Rehab Kareem Kadthim *et al*, International Journal of Computer Science and Mobile Applications,
Vol.10 Issue. 9, September- 2022, pg. 61-71 ISSN: 2321-8363
Impact Factor: 6.308
(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

Cheating Detection in E-exam using Similarity Methods

Rehab Kareem Kadthim[1]*; Zuhair Hussein Ali[2]

[1,2]Department of Computer Science, College of Education, Mustansiriyah University, Iraq

[1] rehabkareem91aa@uomustansiriyah.edu.iq

[2] Zuhair72h@uomustansiriyah.edu.iq

DOI: 10.5281/zenodo.7133121

Abstract: With the growth of the Internet and technology, the use of e-learning has expanded significantly and electronic exams have become an alternative to traditional exams. The phenomenon of cheating was considered to reduce educational sobriety and violate electronic integrity. This research proposed a system for detecting cheating in the answer of students in online exams. The proposed system is based on similarity measurement. where the model uses similarity (cosine, Jaccard, and overlap) algorithms to find cheating in essay questions only. The results of this system showed that using the cosine similarity method is better than the rest of the other methods. that the proposed system provides a reliable test via the Internet and a fair system that maintains electronic safety.

Keywords: student cheating, similarity methods, online exams, Essay Answer.

1. Introduction

The fundamental advantage of e-learning is that everyone may use it, regardless of age, location, or available study time [1]. Many educational institutions utilize the LMS as a platform to access E-learning resources since it is a crucial component of an E-learning system.

Evaluating students' performance during online tests is the hardest part of e-learning. Specifically, online exams are typically taken on e-learning platforms without students and professors being present in the same place physically. Due to these flaws, online tests lose part of their credibility [2].

The existence of many examination types gives the students a wide range of opportunities for cheating, such as utilizing the student use of screen sharing, cellular phones, and students meeting in the same room [3]. Therefore, there must be ways to prevent or at least detect cheating. Various methods exist for the online cheating detection some of these are direct while others are indirect [4]. Direct detection required the existence of physical equipment, such as the use of online proctoring with a webcam or a camera designed with special techniques [5]. In this paper, work is done to find cheating by identifying the similarity in the students' answers, which are of the type of essay questions. The fundamental purpose of using similarity measures is to quantify the degree of similarity between two answers. This paper used three different forms of similarity algorithms (cosine, Jaccard, and overlap).

2. Related Work

Recently, researchers have made many efforts to develop a realistic approach to identify ways to detect cheating in electronic exams. Studies are still expanding to provide the aspect of monitoring cheating cases. The past few decades present a set of studies related to approach.

Cavalcanti et al., 2012 [6] used a decision tree to build two classification models to detect cheating: one based on the cosine similarity, and the other based on the overlap coefficient. The results of both classifiers were compared against the results produced by a domain expert. The results proved that the decision tree with overlap coefficient is better to use for the purposed of cheating detection.

Li, 2019, [7], introduces a novel probabilistic strategy for detecting cheaters. The proposed method is based on analyzing the question-answering record and discovering that the incorrect answers have an extremely low chance under the condition of passing the test. The proposed model combines both probability estimation and feature bagging. Probability estimation is used to detect outlier detection, where every answer time exceeds an hour or the number of answers less than one hundred.

Golden et al., 2021 [8], used a case study to assess the prevalence of potential cheating and propose preventive measures that may be implemented. They utilized cheating intelligence agents as a tool for identifying online cheating behaviors, which is made up of two key modules: IP

ISSN: 2321-8363
Impact Factor: 6.308
(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

detector and behavior detector. The intelligence agent observes the students' behavior and can prevent and detect any malicious conduct. It can be used to give randomized multiple-choice questions in a course examination and can be combined with online learning programs to monitor students' behavior. The proposed method was validated by testing it on a variety of datasets (mid-term and final-term exams).

3. Proposed Methodology for Cheating Detecting

This section introduces the proposed method for encountering student cheating in exams. dataset consists of (30) students' answers for five different subjects. Each student answered one essay question for each subject. These answers are stored as file names a1, a2, and a3, and a4, a5 associated with student numbers. The model based on similarity methods (cosine, Jaccard, and overlap) used to determine the similarity between students' answers. The document's text passed through two phases, preprocessing and similarity methods. The preprocessing consists of two main stages tokenization and punctuation removal. Figure (1) depicts the model's phases of the application:

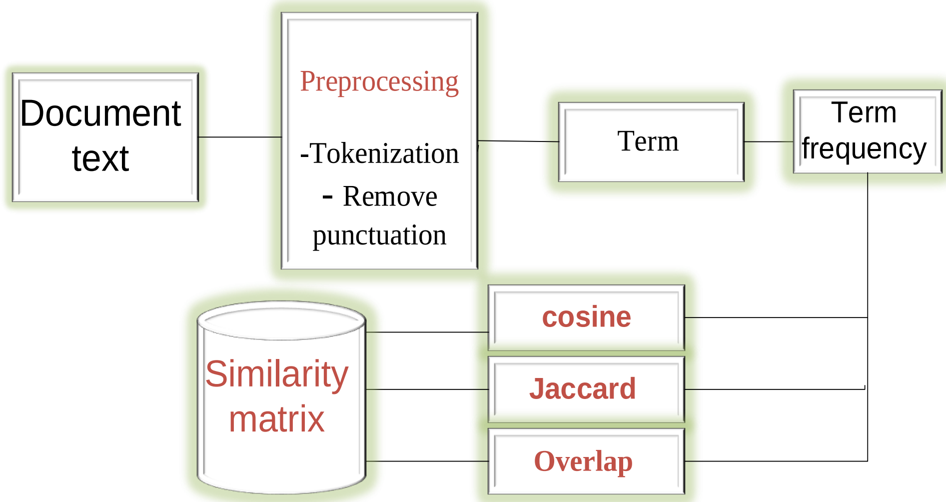

Figure (1): Block Diagram for Proposed Model.

3.1 Dataset Collection

One of the important things that are worth to be mentioned before presenting the experimental results to evaluate the proposed system is providing a brief description of the dataset. These dataset collections were obtained from Iraqi universities. The dataset consists of (30) students' answers for two different subjects. Each student answered one essay question for each subject. These answers are stored as file names a1, and a2 associated with student numbers. For example, a15, refers to the answer of the first subject for student number 5, a27the refer to answer the question of the second subject by student number 7.

3.2 Preprocessing

Preprocessing is a critical step. It attempts to split a set of texts into phrases that allow for the use of similarity assessment. the Answers should have gone through numerous phases, including tokenization and punctuation. The tokenization process includes separating a stream of text into tokens, which can be terms, symbols, or other significant items. The goal of tokenization is to explore each token individually. The token list is used as input for further processing. Figure (3) shows an example of tokenization process.

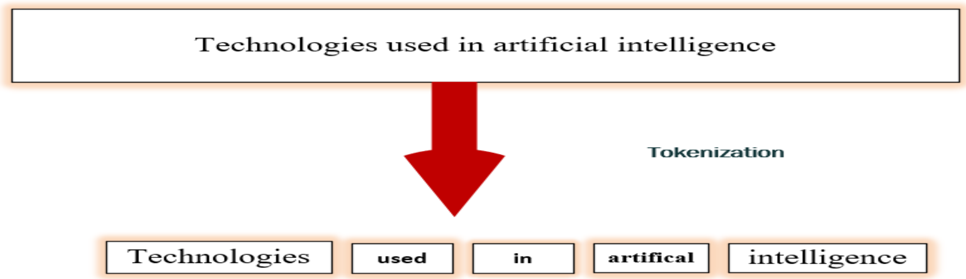

Figure (3): Tokenization Example.

While punctuation Removal This stage is concerned with punctuation removed from the set of tokens These aids in cleaning the data and treating each text equally. Data cleaning depends on the type of data, and it is especially important if the data is textual. Any punctuation that appears in the token list will be removed. Some examples of punctuation are {"(",",", ".", "!")}.

3.2 Similarity Methods

This section is concerned with describing the similarity methods used in the proposed model.

3.2.1 Cosine Similarity

These were appropriate for sparse data sources such as documents (and employed Jaccard and Cosine metrics in the proposed system). In this case, the similarity is defined as 1 if the attribute values are identical and 0 otherwise. Dissimilarity would be provided differently: 0 if the attribute values are the same, 1 otherwise.

Cosine similarity is a metric that defines how closely connected papers are independent of size. It calculates the cosine of the angle formed by two vectors projected in multidimensional space [9]. A, and B's cosine similarity is referred to as:

$$C(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}} \qquad (1)$$

Where,

A: defines the first set.

B: defines the second set.

$A \cdot B$: shows the dot product of the two sets (A and B).

$\|A\| \times \|B\|$: acts as the product of the two sets' lengths (A and B).

Cosine's value varies between [-1, 1]. When two papers are identical, their vectors point in the same direction, resulting in a tiny angle with a cosine value close to one. When two vectors point in opposite directions from the origin, a big angle is formed, and the cosine value is near -1;

Consequently, the documents are different, and no similarity is mapped [10, 11].

3.2.2 Jaccard Similarity

Jaccard similarity measures the similarity of two sets. It is defined as the intersection size of two sets divided by their union size. A and B's Jaccard similarity is referred to as:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \qquad (2)$$

Where, A: represents the first set.

B: represents the second set.

|A∩B|: shows the magnitude of the intersection between A and B.

|A ∪ B|: shows the union between A and B.

The amount of resemblance is represented by a number between 0 and 1. The two texts are identical when the value is 1, and they are distinct when the value is 0 [10, 11]. In the following algorithm, the stages of the Jaccard similarity algorithm are shown Equation (2)

3.2.3 Overlap Similarity

Is overlap similarity a measure of how similar two sets are? It's calculated by dividing the intersection size of two sets by the smaller of the two.
A full match occurs when one set is a subset of the other [12]. The overlap similarity is defined as follows for two sets A and B

$$Sim(A, B) = \frac{A \cap B}{Min(A,B)} \qquad (3)$$

Where,

A: represents the first set.

B: represents the second set.

|A∩B|: shows the magnitude of the intersection between A and B.

min (|A|, |B|): acts the smaller size of A and B.

The degree of resemblance is measured on a scale of 0 to 1. The answer is 1 when the two papers are identical or one is a subset of the other, and 0 when the two documents are completely different [13]

IJCSMA

Rehab Kareem Kadthim *et al*, International Journal of Computer Science and Mobile Applications,
Vol.10 Issue. 9, September- 2022, pg. 61-71 ISSN: 2321-8363
Impact Factor: 6.308
(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

4. Result

In this section, a series of experiments and tests are executed in order to assess the performance of cheating detection in online exams using proposed model. Python programming language version 3.8.5 was used to implement the proposed model. This section concerned with evaluating the performance other f model1, where different algorithms applied to compute the similarity between text documents. These documents represent students' answers for five subjects. The questions are the essay type.

4.2 Cosine Similarity Results

Tables (1), (2) show the results of the cosine similarity

Subject1:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	1	0.96	0.64	0.4	0.17	0.47	0.39	0.92	0.59	0.56	0.58	0.4	0.52	0.47	0.11	0.09	0.94	0.63	0.9	0.67	0.1	0.56	0.94	0.76	0.45	0.11	0.95	0.97	0.1	0.09
2	0.96	1	0.64	0.42	0.17	0.47	0.39	0.89	0.59	0.56	0.53	0.42	0.52	0.47	0.11	0.09	0.9	0.66	0.9	0.67	0.1	0.56	0.94	0.74	0.45	0.11	0.95	0.93	0.1	0.09
3	0.64	0.64	1	0.51	0.18	0.66	0.41	0.6	0.92	0.85	0.77	0.51	0.64	0.6	0.1	0.1	0.62	0.94	0.68	0.46	0.13	0.71	0.62	0.63	0.57	0.1	0.66	0.66	0.14	0.1
4	0.4	0.42	0.51	1	0.23	0.3	0.22	0.3	0.53	0.45	0.37	0.99	0.41	0.35	0.11	0.11	0.38	0.58	0.51	0.27	0.14	0.46	0.42	0.48	0.32	0.11	0.41	0.4	0.14	0.11
5	0.17	0.17	0.18	0.23	1	0.2	0.06	0.15	0.19	0.17	0.19	0.23	0.35	0.19	0.94	0.96	0.17	0.19	0.13	0.31	0.69	0.33	0.17	0.21	0.17	0.96	0.17	0.22	0.83	0.91
6	0.47	0.47	0.66	0.3	0.2	1	0.41	0.45	0.67	0.66	0.61	0.3	0.55	0.45	0.18	0.17	0.46	0.59	0.43	0.39	0.21	0.6	0.44	0.37	0.42	0.17	0.47	0.5	0.21	0.17
7	0.39	0.39	0.41	0.22	0.06	0.41	1	0.37	0.35	0.39	0.38	0.22	0.44	0.39	0.03	0.03	0.35	0.39	0.38	0.34	0.03	0.47	0.39	0.31	0.37	0.03	0.38	0.34	0.03	0.03
8	0.92	0.89	0.6	0.3	0.15	0.45	0.37	1	0.53	0.55	0.56	0.3	0.51	0.46	0.11	0.09	0.94	0.59	0.82	0.68	0.1	0.54	0.91	0.71	0.44	0.11	0.91	0.91	0.11	0.09
9	0.59	0.59	0.92	0.53	0.19	0.67	0.35	0.53	1	0.91	0.78	0.53	0.64	0.6	0.11	0.11	0.56	0.89	0.64	0.42	0.14	0.71	0.57	0.6	0.56	0.11	0.6	0.61	0.14	0.11
10	0.56	0.56	0.85	0.45	0.17	0.66	0.39	0.55	0.91	1	0.78	0.46	0.65	0.63	0.11	0.11	0.52	0.83	0.58	0.48	0.14	0.73	0.61	0.58	0.59	0.11	0.6	0.55	0.15	0.11
11	0.58	0.53	0.77	0.37	0.19	0.61	0.38	0.56	0.78	0.78	1	0.37	0.64	0.56	0.15	0.16	0.56	0.72	0.56	0.44	0.17	0.7	0.53	0.57	0.53	0.14	0.54	0.58	0.17	0.16
12	0.4	0.42	0.51	0.99	0.23	0.3	0.22	0.3	0.53	0.46	0.37	1	0.41	0.35	0.12	0.11	0.38	0.59	0.52	0.27	0.14	0.46	0.42	0.49	0.32	0.11	0.42	0.4	0.14	0.11
13	0.52	0.52	0.64	0.41	0.35	0.55	0.44	0.51	0.64	0.65	0.64	0.41	1	0.62	0.31	0.3	0.51	0.65	0.48	0.58	0.32	0.97	0.52	0.51	0.58	0.3	0.53	0.52	0.33	0.3
14	0.47	0.47	0.6	0.35	0.19	0.45	0.39	0.46	0.6	0.63	0.56	0.35	0.62	1	0.15	0.13	0.45	0.61	0.48	0.53	0.18	0.67	0.47	0.48	0.96	0.15	0.48	0.47	0.19	0.13
15	0.11	0.11	0.1	0.11	0.94	0.18	0.03	0.11	0.11	0.11	0.15	0.12	0.31	0.15	1	0.92	0.11	0.1	0.04	0.3	0.77	0.29	0.11	0.14	0.14	0.97	0.11	0.16	0.81	0.92
16	0.09	0.09	0.1	0.11	0.96	0.17	0.03	0.09	0.11	0.11	0.16	0.11	0.3	0.13	0.92	1	0.08	0.1	0.02	0.25	0.68	0.28	0.09	0.13	0.12	0.95	0.09	0.13	0.82	0.95
17	0.94	0.9	0.62	0.38	0.17	0.46	0.35	0.94	0.56	0.52	0.56	0.38	0.51	0.45	0.11	0.08	1	0.61	0.88	0.67	0.1	0.54	0.88	0.72	0.43	0.11	0.9	0.93	0.1	0.08
18	0.63	0.66	0.94	0.58	0.19	0.59	0.39	0.59	0.89	0.83	0.72	0.59	0.65	0.61	0.1	0.1	0.61	1	0.72	0.46	0.13	0.72	0.63	0.64	0.57	0.1	0.65	0.65	0.13	0.1
19	0.9	0.9	0.68	0.51	0.13	0.43	0.38	0.82	0.64	0.58	0.56	0.52	0.48	0.48	0.04	0.02	0.88	0.72	1	0.58	0.04	0.54	0.86	0.76	0.45	0.04	0.87	0.9	0.04	0.02
20	0.67	0.67	0.46	0.27	0.31	0.39	0.34	0.68	0.42	0.48	0.44	0.27	0.58	0.53	0.3	0.25	0.67	0.46	0.58	1	0.32	0.6	0.67	0.55	0.48	0.29	0.65	0.63	0.3	0.25
21	0.1	0.1	0.13	0.14	0.69	0.21	0.03	0.1	0.14	0.14	0.17	0.14	0.32	0.18	0.77	0.68	0.1	0.13	0.04	0.32	1	0.3	0.1	0.15	0.17	0.7	0.1	0.15	0.77	0.68
22		0.56	0.71	0.46	0.33	0.6	0.47	0.54	0.71	0.73	0.7	0.46	0.97	0.67	0.29	0.28	0.54	0.72	0.54	0.6	0.3	1	0.56	0.55	0.62	0.28	0.57	0.56	0.3	0.28
23	0.94	0.94	0.62	0.42	0.17	0.44	0.39	0.91	0.57	0.61	0.53	0.42	0.52	0.47	0.11	0.09	0.88	0.63	0.86	0.67	0.1	0.56	1	0.76	0.45	0.11	0.96	0.91	0.1	0.09
24	0.76	0.74	0.63	0.48	0.21	0.37	0.31	0.71	0.6	0.58	0.57	0.49	0.51	0.48	0.14	0.13	0.72	0.64	0.76	0.55	0.15	0.55	0.76	1	0.47	0.13	0.77	0.78	0.15	0.13
25	0.45	0.45	0.57	0.32	0.17	0.42	0.37	0.44	0.56	0.59	0.53	0.32	0.58	0.96	0.14	0.12	0.43	0.57	0.45	0.48	0.17	0.62	0.45	0.47	1	0.13	0.46	0.45	0.17	0.12
26	0.11	0.11	0.1	0.11	0.96	0.17	0.03	0.11	0.11	0.11	0.14	0.11	0.3	0.15	0.97	0.95	0.11	0.1	0.04	0.29	0.7	0.28	0.11	0.13	0.13	1	0.11	0.16	0.84	0.95
27	0.95	0.95	0.66	0.41	0.17	0.47	0.38	0.91	0.6	0.6	0.54	0.42	0.53	0.48	0.11	0.09	0.9	0.65	0.87	0.65	0.1	0.57	0.96	0.77	0.46	0.11	1	0.94	0.1	0.09
28	0.97	0.93	0.66	0.4	0.22	0.5	0.34	0.91	0.61	0.55	0.58	0.4	0.52	0.47	0.16	0.13	0.93	0.65	0.9	0.63	0.15	0.56	0.91	0.78	0.45	0.16	0.94	1	0.15	0.13
29	0.1	0.1	0.14	0.14	0.83	0.21	0.03	0.11	0.14	0.15	0.17	0.14	0.33	0.19	0.81	0.82	0.1	0.13	0.04	0.3	0.77	0.3	0.1	0.15	0.17	0.84	0.1	0.15	1	0.79
30	0.09	0.09	0.1	0.11	0.91	0.17	0.03	0.09	0.11	0.11	0.16	0.11	0.3	0.13	0.92	0.95	0.08	0.1	0.02	0.25	0.68	0.28	0.09	0.13	0.12	0.95	0.09	0.13	0.79	1

Subject2:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	1	0.96	0.35	0.39	0.31	0.08	0.25	0.96	0.34	0.65	0.29	0.23	0.41	0.18	0.29	0.29	0.95	0.03	0.17	0.44	0.25	0.05	0.95	0.42	0.18	0.29	0.92	0.96	0.45	0.29
2	0.96	1	0.36	0.4	0.32	0.08	0.25	0.91	0.33	0.69	0.29	0.24	0.4	0.19	0.29	0.3	0.91	0.03	0.17	0.44	0.26	0.05	0.9	0.41	0.19	0.3	0.87	0.91	0.44	0.3
3	0.35	0.36	1	0.74	0.33	0.18	0.29	0.34	0.85	0.23	0.3	0.3	0.35	0.46	0.3	0.32	0.38	0.18	0.43	0.33	0.31	0.16	0.36	0.35	0.51	0.3	0.33	0.35	0.38	0.31
4	0.39	0.4	0.74	1	0.77	0.34	0.33	0.37	0.63	0.24	0.8	0.82	0.5	0.38	0.8	0.8	0.4	0.26	0.35	0.37	0.8	0.22	0.38	0.51	0.4	0.8	0.38	0.38	0.48	0.81
5	0.31	0.32	0.33	0.77	1	0.25	0.24	0.3	0.28	0.19	0.93	0.84	0.47	0.2	0.93	0.96	0.3	0.32	0.18	0.39	0.85	0.33	0.29	0.47	0.21	0.92	0.31	0.31	0.46	0.93
6	0.08	0.08	0.18	0.34	0.25	1	0.13	0.08	0.17	0.03	0.33	0.38	0.61	0.2	0.33	0.3	0.07	0.24	0.09	0.16	0.35	0.2	0.08	0.61	0.18	0.33	0.12	0.07	0.49	0.33
7	0.25	0.25	0.29	0.33	0.24	0.13	1	0.23	0.25	0.22	0.24	0.25	0.27	0.23	0.24	0.26	0.26	0.11	0.22	0.24	0.26	0.12	0.25	0.27	0.22	0.24	0.26	0.25	0.28	0.25
8	0.96	0.91	0.34	0.37	0.3	0.08	0.23	1	0.36	0.63	0.28	0.23	0.4	0.12	0.28	0.28	0.96	0	0.15	0.43	0.25	0.02	0.9	0.41	0.15	0.28	0.89	0.91	0.44	0.28
9	0.34	0.33	0.85	0.63	0.28	0.17	0.25	0.36	1	0.22	0.26	0.26	0.32	0.37	0.26	0.28	0.37	0.16	0.35	0.3	0.26	0.16	0.35	0.32	0.44	0.26	0.34	0.34	0.35	0.26
10	0.65	0.69	0.23	0.24	0.19	0.03	0.22	0.63	0.22	1	0.17	0.18	0.32	0.14	0.17	0.2	0.66	0.04	0.15	0.37	0.22	0.08	0.68	0.33	0.15	0.17	0.55	0.62	0.35	0.18
11	0.29	0.29	0.3	0.8	0.93	0.33	0.24	0.28	0.26	0.17	1	0.94	0.45	0.19	1	0.97	0.28	0.25	0.17	0.3	0.91	0.23	0.27	0.46	0.2	0.97	0.29	0.28	0.41	0.98
12	0.23	0.24	0.3	0.82	0.84	0.38	0.25	0.23	0.26	0.18	0.94	1	0.47	0.18	0.94	0.92	0.24	0.25	0.18	0.27	0.97	0.19	0.24	0.48	0.2	0.92	0.24	0.23	0.41	0.93
13	0.41	0.4	0.35	0.5	0.47	0.61	0.27	0.4	0.32	0.32	0.45	0.47	1	0.33	0.45	0.48	0.43	0.72	0.29	0.82	0.49	0.69	0.39	0.99	0.37	0.46	0.39	0.4	0.97	0.46
14	0.18	0.19	0.46	0.38	0.2	0.2	0.23	0.12	0.37	0.14	0.19	0.18	0.33	1	0.19	0.2	0.18	0.29	0.84	0.31	0.2	0.28	0.18	0.32	0.86	0.19	0.15	0.18	0.34	0.19
15	0.29	0.29	0.3	0.8	0.93	0.33	0.24	0.28	0.26	0.17	1	0.94	0.45	0.19	1	0.97	0.28	0.25	0.17	0.3	0.91	0.23	0.27	0.46	0.2	0.97	0.29	0.28	0.41	0.98
16	0.29	0.3	0.32	0.8	0.96	0.3	0.26	0.28	0.28	0.2	0.97	0.92	0.48	0.2	0.97	1	0.29	0.29	0.18	0.35	0.93	0.29	0.3	0.48	0.21	0.95	0.29	0.29	0.45	0.96
17	0.95	0.91	0.38	0.4	0.3	0.07	0.26	0.96	0.37	0.66	0.28	0.24	0.43	0.18	0.28	0.29	1	0.03	0.19	0.45	0.26	0.04	0.9	0.44	0.2	0.28	0.87	0.91	0.46	0.29
18	0.03	0.03	0.18	0.26	0.32	0.24	0.11	0	0.16	0.04	0.25	0.25	0.72	0.29	0.25	0.29	0.03	1	0.27	0.82	0.28	0.95	0.03	0.66	0.36	0.23	0.02	0.03	0.76	0.23
19	0.17	0.17	0.43	0.35	0.18	0.09	0.22	0.15	0.35	0.15	0.17	0.18	0.29	0.84	0.17	0.18	0.19	0.27	1	0.31	0.2	0.24	0.17	0.28	0.82	0.17	0.13	0.17	0.31	0.17
20	0.44	0.44	0.33	0.37	0.39	0.16	0.24	0.43	0.3	0.37	0.3	0.27	0.82	0.31	0.3	0.35	0.45	0.82	0.31	1	0.33	0.82	0.41	0.78	0.38	0.28	0.38	0.43	0.89	0.28
21	0.25	0.26	0.31	0.8	0.85	0.35	0.26	0.25	0.26	0.22	0.91	0.97	0.49	0.2	0.91	0.93	0.26	0.28	0.2	0.33	1	0.24	0.28	0.5	0.22	0.91	0.25	0.25	0.44	0.9
22	0.05	0.05	0.16	0.22	0.33	0.2	0.12	0.02	0.16	0.08	0.23	0.19	0.69	0.28	0.23	0.29	0.04	0.95	0.24	0.82	0.24	1	0.05	0.64	0.32	0.21	0.05	0.04	0.73	0.21
23	0.95	0.9	0.36	0.38	0.29	0.08	0.25	0.9	0.35	0.68	0.27	0.24	0.39	0.18	0.27	0.3	0.9	0.03	0.17	0.41	0.28	0.05	1	0.4	0.18	0.27	0.86	0.9	0.43	0.27
24	0.42	0.41	0.35	0.51	0.47	0.61	0.27	0.41	0.32	0.33	0.46	0.48	0.99	0.32	0.46	0.48	0.44	0.66	0.28	0.78	0.5	0.64	0.4	1	0.34	0.47	0.4	0.42	0.95	0.47
25	0.18	0.19	0.51	0.4	0.21	0.18	0.22	0.15	0.44	0.15	0.2	0.2	0.37	0.86	0.2	0.21	0.2	0.36	0.82	0.38	0.22	0.32	0.18	0.34	1	0.2	0.15	0.18	0.39	0.2
26	0.29	0.3	0.3	0.8	0.92	0.33	0.24	0.28	0.26	0.17	0.97	0.92	0.46	0.19	0.97	0.95	0.28	0.23	0.17	0.28	0.91	0.21	0.27	0.47	0.2	1	0.29	0.29	0.41	0.98
27	0.92	0.87	0.33	0.38	0.31	0.12	0.26	0.89	0.34	0.55	0.29	0.24	0.39	0.15	0.29	0.29	0.87	0.02	0.13	0.38	0.25	0.05	0.86	0.4	0.15	0.29	1	0.88	0.42	0.29
28	0.96	0.91	0.35	0.38	0.31	0.07	0.25	0.91	0.34	0.62	0.28	0.23	0.4	0.18	0.28	0.29	0.91	0.03	0.17	0.43	0.25	0.04	0.9	0.42	0.18	0.29	0.88	1	0.44	0.29
29	0.45	0.44	0.38	0.48	0.46	0.49	0.28	0.44	0.35	0.35	0.41	0.41	0.97	0.34	0.41	0.45	0.46	0.76	0.31	0.89	0.44	0.73	0.43	0.95	0.39	0.41	0.42	0.44	1	0.42
30	0.29	0.3	0.31	0.81	0.93	0.33	0.25	0.28	0.26	0.18	0.98	0.93	0.46	0.19	0.98	0.96	0.29	0.23	0.17	0.28	0.9	0.21	0.27	0.47	0.2	0.98	0.29	0.29	0.42	1

IJCSMA
Rehab Kareem Kadthim *et al*, International Journal of Computer Science and Mobile Applications,
Vol.10 Issue. 9, September- 2022, pg. 61-71 ISSN: 2321-8363
Impact Factor: 6.308
(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

By following the above table, it was noted that there are many similarities in the tables, which can be happened in this type of question. But the interesting thing is that there are two groups: the first group shaded with red and the other shaded with green. For students with numbers (1, 2, 8, 17,23,27,28) and (5, 15, 16,26,30) respectively to each group. As the students of these two groups have a similar percentage in all subjects, which gives the impression that there are cases of cheating among them.
It is noted the efficiency of the cosine similarity method for this type of applications. Cosine similarity used to measure how similar the documents are irrespective of their size. The advantage of the cosine similarity is to calculate the frequency of words and take it into account when calculating the similarity. This is a very important thing for fraud *detection*. Where some students try to preserve the original text of the answer with the addition of some unnecessary sentences. These unnecessary sentences reduce the similarity between the students' answers, but at the same time, the original answers contain a high percentage of similarity. Adding some sentences is a way some students use to confuse cheating detections. In order for the answers to appear different from each other

4.3 Jaccard Similarity Result

Tables (3),(4) show the results of the jaccard similarity

Subject1:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	1	0.87	0.33	0.07	0.08	0.27	0.15	0.79	0.29	0.22	0.28	0.07	0.09	0.16	0.06	0.04	0.81	0.3	0.74	0.39	0.06	0.09	0.81	0.57	0.17	0.06	0.82	0.94	0.06	0.04
2	0.87	1	0.33	0.09	0.08	0.27	0.15	0.69	0.29	0.22	0.24	0.09	0.09	0.16	0.06	0.04	0.71	0.36	0.74	0.39	0.06	0.09	0.81	0.53	0.17	0.06	0.82	0.82	0.06	0.04
3	0.33	0.33	1	0.1	0.08	0.46	0.16	0.31	0.82	0.64	0.54	0.1	0.18	0.29	0.06	0.06	0.33	0.83	0.29	0.22	0.08	0.17	0.3	0.26	0.31	0.06	0.34	0.34	0.08	0.06
4	0.07	0.09	0.1	1	0.05	0.1	0.08	0.07	0.11	0.11	0.08	0.94	0.07	0.08	0.03	0.03	0.06	0.1	0.06	0.09	0.03	0.08	0.09	0.09	0.09	0.03	0.09	0.07	0.03	0.03
5	0.08	0.08	0.08	0.05	1	0.07	0.04	0.08	0.07	0.07	0.06	0.05	0.04	0.05	0.84	0.88	0.08	0.08	0.06	0.13	0.42	0.05	0.08	0.09	0.06	0.88	0.08	0.1	0.6	0.73
6	0.27	0.27	0.46	0.1	0.07	1	0.19	0.28	0.5	0.45	0.4	0.1	0.2	0.24	0.06	0.06	0.27	0.41	0.23	0.18	0.07	0.19	0.25	0.17	0.26	0.06	0.29	0.31	0.07	0.06
7	0.15	0.15	0.16	0.08	0.04	0.19	1	0.14	0.15	0.15	0.14	0.08	0.13	0.14	0.02	0.02	0.12	0.16	0.15	0.13	0.02	0.12	0.15	0.13	0.15	0.02	0.14	0.14	0.02	0.02
8	0.79	0.69	0.31	0.07	0.08	0.28	0.14	1	0.28	0.25	0.29	0.07	0.1	0.17	0.06	0.04	0.85	0.31	0.63	0.4	0.06	0.09	0.74	0.53	0.18	0.06	0.75	0.8	0.06	0.04
9	0.29	0.29	0.82	0.11	0.07	0.5	0.15	0.28	1	0.8	0.6	0.11	0.21	0.32	0.05	0.05	0.29	0.72	0.29	0.2	0.07	0.2	0.27	0.26	0.35	0.05	0.31	0.31	0.07	0.05
10	0.22	0.22	0.64	0.11	0.07	0.45	0.15	0.25	0.8	1	0.57	0.11	0.21	0.32	0.05	0.05	0.22	0.6	0.22	0.2	0.07	0.21	0.29	0.26	0.35	0.05	0.28	0.24	0.07	0.05
11	0.28	0.24	0.54	0.08	0.06	0.4	0.14	0.29	0.6	0.57	1	0.08	0.22	0.28	0.04	0.04	0.28	0.48	0.26	0.19	0.06	0.21	0.24	0.25	0.3	0.04	0.25	0.3	0.06	0.04
12	0.07	0.09	0.1	0.94	0.05	0.1	0.08	0.07	0.11	0.11	0.08	1	0.07	0.08	0.03	0.03	0.06	0.1	0.06	0.09	0.03	0.08	0.09	0.09	0.09	0.03	0.08	0.07	0.03	0.03
13	0.109	0.09	0.18	0.07	0.04	0.2	0.13	0.1	0.21	0.21	0.22	0.07	1	0.2	0.03	0.03	0.09	0.19	0.09	0.12	0.05	0.89	0.09	0.09	0.19	0.03	0.1	0.1	0.05	0.03
14	0.16	0.16	0.29	0.08	0.05	0.24	0.14	0.17	0.32	0.32	0.28	0.08	0.2	1	0.04	0.04	0.16	0.29	0.16	0.23	0.05	0.21	0.16	0.2	0.83	0.04	0.17	0.17	0.05	0.04
15	0.06	0.06	0.06	0.03	0.84	0.06	0.02	0.06	0.05	0.05	0.04	0.03	0.03	0.04	1	0.8	0.06	0.06	0.04	0.11	0.47	0.04	0.06	0.07	0.04	0.96	0.06	0.08	0.6	0.8
16	0.04	0.04	0.06	0.03	0.88	0.06	0.02	0.04	0.05	0.05	0.04	0.03	0.03	0.04	0.8	1	0.04	0.06	0.02	0.09	0.39	0.04	0.04	0.07	0.04	0.83	0.04	0.06	0.57	0.83
17	0.81	0.71	0.33	0.06	0.08	0.27	0.12	0.85	0.29	0.22	0.28	0.06	0.09	0.16	0.06	0.04	1	0.3	0.69	0.39	0.06	0.09	0.66	0.53	0.17	0.06	0.71	0.82	0.06	0.04
18	0.3	0.36	0.83	0.1	0.08	0.41	0.16	0.31	0.72	0.6	0.48	0.1	0.19	0.29	0.06	0.06	0.3	1	0.32	0.22	0.08	0.18	0.3	0.26	0.31	0.06	0.31	0.31	0.08	0.06
19	0.74	0.74	0.29	0.06	0.06	0.23	0.15	0.63	0.29	0.22	0.26	0.06	0.09	0.16	0.04	0.02	0.69	0.32	1	0.35	0.04	0.08	0.59	0.51	0.17	0.04	0.65	0.74	0.04	0.02
20	0.39	0.39	0.22	0.09	0.13	0.18	0.13	0.4	0.2	0.2	0.19	0.09	0.12	0.23	0.11	0.09	0.39	0.22	0.35	1	0.11	0.11	0.39	0.35	0.19	0.11	0.37	0.37	0.1	0.09
21	0.06	0.06	0.08	0.03	0.42	0.07	0.02	0.06	0.07	0.07	0.06	0.03	0.05	0.05	0.47	0.39	0.06	0.08	0.04	0.11	1	0.04	0.06	0.06	0.06	0.44	0.06	0.08	0.58	0.39
22	0.09	0.09	0.17	0.08	0.05	0.19	0.12	0.09	0.2	0.21	0.21	0.08	0.89	0.21	0.04	0.04	0.09	0.18	0.08	0.11	0.04	1	0.09	0.09	0.19	0.04	0.09	0.1	0.04	0.04
23	0.81	0.81	0.3	0.09	0.08	0.25	0.15	0.74	0.27	0.29	0.24	0.09	0.09	0.16	0.06	0.04	0.66	0.3	0.59	0.39	0.06	0.09	1	0.57	0.17	0.06	0.88	0.76	0.06	0.04
24	0.57	0.53	0.26	0.09	0.09	0.17	0.13	0.53	0.26	0.26	0.25	0.09	0.09	0.2	0.07	0.07	0.53	0.26	0.51	0.35	0.06	0.09	0.57	1	0.22	0.07	0.58	0.58	0.06	0.07
25	0.17	0.17	0.31	0.09	0.06	0.26	0.15	0.18	0.35	0.35	0.3	0.09	0.19	0.83	0.04	0.04	0.17	0.31	0.17	0.19	0.06	0.19	0.17	0.22	1	0.04	0.19	0.19	0.05	0.04
26	0.06	0.06	0.06	0.03	0.88	0.06	0.02	0.06	0.05	0.05	0.04	0.03	0.03	0.04	0.96	0.83	0.06	0.06	0.04	0.11	0.44	0.04	0.06	0.07	0.04	1	0.06	0.08	0.62	0.83
27	0.82	0.82	0.34	0.09	0.08	0.29	0.14	0.75	0.31	0.28	0.25	0.08	0.1	0.17	0.06	0.04	0.71	0.31	0.65	0.37	0.06	0.09	0.88	0.58	0.19	0.06	1	0.82	0.06	0.04
28	0.94	0.82	0.34	0.07	0.1	0.31	0.14	0.8	0.31	0.24	0.3	0.07	0.1	0.17	0.08	0.06	0.82	0.31	0.74	0.37	0.08	0.1	0.76	0.58	0.19	0.08	0.82	1	0.08	0.06
29	0.06	0.06	0.08	0.03	0.6	0.07	0.02	0.06	0.07	0.07	0.06	0.03	0.05	0.05	0.6	0.57	0.06	0.08	0.04	0.1	0.58	0.04	0.06	0.06	0.05	0.62	0.06	0.08	1	0.52
30	0.04	0.04	0.06	0.03	0.73	0.06	0.02	0.04	0.05	0.05	0.04	0.03	0.03	0.04	0.8	0.83	0.04	0.06	0.02	0.09	0.39	0.04	0.04	0.07	0.04	0.83	0.04	0.06	0.52	1

Subject2:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	1	0.9	0.09	0.1	0.08	0.04	0.09	0.92	0.1	0.45	0.06	0.06	0.36	0.07	0.06	0.06	0.88	0.02	0.07	0.38	0.06	0.03	0.88	0.37	0.08	0.07	0.82	0.88	0.36	0.06
2	0.9	1	0.09	0.1	0.08	0.04	0.09	0.83	0.1	0.51	0.06	0.06	0.34	0.07	0.06	0.06	0.79	0.02	0.07	0.35	0.06	0.03	0.79	0.35	0.08	0.07	0.73	0.79	0.34	0.06
3	0.09	0.09	1	0.47	0.07	0.06	0.08	0.08	0.65	0.09	0.07	0.09	0.12	0.23	0.07	0.07	0.1	0.1	0.23	0.13	0.08	0.08	0.09	0.12	0.25	0.07	0.08	0.09	0.12	0.07
4	0.1	0.1	0.47	1	0.5	0.09	0.06	0.09	0.34	0.08	0.51	0.54	0.17	0.14	0.51	0.49	0.1	0.13	0.14	0.16	0.51	0.11	0.08	0.17	0.15	0.48	0.09	0.09	0.17	0.49
5	0.08	0.08	0.07	0.5	1	0.14	0.05	0.07	0.06	0.07	0.93	0.81	0.17	0.1	0.93	0.93	0.08	0.18	0.1	0.16	0.8	0.16	0.06	0.17	0.09	0.87	0.09	0.08	0.17	0.89
6	0.04	0.04	0.06	0.09	0.14	1	0.06	0.04	0.04	0.02	0.14	0.12	0.25	0.07	0.14	0.14	0.04	0.15	0.04	0.1	0.12	0.12	0.03	0.25	0.07	0.12	0.06	0.03	0.24	0.12
7	0.09	0.09	0.08	0.06	0.05	0.06	1	0.08	0.07	0.08	0.05	0.07	0.11	0.09	0.05	0.05	0.09	0.07	0.11	0.11	0.07	0.07	0.09	0.09	0.11	0.05	0.08	0.09	0.1	0.05
8	0.92	0.83	0.08	0.09	0.07	0.04	0.08	1	0.08	0.43	0.05	0.05	0.37	0.05	0.05	0.05	0.9	0	0.05	0.38	0.05	0.02	0.81	0.37	0.07	0.05	0.79	0.81	0.36	0.05
9	0.1	0.1	0.65	0.34	0.06	0.04	0.07	0.08	1	0.09	0.06	0.07	0.11	0.2	0.06	0.06	0.1	0.08	0.2	0.1	0.07	0.08	0.1	0.11	0.21	0.06	0.09	0.1	0.11	0.06
10	0.45	0.51	0.09	0.08	0.07	0.02	0.08	0.43	0.09	1	0.06	0.07	0.25	0.06	0.06	0.07	0.46	0.02	0.08	0.29	0.09	0.06	0.5	0.25	0.07	0.06	0.34	0.41	0.25	0.06
11	0.06	0.06	0.07	0.51	0.93	0.14	0.05	0.05	0.06	0.06	1	0.87	0.17	0.09	1	0.96	0.07	0.17	0.09	0.15	0.85	0.15	0.05	0.17	0.09	0.93	0.07	0.06	0.17	0.96
12	0.06	0.06	0.09	0.54	0.81	0.12	0.07	0.05	0.07	0.07	0.87	1	0.16	0.08	0.87	0.87	0.07	0.15	0.09	0.14	0.89	0.13	0.06	0.16	0.09	0.81	0.07	0.06	0.15	0.83
13	0.36	0.34	0.12	0.17	0.17	0.25	0.11	0.37	0.11	0.25	0.17	0.16	1	0.12	0.17	0.17	0.39	0.38	0.13	0.77	0.18	0.35	0.34	0.98	0.13	0.16	0.31	0.36	0.95	0.16
14	0.07	0.07	0.23	0.14	0.1	0.07	0.09	0.05	0.2	0.06	0.09	0.08	0.12	1	0.09	0.09	0.07	0.14	0.73	0.12	0.09	0.11	0.07	0.1	0.8	0.1	0.07	0.07	0.12	0.09
15	0.06	0.06	0.07	0.51	0.93	0.14	0.05	0.05	0.06	0.06	1	0.87	0.17	0.09	1	0.96	0.07	0.17	0.09	0.15	0.85	0.15	0.05	0.17	0.09	0.93	0.07	0.06	0.17	0.96
16	0.06	0.06	0.07	0.49	0.93	0.14	0.05	0.05	0.06	0.07	0.96	0.87	0.17	0.09	0.96	1	0.07	0.17	0.09	0.15	0.85	0.15	0.06	0.17	0.09	0.89	0.07	0.06	0.17	0.91
17	0.88	0.79	0.1	0.1	0.08	0.04	0.09	0.9	0.1	0.46	0.07	0.07	0.39	0.07	0.07	0.07	1	0.02	0.07	0.38	0.06	0.03	0.77	0.39	0.08	0.07	0.71	0.77	0.36	0.07
18	0.02	0.02	0.1	0.13	0.18	0.15	0.07	0	0.08	0.02	0.17	0.15	0.38	0.14	0.17	0.17	0.02	1	0.16	0.46	0.17	0.85	0.02	0.36	0.13	0.16	0.02	0.02	0.37	0.15
19	0.07	0.07	0.23	0.14	0.1	0.04	0.11	0.05	0.2	0.08	0.09	0.09	0.13	0.73	0.09	0.09	0.07	0.16	1	0.16	0.11	0.14	0.07	0.12	0.74	0.1	0.07	0.07	0.13	0.09
20	0.38	0.35	0.13	0.16	0.16	0.1	0.11	0.38	0.1	0.29	0.15	0.14	0.77	0.12	0.15	0.15	0.38	0.46	0.16	1	0.17	0.46	0.35	0.75	0.14	0.14	0.32	0.37	0.75	0.14
21	0.06	0.06	0.08	0.51	0.8	0.12	0.07	0.05	0.07	0.09	0.85	0.89	0.18	0.09	0.85	0.85	0.06	0.17	0.11	0.17	1	0.15	0.08	0.18	0.11	0.83	0.07	0.06	0.18	0.82
22	0.03	0.03	0.08	0.11	0.16	0.12	0.07	0.02	0.08	0.06	0.15	0.13	0.35	0.11	0.15	0.15	0.03	0.85	0.14	0.46	0.15	1	0.03	0.34	0.11	0.14	0.04	0.03	0.35	0.13
23	0.88	0.79	0.09	0.08	0.06	0.03	0.09	0.81	0.1	0.5	0.05	0.06	0.34	0.07	0.05	0.06	0.77	0.02	0.07	0.35	0.08	0.03	1	0.34	0.08	0.05	0.71	0.77	0.33	0.05
24	0.37	0.35	0.12	0.17	0.17	0.25	0.09	0.37	0.11	0.25	0.17	0.16	0.98	0.1	0.17	0.17	0.39	0.36	0.12	0.75	0.18	0.34	0.34	1	0.12	0.16	0.31	0.36	0.94	0.16
25	0.08	0.08	0.25	0.15	0.09	0.07	0.11	0.07	0.21	0.07	0.09	0.09	0.13	0.8	0.09	0.09	0.08	0.13	0.74	0.14	0.11	0.11	0.08	0.12	1	0.09	0.09	0.08	0.13	0.09
26	0.07	0.07	0.07	0.48	0.87	0.12	0.05	0.05	0.06	0.06	0.93	0.81	0.16	0.1	0.93	0.89	0.07	0.16	0.1	0.14	0.83	0.14	0.05	0.16	0.09	1	0.07	0.06	0.16	0.93
27	0.82	0.73	0.08	0.09	0.09	0.06	0.08	0.79	0.09	0.34	0.07	0.07	0.31	0.07	0.07	0.07	0.71	0.02	0.07	0.32	0.07	0.04	0.71	0.31	0.09	0.07	1	0.71	0.31	0.07
28	0.88	0.79	0.09	0.09	0.08	0.03	0.09	0.81	0.1	0.41	0.06	0.06	0.36	0.07	0.06	0.06	0.77	0.02	0.07	0.37	0.06	0.03	0.77	0.36	0.08	0.06	0.71	1	0.35	0.06
29	0.36	0.34	0.12	0.17	0.17	0.24	0.1	0.36	0.11	0.25	0.17	0.15	0.95	0.12	0.17	0.17	0.36	0.37	0.13	0.75	0.18	0.35	0.33	0.94	0.13	0.16	0.31	0.35	1	0.15
30	0.06	0.06	0.07	0.49	0.89	0.12	0.05	0.05	0.06	0.06	0.96	0.83	0.16	0.09	0.96	0.91	0.07	0.15	0.09	0.14	0.82	0.13	0.05	0.16	0.09	0.93	0.07	0.06	0.15	1

Rehab Kareem Kadthim *et al*, International Journal of Computer Science and Mobile Applications,
Vol.10 Issue. 9, September- 2022, pg. 61-71 ISSN: 2321-8363
Impact Factor: 6.308
(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

It's clear from these tables that the two mentioned groups (red, green) refer to the same students as indicated by cosine similarity. Noting that the percentage of similarity was less than that of the cosine similarity. The reason for this difference is that the Jaccard similarity does not take the word repetition into account. As is known, the Jaccard similarity depends on dividing the number of words that exist in both documents (intersection) by the number of words unions between them. Jaccard similarity is greatly affected by the size of data. Large data sets can have a significant impact on the index as they can increase union exponentially while keeping intersection similar. Therefore, existing of unnecessary sentences or some added stop words by the student will decrease the similarity. Which can be one of the weak points for Jaccard similarity in this type of applications.

4.4 Overlap Similarity Result

Tables (5),(6) show the results of the overlap similarity for each subject

Subject1:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	1	0.93	0.52	0.17	0.17	0.59	0.28	0.93	0.52	0.41	0.59	0.17	0.45	0.41	0.13	0.09	0.9	0.48	0.86	0.57	0.12	0.41	0.9	0.77	0.41	0.14	0.93	1	0.12	0.09
2	0.93	1	0.52	0.21	0.17	0.59	0.28	0.86	0.52	0.41	0.52	0.21	0.45	0.41	0.13	0.09	0.83	0.55	0.86	0.57	0.12	0.41	0.9	0.73	0.41	0.14	0.93	0.93	0.12	0.09
3	0.52	0.52	1	0.22	0.17	0.81	0.32	0.47	0.97	0.84	0.88	0.22	0.75	0.62	0.13	0.14	0.52	0.91	0.47	0.39	0.17	0.72	0.48	0.46	0.62	0.14	0.52	0.52	0.16	0.14
4	0.17	0.21	0.22	1	0.13	0.2	0.2	0.16	0.22	0.22	0.16	0.98	0.27	0.18	0.09	0.09	0.14	0.22	0.13	0.21	0.08	0.27	0.21	0.23	0.18	0.09	0.19	0.16	0.08	0.09
5	0.17	0.17	0.17	0.13	1	0.22	0.09	0.17	0.17	0.17	0.17	0.13	0.26	0.17	0.91	0.95	0.17	0.17	0.13	0.26	0.61	0.3	0.17	0.17	0.17	0.95	0.17	0.22	0.78	0.86
6	0.59	0.59	0.81	0.2	0.22	1	0.48	0.56	0.78	0.73	0.58	0.2	0.58	0.42	0.17	0.18	0.59	0.75	0.5	0.43	0.21	0.56	0.55	0.42	0.42	0.18	0.58	0.61	0.2	0.18
7	0.28	0.28	0.32	0.2	0.09	0.48	1	0.28	0.32	0.32	0.36	0.2	0.68	0.4	0.04	0.05	0.24	0.32	0.28	0.24	0.04	0.64	0.28	0.24	0.4	0.05	0.28	0.28	0.04	0.05
8	0.93	0.86	0.47	0.16	0.17	0.56	0.28	1	0.47	0.44	0.56	0.16	0.44	0.41	0.13	0.09	0.97	0.47	0.8	0.61	0.12	0.41	0.9	0.77	0.41	0.14	0.87	0.9	0.12	0.09
9	0.52	0.52	0.97	0.22	0.17	0.78	0.32	0.47	1	0.89	0.86	0.22	0.76	0.62	0.13	0.14	0.52	0.91	0.5	0.39	0.17	0.73	0.48	0.5	0.62	0.14	0.52	0.52	0.16	0.14
10	0.41	0.41	0.84	0.22	0.17	0.73	0.32	0.44	0.89	1	0.84	0.22	0.78	0.62	0.13	0.14	0.41	0.81	0.4	0.39	0.17	0.76	0.52	0.5	0.62	0.14	0.48	0.42	0.16	0.14
11	0.59	0.52	0.88	0.16	0.17	0.58	0.36	0.56	0.86	0.84	1	0.15	0.67	0.48	0.13	0.14	0.59	0.81	0.53	0.43	0.17	0.62	0.52	0.58	0.48	0.14	0.52	0.58	0.16	0.14
12	0.17	0.21	0.22	0.98	0.13	0.2	0.2	0.16	0.22	0.22	0.15	1	0.26	0.17	0.09	0.09	0.14	0.22	0.13	0.21	0.08	0.26	0.21	0.23	0.17	0.09	0.19	0.16	0.08	0.09
13	0.45	0.45	0.75	0.27	0.26	0.58	0.68	0.44	0.76	0.78	0.67	0.26	1	0.53	0.22	0.23	0.45	0.78	0.43	0.57	0.29	0.95	0.45	0.5	0.54	0.23	0.45	0.48	0.28	0.23
14	0.41	0.41	0.62	0.18	0.17	0.42	0.4	0.41	0.62	0.62	0.48	0.17	0.53	1	0.13	0.14	0.41	0.62	0.4	0.57	0.17	0.53	0.41	0.54	0.96	0.14	0.42	0.42	0.16	0.14
15	0.13	0.13	0.13	0.09	0.91	0.17	0.04	0.13	0.13	0.13	0.13	0.09	0.22	0.13	1	0.91	0.13	0.13	0.09	0.22	0.65	0.26	0.13	0.13	0.13	1	0.13	0.17	0.78	0.91
16	0.09	0.09	0.14	0.09	0.95	0.18	0.05	0.09	0.14	0.14	0.14	0.09	0.23	0.14	0.91	1	0.09	0.14	0.05	0.18	0.59	0.27	0.09	0.14	0.14	0.91	0.09	0.14	0.77	0.91
17	0.9	0.83	0.52	0.14	0.17	0.59	0.24	0.97	0.52	0.41	0.59	0.14	0.45	0.41	0.13	0.09	1	0.48	0.83	0.57	0.12	0.41	0.79	0.73	0.41	0.14	0.86	0.93	0.12	0.09
18	0.48	0.55	0.91	0.22	0.17	0.75	0.32	0.47	0.91	0.81	0.81	0.22	0.78	0.62	0.13	0.14	0.48	1	0.5	0.39	0.17	0.75	0.48	0.46	0.62	0.14	0.48	0.48	0.16	0.14
19	0.86	0.86	0.47	0.13	0.13	0.5	0.28	0.8	0.5	0.4	0.53	0.13	0.43	0.4	0.09	0.05	0.83	0.5	1	0.54	0.08	0.4	0.76	0.73	0.4	0.09	0.8	0.87	0.08	0.05
20	0.57	0.57	0.39	0.21	0.26	0.43	0.24	0.61	0.39	0.39	0.43	0.21	0.57	0.57	0.22	0.18	0.57	0.39	0.54	1	0.21	0.54	0.57	0.54	0.46	0.23	0.57	0.57	0.2	0.18
21	0.12	0.12	0.17	0.08	0.61	0.21	0.04	0.12	0.17	0.17	0.17	0.08	0.29	0.17	0.65	0.59	0.12	0.17	0.08	0.21	1	0.25	0.12	0.12	0.17	0.64	0.12	0.17	0.75	0.59
22	0.41	0.41	0.72	0.27	0.3	0.56	0.64	0.41	0.73	0.76	0.62	0.26	0.95	0.53	0.26	0.27	0.41	0.75	0.4	0.54	0.25	1	0.41	0.46	0.54	0.27	0.42	0.45	0.24	0.27
23	0.9	0.9	0.48	0.21	0.17	0.55	0.28	0.9	0.48	0.52	0.52	0.21	0.45	0.41	0.13	0.09	0.79	0.48	0.76	0.57	0.12	0.41	1	0.77	0.41	0.14	0.97	0.9	0.12	0.09
24	0.77	0.73	0.46	0.23	0.17	0.42	0.24	0.77	0.5	0.5	0.58	0.23	0.5	0.54	0.13	0.14	0.73	0.46	0.73	0.54	0.12	0.46	0.77	1	0.54	0.14	0.81	0.81	0.12	0.14
25	0.41	0.41	0.62	0.18	0.17	0.42	0.4	0.41	0.62	0.62	0.48	0.17	0.54	0.96	0.13	0.14	0.41	0.62	0.4	0.46	0.17	0.54	0.41	0.54	1	0.14	0.42	0.42	0.16	0.14
26	0.14	0.14	0.14	0.09	0.95	0.18	0.05	0.14	0.14	0.14	0.14	0.09	0.23	0.14	1	0.91	0.14	0.14	0.09	0.23	0.64	0.27	0.14	0.14	0.14	1	0.14	0.18	0.82	0.91
27	0.93	0.93	0.52	0.19	0.17	0.58	0.28	0.87	0.52	0.48	0.52	0.19	0.45	0.42	0.13	0.09	0.86	0.48	0.8	0.57	0.12	0.42	0.97	0.81	0.42	0.14	1	0.9	0.12	0.09
28	1	0.93	0.52	0.16	0.22	0.61	0.28	0.9	0.52	0.42	0.58	0.16	0.48	0.42	0.17	0.14	0.93	0.48	0.87	0.57	0.17	0.45	0.9	0.81	0.42	0.18	0.9	1	0.16	0.14
29	0.12	0.12	0.16	0.08	0.78	0.2	0.04	0.12	0.16	0.16	0.16	0.08	0.28	0.16	0.78	0.77	0.12	0.16	0.08	0.2	0.75	0.24	0.12	0.12	0.16	0.82	0.12	0.16	1	0.73

IJCSMA

Rehab Kareem Kadthim *et al*, International Journal of Computer Science and Mobile Applications,
Vol.10 Issue. 9, September- 2022, pg. 61-71 ISSN: 2321-8363
Impact Factor: 6.308
(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

Subject2:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	0.95	0.19	0.24	0.16	0.1	0.18	0.97	0.21	0.7	0.13	0.13	0.68	0.15	0.13
2	0.95	1	0.19	0.24	0.16	0.1	0.18	0.92	0.21	0.77	0.13	0.13	0.66	0.15	0.13
3	0.19	0.19	1	0.97	0.16	0.14	0.16	0.16	0.83	0.17	0.16	0.19	0.31	0.42	0.16
4	0.24	0.24	0.97	1	0.84	0.33	0.18	0.22	0.83	0.23	0.84	0.86	0.3	0.42	0.84
5	0.16	0.16	0.16	0.84	1	0.38	0.12	0.14	0.14	0.17	0.98	0.91	0.35	0.23	0.98
6	0.1	0.1	0.14	0.33	0.38	1	0.14	0.1	0.1	0.05	0.38	0.33	0.76	0.14	0.38
7	0.18	0.18	0.16	0.18	0.12	0.14	1	0.15	0.14	0.17	0.12	0.15	0.26	0.19	0.12
8	0.97	0.92	0.16	0.22	0.14	0.1	0.15	1	0.17	0.67	0.11	0.11	0.7	0.12	0.11
9	0.21	0.21	0.83	0.83	0.14	0.1	0.14	0.17	1	0.17	0.14	0.17	0.31	0.35	0.14
10	0.7	0.77	0.17	0.23	0.17	0.05	0.17	0.67	0.17	1	0.13	0.17	0.6	0.12	0.13
11	0.13	0.13	0.16	0.84	0.98	0.38	0.12	0.11	0.14	0.13	1	0.93	0.34	0.23	1
12	0.13	0.13	0.19	0.86	0.91	0.33	0.15	0.11	0.17	0.17	0.93	1	0.32	0.19	0.93
13	0.68	0.66	0.31	0.3	0.35	0.76	0.26	0.7	0.31	0.6	0.34	0.32	1	0.35	0.34
14	0.15	0.15	0.42	0.42	0.23	0.14	0.19	0.12	0.35	0.12	0.23	0.19	0.35	1	0.23
15	0.13	0.13	0.16	0.84	0.98	0.38	0.12	0.11	0.14	0.13	1	0.93	0.34	0.23	1
16	0.13	0.13	0.16	0.82	0.98	0.38	0.12	0.11	0.14	0.17	0.98	0.93	0.34	0.23	0.98
17	0.95	0.89	0.19	0.24	0.16	0.1	0.18	0.95	0.21	0.7	0.14	0.14	0.73	0.15	0.14
18	0.04	0.04	0.21	0.42	0.42	0.29	0.17	0	0.17	0.04	0.42	0.38	0.96	0.25	0.42
19	0.15	0.15	0.42	0.42	0.23	0.1	0.23	0.12	0.35	0.15	0.23	0.23	0.38	0.85	0.23
20	0.61	0.58	0.28	0.33	0.28	0.29	0.24	0.62	0.24	0.57	0.27	0.25	1	0.31	0.27
21	0.13	0.13	0.19	0.82	0.91	0.33	0.15	0.11	0.17	0.2	0.93	0.95	0.36	0.23	0.93
22	0.08	0.08	0.17	0.38	0.38	0.24	0.17	0.04	0.17	0.12	0.38	0.33	0.92	0.21	0.38
23	0.95	0.89	0.19	0.21	0.13	0.1	0.18	0.92	0.21	0.77	0.1	0.13	0.64	0.15	0.1
24	0.68	0.66	0.31	0.31	0.35	0.76	0.24	0.7	0.31	0.6	0.34	0.32	1	0.31	0.34
25	0.18	0.18	0.43	0.43	0.21	0.14	0.21	0.14	0.36	0.14	0.21	0.21	0.36	0.92	0.21
26	0.13	0.13	0.16	0.81	0.93	0.33	0.12	0.11	0.14	0.13	0.98	0.91	0.33	0.23	0.98
27	0.97	0.91	0.16	0.24	0.18	0.14	0.15	0.94	0.17	0.53	0.15	0.15	0.67	0.15	0.15
28	0.95	0.89	0.19	0.23	0.15	0.1	0.18	0.92	0.21	0.67	0.13	0.13	0.67	0.15	0.13
29	0.68	0.66	0.31	0.3	0.35	0.76	0.26	0.7	0.31	0.6	0.34	0.32	0.98	0.35	0.34

	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	0.13	0.95	0.04	0.15	0.61	0.13	0.08	0.95	0.68	0.18	0.13	0.97	0.95	0.68	0.13
2	0.13	0.89	0.04	0.15	0.58	0.13	0.08	0.89	0.66	0.18	0.13	0.91	0.89	0.66	0.13
3	0.16	0.19	0.21	0.42	0.28	0.19	0.17	0.19	0.31	0.43	0.16	0.16	0.19	0.31	0.16
4	0.82	0.24	0.42	0.42	0.33	0.82	0.38	0.21	0.31	0.43	0.81	0.24	0.23	0.3	0.82
5	0.98	0.16	0.42	0.23	0.28	0.91	0.38	0.13	0.35	0.21	0.93	0.18	0.15	0.35	0.95
6	0.38	0.1	0.29	0.1	0.29	0.33	0.24	0.1	0.76	0.14	0.33	0.14	0.1	0.76	0.33
7	0.12	0.18	0.17	0.23	0.24	0.15	0.17	0.18	0.24	0.21	0.12	0.15	0.18	0.26	0.12
8	0.11	0.95	0.08	0.12	0.62	0.11	0.04	0.92	0.7	0.14	0.11	0.94	0.92	0.7	0.11
9	0.14	0.21	0.17	0.35	0.24	0.17	0.17	0.21	0.31	0.36	0.14	0.17	0.21	0.31	0.14
10	0.17	0.7	0.04	0.15	0.57	0.2	0.12	0.77	0.6	0.14	0.13	0.53	0.67	0.6	0.13
11	0.98	0.14	0.42	0.23	0.27	0.93	0.38	0.1	0.34	0.21	0.98	0.15	0.13	0.34	0.98
12	0.93	0.14	0.38	0.23	0.25	0.95	0.33	0.13	0.32	0.21	0.91	0.15	0.13	0.32	0.91
13	0.34	0.73	0.96	0.38	1	0.36	0.92	0.64	1	0.36	0.33	0.67	0.67	0.98	0.32
14	0.23	0.15	0.25	0.85	0.31	0.23	0.21	0.15	0.31	0.92	0.23	0.15	0.15	0.35	0.23
15	0.98	0.14	0.42	0.23	0.27	0.93	0.38	0.1	0.34	0.21	0.98	0.15	0.13	0.34	0.98
16	1	0.14	0.42	0.23	0.27	0.93	0.38	0.13	0.34	0.21	0.95	0.15	0.13	0.34	0.95
17	0.14	1	0.04	0.15	0.62	0.14	0.08	0.89	0.73	0.18	0.14	0.88	0.89	0.7	0.14
18	0.42	0.04	1	0.29	0.92	0.42	0.92	0.04	0.92	0.25	0.38	0.04	0.04	0.96	0.38
19	0.23	0.15	0.29	1	0.38	0.27	0.25	0.15	0.35	0.88	0.23	0.15	0.15	0.38	0.23
20	0.27	0.62	0.92	0.38	1	0.29	0.92	0.56	0.98	0.32	0.26	0.58	0.59	1	0.25
21	0.93	0.14	0.42	0.27	0.29	1	0.38	0.15	0.36	0.25	0.93	0.15	0.13	0.36	0.91
22	0.38	0.08	0.92	0.25	0.92	0.38	1	0.08	0.88	0.21	0.33	0.08	0.08	0.92	0.33
23	0.13	0.89	0.04	0.15	0.56	0.15	0.08	1	0.64	0.18	0.1	0.91	0.87	0.64	0.1
24	0.34	0.73	0.92	0.36	0.98	0.36	0.88	0.64	1	0.32	0.33	0.67	0.67	0.98	0.32
25	0.21	0.18	0.25	0.88	0.32	0.25	0.21	0.18	0.32	1	0.21	0.18	0.18	0.36	0.21
26	0.95	0.14	0.38	0.23	0.26	0.93	0.33	0.1	0.33	0.21	1	0.15	0.13	0.33	0.98
27	0.15	0.88	0.04	0.15	0.58	0.15	0.08	0.91	0.67	0.18	0.15	1	0.91	0.67	0.15
28	0.13	0.89	0.04	0.15	0.59	0.13	0.08	0.87	0.67	0.18	0.13	0.91	1	0.67	0.13
29	0.34	0.7	0.96	0.38	1	0.36	0.92	0.64	0.98	0.36	0.33	0.67	0.67	1	0.32

As it's clear from the tables the sam`e students who were identified by the cosine and Jaccard methods were identified by the overlap method. Also, it's shown that overlap similarity is closer to Jaccard than the cosine in terms of results. The only different between overlap and Jaccard is the calculating of the denominator, while the numerator is same. The length of the smaller text is taken from the two compared texts to be a denominator. For this reason, the effect of document size of overlap is more than Jaccard. For example, suppose there are two documents both of size (100) terms and there are 50 terms share between them, when we change the document size as follows.

5. Conclusion

The interest of educational institutions in e-learning and its application has increased, and students are directed to rely entirely on the practice of distance learning during Covid19. This methodology has been applied in the whole world due to the flexibility, accessibility, and ease of use of e-learning, but the main challenge remains in education via the Internet. It is the evaluation in exams because the phenomenon of cheating has become increasing along with the technological development taking place in education and it is considered an important issue. In this paper, a solution is proposed to discover cheating by revealing the similarity in the text answers provided by the student to a specific type of question. 1. The number of essay questions on online platforms and the number of similar answers are important features for identifying potential cheating cases. The model was distinguished by detecting a high rate of cheating in the essay questions because most students cheat with essay questions. The cosine similarity is the best method in the first model because it is not affected by the extra sentences put in by the students. Cosine similarity required more operations than Jaccard and overlap.

REFERENCES

[1] A. Moubayed, M. Injadat, A. B. Nassif, H. Lutfiyya, and A. Shami, "E-Learning: Challenges and Research Opportunities Using Machine Learning Data Analytics," IEEE Access, vol. 6, pp. 39117–39138, 2018.

[2] K. Andersen, S. E. Thorsteinsson, H. Thorbergsson, and K. S. Gudmundsson, "Adapting engineering examinations from paper to online," in IEEE Global Engineering Education Conference, EDUCON, 2020, vol. 2020-April, pp. 1891–1895, 2020.

[3] O. Holden, V. Kuhlmeier, and M. Norris, "Academic integrity in online testing: A research review", 2021.

[4] K. Curran, G. Middleton, and C. Doherty, "Cheating in exams with technology," 2011. *Int. J. Cyber Ethics Educ.*, vol. 1, no. 2, pp. 54–62. 2011.

[5] A. M. Duhaim, S. O. Al-Mamory, and M. S. Mahdi, "Cheating detection in online exams during Covid-19 pandemic using data mining techniques" 2021.

[6] E. R. Cavalcanti, C. E. Pires, E. P. Cavalcanti, and V. F. Pires "Detection and evaluation of cheating on college exams using supervised classification," 2012. *Informatics Educ.*, vol. 11, no. 2, pp. 169–190.

[7] L. Li, C. Qiu, and A. Fang, "Detection of test cheating by outlier analysis approach using probability estimation" 2019. *PervasiveHealth Pervasive Comput. Technol. Healthc*.

[8] G. Jain, T. Mahara, and K. N. Tripathi, "A Survey of Similarity Measures for Collaborative Filtering-Based Recommender System," in *Advances in Intelligent Systems and Computing*, vol. 1053, pp. 343–352, Springer, 2020.

[9] K. P. Reddy, T. R. Reddy, G. A. Naidu, and B. Vishnu, "Impact of Similarity Measures in Information Retrieval," *Int. J. Comput. Eng. Res.*, vol.08, no.6, pp. 54–59, 2018.

[10] M. Afzali and S. Kumar, "An Extensive Study of Similarity and Dissimilarity Measures Used for Text Document Clustering using K-means Algorithm," *Int.J. Inf. Technol. Comput. Sci.*, vol. 10, no. 9, pp. 64–73, 2018.

[11] W. H.Gomaa and A. A. Fahmy, "A Survey of Text Similarity Approaches," *Int. J. Comput. Appl.*, vol. 68, no. 13, pp. 13–18, 2013.

[12] V. M.K and K. K, "A Survey on Similarity Measures in Text Mining," *Mach. Learn. Appl. An Int. J.*, vol. 3, no. 1, pp. 19-28, 2016.