
  "context_id": "sorting_1",
  "question_id": "5",
  "correct_answer": "A sorting algorithm is stable if elements with equal keys maintain their relative order after sorting.",
  "correct_score": 3,
  "partially_correct_answer": "Stable sorting keeps equal elements together.",
  "partially_correct_score": 2,
  "weak_answer": "Stable sorting is fast.",
```

```
[
  {
    "context_id": "recursion_1",
    "paragraph": "Any function which calls itself is called recursive. A recursive method solves a problem by calling a copy of itself to work on a
  }
,
  {
    "context_id": "sorting_1",
    "paragraph": "10.1 What is Sorting?\nSorting is an algorithm that arranges the elements of a list in a certain order (either ascending or descen
  }
,
{
  "context_id": "sorting_3",
  "paragraph": "10.8 Analysis of Insertion Sort\nWorst Case Analysis\nWorst case occurs when for every i the inner loop has to move all elements A[1
}]
```

Other Evaluation systems till now

Paper / System	Year	Approach Used	Dataset	Evaluation Metric	Reported Accuracy / Score
Burrows et al. – Survey of ASAG	2015	NLP + ML (TF-IDF, cosine, syntactic	Multiple (ASAP, C-Rater)	Accuracy, QWK	60% – 75%
AutoSAS	2020	Ensemble ML + NLP features	ASAP-SAS	Quadratic Weighted Kappa	**0.73 – 0.78**
Ans2Vec	2019	Sentence Embeddings (Skip-	Proprietary	Accuracy	~70%
Traditional TF-IDF + Cosine Similarity	2016–2019	Bag-of-Words similarity	ASAP-SAS	Accuracy	**55% – 65%**
WordNet-based Semantic	2018	Ontology + semantic overlap	Custom academic datasets	Accuracy	~68%
SBERT-based Grading	2021	Sentence-BERT embeddings	ASAP-SAS	Pearson / Accuracy	**78% – 85%**
LLM-based Graders (GPT-like)	2023–2024	Large Language Models + Rubrics	Multiple	Human Agreement	**85% – 92%**
OCR + NLP Grading System	2024	OCR + TF-IDF + rules	University exams	Accuracy	65% – 72%
Rule-Based Keyword Matching	2015	Keywords + heuristics	Custom	Accuracy	50% – 60%

Subjective Answer Evaluation System

Choose File No file chosen define recursion

iteration

Evaluate

Score: 0

Grade: Incorrect

Feedback: Answer does not cover required concepts.

07

Subjective Answer Evaluation System

Choose File No file chosen define sorting

sorting is arranging data in ascending and descending order

Evaluate

Score: 1

Grade: Partially Correct

Feedback: Definition is incomplete.

Student Answer Evaluation System

Evaluate Answer

hii

Choose File Aryan.pdf

Extract (OCR)

```
ARYAN RAJPUT

aryanrajput039@gmail.com | +91 9717075101 | New Delhi, India
LeetCode | Portfolio | GitHub | LinkedIn

EDUCATION
Banaras Hindu University 9.23 CGPA
Master of Computer Applications (MCA) 2026
PROJECTS
```

Evaluate

Evaluation Result

Question:

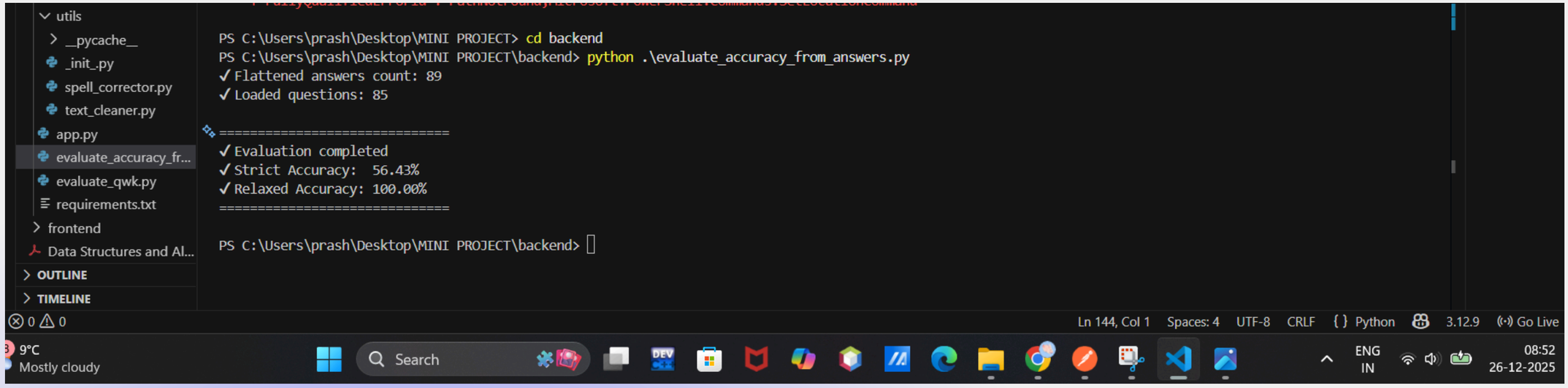
Score: 0

Grade: Manual Review Recommended

Feedback: Question not found in dataset

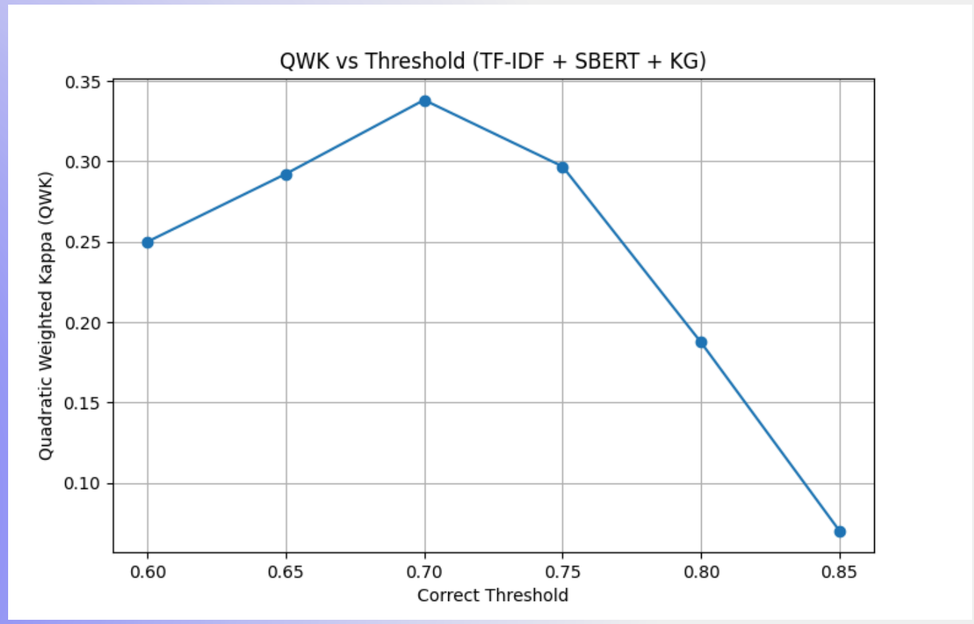
10

Result Analysis



Strict Accuracy: |Human-System|=0

Relaxed Accuracy:|Human-System|<=1



Thank You