

Izmir University of Economics
Department of Software Engineering

SE 420 Project II Report

Defne Yılmaz - 20220601078
Hüseyin Yontar - 20210601067

1 Project Definition

The objective of the project is to evaluate the students' answers to the short questions by comparing the student answers to the answers in answer pool (Short Answer Automatic Scoring System, SAASS).

There are two files provided; one containing answer key for the given questions and which answer corresponds to which point, the other one containing students' answer to the given questions.

In order to evaluate the answers of the students the program generates embeddings for answers by using the Sentence Transformers model (distiluse-base-multilingual-cased-v1) which is suitable for embedding Turkish sentences. After embedding all answers, the program calculates the similarities based on cosine distance between students' answers and all answers to a specific question. Then, the program checks which answer is the most similar to the student's answer and assigns a point between 1-3 related with the given answer. Based on the similarities, the system assigns scores, computes the total score for each student, and outputs the results to an Excel file (results.xlsx).

An example for asked questions and its answers & corresponding scores:

Question- ID	Question	SCORE-1	SCORE-2	SCORE-3
Q1	İklim nedir?	Hava durumu, Havanın bir önceki yıla göre aynı olmaması, Hava durumunun bir bölgede farklı zamanlarda aynı olmaması	Bir bölgedeki hava olayları, Bir bölgedeki hava olaylarının ortalaması, Uzun süreli hava olayları, Uzun süreli hava durumu	Belli bir bölgedeki hava olaylarının uzun süreli ortalaması, Belli bir bölgede gözlemlenen ortalama hava olayları, Bir bölgeye özgü uzun süreli gözlemlenen hava olaylarının ortalaması
Q3	Albedo etkisi nedir ve küresel sıcaklıklara nasıl etki eder açıklayınız.	yüzeye düşen güneş ışınları sıcaklığı artırır, küresel sıcaklığı artırır, küresel ısınmayı engeller	Yüzeyin ışığı yansıtması, Yansıtma kapasitesi, Soğumaya neden olur	Herhangi bir yüzeyin üzerine düşen güneş ışığını yansıtma kapasitesine albedo denir, Albedo yüzeyin yansıtma gücü ya da yansıtma yeteneği olarak da ifade edilir. Yüzey ne kadar çok yansıtırsa o kadar soğur.

An example for a student's answer to an asked question:

	Q1-answer	Q2-answer	Q3-answer	Q4-answer
ST1	Belirli bir bölgede uzun bir sürede gözlemlenen ortalama hava koşullarının belirli döngüsel olaylarını ifade eden kavramdır.	İklim değişikliği ile dünya genelinde ortalama sıcaklıklar artar, atmosferdeki su buharı miktarından kaynaklı yağışlar artar gibi etkileri bulunmaktadır.	Albedo etkisi bir yüzeyin üzerine düşen güneş ışığını yansıtma kapasitesidir. Albedo etkisinin yüksek olması küresel sıcaklıkların düşük, düşük olması küresel sıcaklıkların yüksek olmasına neden olur.	Antropojenik kökenli karbondioksitin atmosferden emilmesi, okyanus suyunun pH seviyelerinde düşüşe neden olur. Bu da okyanus asidifikasyonu olarak ifade edilir.

A sample result for the answers of students for question 1 after running the program:

Student ID	Student Total Score	Q1 - Predicted Score	Q1 - Matching Answer	Q1 - Cosine Similarity
ST1	11	3	Bir bölgeye özgü uzun süreli gözlemlenen hava olaylarının ortalaması	0.73571765
ST2	9	1	Hava durumu	0.3062009
ST3	8	1	Hava durumu	0.3062009
ST4	10	3	Belli bir bölgedeki hava olaylarının uzun süreli ortalaması	0.7808541
ST5	7	2	Bir bölgedeki hava olayları	0.5443083

2 Sentence Transformers Model

In the project “distiluse-base-multilingual-cased-v1” is used. The model maps sentences & paragraphs to a 512-dimensional dense vector space and can be used for tasks like clustering or semantic search (adapted from “<https://huggingface.co/sentence-transformers/distiluse-base-multilingual-cased-v1>”). It is a version of multilingual Universal Sentence Encoder and supports 15 languages: Arabic, Chinese, Dutch, English, French, German, Italian, Korean, Polish, Portuguese, Russian, Spanish, Turkish.

3 Possible Improvement Ideas on SAASS

- The empty answers of the student answers file can be evaluated differently. The system can automatically assign 0 point to these empty answers. In current system there is no configuration to handle that. The system assigns a point by looking which answer has the least lexeme. Because less lexeme is closer to the empty answer.

- The answers can be edited to correct spelling and grammatical errors using tools like spaCy (free, open-source Python library that provides tokenization, lemmatization, etc.) or SymSpell (Symmetric Spell Checker, fast and efficient spell-checking and autocorrect library.).

- Semantic relevance and term-based relevance can be combined to improve SAASS. TF-IDF can handle exact keyword matching and semantic model (in the case this is “distiluse-base-multilingual-cased-v1”) can handle semantic similarity.

- A more accurate method for similarity calculation, if available, could be used to obtain more precise and accurate similarity results, thus enhancing the overall accuracy of the system.

- Domain specific data sets can be provided for the situations that the used model is not sufficient to evaluate answers precisely.