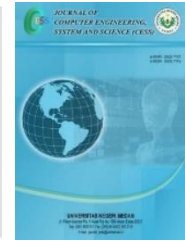


Contents list available at [www.jurnal.unimed.ac.id](http://www.jurnal.unimed.ac.id)

**CESS**  
**(Journal of Computing Engineering, System and Science)**

journal homepage: <https://jurnal.unimed.ac.id/2012/index.php/cess>



---

## **Metode Cosine Similarity pada Penilaian Essay Perkaderan Ikatan Mahasiswa Muhammadiyah**

### ***Cosine Similarity Method in Muhammadiyah Student Association Cadre Essay Assessment***

**Novi Kusuma Anggraini<sup>1</sup>, Uky Yudatama<sup>2</sup>, Maimunah<sup>3\*</sup>**

<sup>1,2,3</sup>Universitas Muhammadiyah Magelang, Jawa Tengah, Indonesia

email: <sup>1</sup>[1805040034@student.ummql.ac.id](mailto:1805040034@student.ummql.ac.id), <sup>2</sup>[uky@ummql.ac.id](mailto:uky@ummql.ac.id), <sup>3</sup>[maimunah@unimma.ac.id](mailto:maimunah@unimma.ac.id)

---

#### **ABSTRAK**

Perkaderan persyarikatan merupakan suatu program yang dilaksanakan oleh Muhammadiyah baik ditingkat pusat maupun bawahnya dan menjadi bagian inti dalam perjalanan organisasi Muhammadiyah. Perkaderan utama di Muhammadiyah dibagi menjadi 2 (dua) yaitu Darul Arqam dan Baitul Arqam. Dalam pelaksanaannya juga terdapat evaluasi pembelajaran yaitu *pre-test* dan *post-test*. *Pre-Test* dilaksanakan sebelum materi disampaikan, sedangkan *post-test* dilaksanakan setelah materi disampaikan. Evaluasi ini berbentuk soal esai yang bersangkutan pautan dengan materi yang akan diberikan. Jawaban peserta yang bervariasi dan waktu yang padat menjadi tantangan tersendiri untuk mengoreksinya, dan membutuhkan waktu yang lebih banyak lagi ditambah dengan harus memasukkan nilai ke sistem perkaderan secara manual. Penelitian ini menerapkan *text mining* dengan menggunakan metode *cosine similarity* untuk melakukan penilaian otomatis pada soal esai *pre-test* dan *post-test*. Hasil perhitungan pada soal *pre-test* mendapatkan hasil tertinggi dengan tingkat persamaan 77,46% oleh peserta 1 pada soal 4 dan peserta 2 pada soal 3, dan *post-test* dengan hasil tertinggi 100% oleh peserta 1 pada soal 4. Sedangkan untuk hasil 0,0 yang diperoleh beberapa peserta dikarenakan *term* antar kunci jawaban dan jawaban peserta tidak memiliki persamaan.

**Kata Kunci:** *Perkaderan, Pre Test, Post Test, Text Mining, Cosine Similarity.*

---

#### **ABSTRACT**

Persyarikatan cadre is a program implemented by Muhammadiyah both at the central and lower levels and is a core part of the journey of the Muhammadiyah organization. The main cadres in Muhammadiyah are divided into 2 (two), namely Darul Arqam and Baitul Arqam. In its implementation there is also a learning evaluation, namely *pre-test* and *post-test*. The *pre-test* is carried out before the material is delivered, while the *post-test* is carried out after

\*Penulis Korespondensi:

email: [maimunah@unimma.ac.id](mailto:maimunah@unimma.ac.id)

the material is delivered. This evaluation takes the form of essay questions related to the material to be given. Participants' varied answers and time constraints make it a challenge to correct them and require more time plus having to enter scores into the cadre system manually. This research applies text mining by using the method cosine similarity to carry out automatic assessments on pre-test and post-test essay questions. Calculation results in the problem pre-test get the highest results with a similarity level of 77.46% by participant 1 on question 4 and participant 2 on question 3, and post-test with the highest result of 100% by participant 1 on question 4. Meanwhile, the result of 0.0 obtained by several participants was due to term There are no similarities between the answer key and the participants' answers.

**Keywords:** *Cadre, Pre Test, Post Test, Text Mining, Cosine Similarity.*

---

## 1. PENDAHULUAN

Perkaderan merupakan upaya merumuskan dan mengimplementasikan strategi, langkah, dan bentuk untuk membina kader [1]. Perkaderan persyarikatan merupakan suatu program yang dilaksanakan oleh Muhammadiyah baik ditingkat pusat maupun bawahnya dan menjadi bagian inti dalam perjalanan organisasi Muhammadiyah. Perkaderan utama di Muhammadiyah dibagi menjadi 2 (dua) yaitu Darul Arqam dan Baitul Arqam. Sama halnya dengan pendidikan dibangku kuliah, dalam proses perkaderan juga terdapat evaluasi proses pembelajaran yang bertujuan untuk mengetahui tingkat penguasaan materi yang disampaikan dan keberhasilan dalam penyampaian materi. Evaluasi pembelajaran dalam perkaderan ini dibagi menjadi 2 (dua) yaitu *pre-test* dan *post-test*. *Pre-test* ialah tes yang dilaksanakan sebelum materi diberikan dengan tujuan untuk mengetahui sejauh mana peserta mengetahui materi yang akan diberikan dan hasil *pre-test* ini dapat dijadikan acuan dalam penentuan metode pembelajaran. *Post test* ialah tes yang dilaksanakan setelah semua materi diberikan, dengan tujuan untuk mengetahui sejauh mana peserta menguasai dan memahami materi yang diberikan serta menjadi rangkaian penutup kegiatan perkaderan [2].

Bentuk soal dari *pre-test* dan *post-test* itu sendiri ialah esai yang mana soal yang diberikan berisi tentang materi yang akan disampaikan. Esai merupakan ujian dalam bentuk pertanyaan yang bertujuan agar jawaban yang diberikan berupa uraian berdasarkan pemahaman peserta ujian itu sendiri terhadap pertanyaan yang diberikan sesuai dengan apa yang diketahuinya [3]. Jawaban dari tes tersebut akan dikoreksi satu persatu yang kemudian dibandingkan dengan kunci jawaban lalu diberi skor. Adapun cara penilaian pada jawaban tes esai ialah dengan pembobotan, yang jika jawaban yang mendekati dengan makna pada kunci jawaban tidak bernilai salah dan diberi nilai dengan bobot yang sudah ditentukan. Begitu pula dengan jawaban salah dan jawaban kosong. Penilaian terhadap peserta pada pelaksanaan perkaderan sangatlah penting, karena hal ini dilakukan untuk mengevaluasi pemahaman peserta terhadap materi yang akan disampaikan dan sudah disampaikan.

Banyaknya peserta dan jumlah soal yang dikoreksi secara manual, maka hal tersebut membutuhkan waktu yang cukup panjang, belum lagi apabila terdapat tulisan peserta yang tidak dapat terbaca. Bila dilihat lagi, dengan jadwal perkaderan dan tugas instruktur yang padat dapat menyebabkan kualitas penilaian turun, dan harus menginputkan nilai yang telah dihitung ke sistem perkaderan secara manual dan satu-persatu. Dan juga teruntuk *post-test* yang berada diakhir waktu, menyebabkan tidak dapat ternilai dan tidak masuk dapat

terinput dalam sistem perkaderan. Hal tersebut yang melatarbelakangi penelitian ini dilakukan serta bagaimana memanfaatkan teknologi dalam melakukan penilaian *pre-test* dan *post-test* secara otomatis untuk memeriksa tingkat kemiripan jawaban dengan kunci jawaban. Untuk melakukan penilaian otomatis dan mengoreksi jawaban peserta, dapat memanfaatkan text mining. *Text mining* merupakan proses untuk menganalisis, mengekstrak informasi yang berguna untuk tujuan tertentu, *text mining* juga salah satu dari cabang ilmu data mining dan biasanya digunakan dalam pencocokan data, deteksi plagiarisme, penilaian dan otomatis jawaban esai [4]. Terdapat beberapa metode *text mining* diantaranya, *jaccard similarity*, *euclidean distance*, *cosine similarity* dan sebagainya. Dalam penelitian ini untuk melakukan pengecekan jawaban dan penilaian otomatis ialah menggunakan metode *cosine similarity*.

*Cosine similarity* merupakan metode yang digunakan untuk menghitung tingkat kemiripan antar dua buah objek [5]. Perhitungan metode ini berlandaskan pada dua vektor yang mempunyai kemiripan kata pada objek yang dibandingkan. Penerapan metode ini biasanya digunakan untuk mengukur tingkat kemiripan data yang berupa dokumen ataupun gambar [6]. Proses yang dilakukan ialah dengan membandingkan jawaban dari peserta dan kunci jawaban untuk mengukur tingkat kemiripan. Menurut Indra Mawanta, *cosine similarity* ialah metode yang dapat melakukan pendeteksian secara cepat dan tepat [7].

Penelitian terdahulu, melakukan analisis metode *similarity* pada aplikasi ujian online esai mendapatkan hasil dari pengujian diperoleh nilai rata-rata *precision* ialah 0,93 atau 93%, *recall* ialah 0,86 atau 86%, dan *f-measure* ialah 0,89 atau 89% untuk mendekati 1 atau 100% [8]. Hasil perhitungan *cosine similarity* pada penelitian yang dilakukan oleh Daniel, kemiripan dengan nilai tertinggi mencapai angka 90,58%. Berdasarkan penelitian ini menunjukkan bahwa metode *cosine similarity* dapat diterapkan dalam melakukan penilaian soal esai secara otomatis [3]. Metode *cosine similarity* juga dapat diterapkan dalam mendeteksi kemiripan gambar seperti yang dilakukan oleh Niko Surya dalam penelitiannya untuk mengetahui tingkat prosentase kemiripan gambar dan diperoleh hasil tingkat kemiripan sebesar 83% [9].

Dalam penelitian ini melakukan penerapan metode *cosine similarity* dalam melakukan penilaian otomatis pada jawaban *pre-test* dan *post-test* perkaderan. Penelitian ini memfokuskan pada penerapan dalam melakukan pengecekan jawaban sampai dengan penilaian akhir. Hasil dari penelitian ini diharapkan dapat menjadi solusi dalam pelaksanaan perkaderan terutama dalam pelaksanaan *pre-test* dan *post-test* sehingga memudahkan pekerjaan instruktur untuk menilai hasil dalam waktu yang singkat dan langsung terinput dalam sistem perkaderan.

## 2. DASAR/TINJAUAN TEORI

### 2.1 Text Mining

*Text mining* merupakan proses menyaring informasi yang berawal dari data mentah ke data yang lebih inti menggunakan teknik komputasi. *Text mining* terdiri dari beberapa tahapan untuk memproses data dari data yang tidak terstruktur, dengan tujuan untuk menemukan pola, konsep, serta atribut lainnya dalam data sehingga dapat dilakukan analisa antar dokumen. *Text mining* juga termasuk dalam aspek pemrosesan NLP (*Natural Language Processing*) yang berfokus pada analisa informasi berupa teks. Hal ini dikarenakan dalam penerapannya, teknik yang digunakan dalam *text mining* melibatkan pemrosesan bahasa

dan konsep yang terdapat dalam NLP. *Text mining* biasa diterapkan pada pemrosesan klasifikasi dokumen dalam jumlah yang banyak secara otomatis dan terkomputerisasi, karena jika dokumen diproses secara manual akan memerlukan waktu yang lama [10].

## 2.2 Preprocessing

Tahap *preprocessing* merupakan tahap awal dalam menyiapkan data yang akan diolah lebih lanjut. *Preprocessing* ialah proses mengolah data dengan teknik tertentu yang bertujuan untuk mengubah teks menjadi numerik. Tahap ini merupakan tahap yang penting, karena jika tanpa adanya tahap preprocessing, hasil dari kemiripan tidak akurat karena kata umum tidak diolah atau dihilangkan [11]. Terdapat beberapa tahapan dalam *preprocessing* yang diterapkan dalam penelitian ini, diantaranya *lowercasing*, *stopwords removal*, *tokenizing*, dan *stemming*.

### a. Lowercasing

Mengubah semua huruf dalam menjadi huruf kecil.

### b. Stopwords Removal

Menghapus kata umum dan tanda baca yang tidak terlalu berpengaruh dalam kalimat tetapi sering muncul dalam kalimat, seperti kata “yang”, “atau”.

### c. Tokenizing

Memisahkan kata yang tersusun dalam kalimat menjadi per-kata.

### d. Stemming

Mengubah semua kata menjadi bentuk dasar.

## 2.3 Cosine Similarity

*Cosine Similarity* merupakan algoritma yang digunakan untuk mengukur tingkat kemiripan nilai sudut cosinus yang didapat dari perkalian dua vektor yang dibanding. Vektor dinyatakan mirip apabila sudut cosinus bernilai 1, nilai cosinus  $0^\circ$  adalah 1 dan bernilai kurang dari 1 untuk nilai sudut yang lain [12]. Metode *cosine similarity* dapat melakukan pendeteksian secara cepat dan tepat, sebelum perhitungan kemiripan, terdapat beberapa proses *preprocessing text mining* yang harus dilakukan terlebih dahulu, kemudian term dari masing-masing kalimat jawaban peserta akan dilakukan pembobotan TF (*term frekuensi*), setelah itu dilakukan perhitungan jarak cosinus antar term menggunakan cosine similarity [13]. Perhitungan persamaan *cosine similarity* [8].

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2 \sum_{i=1}^n B_i^2}}$$

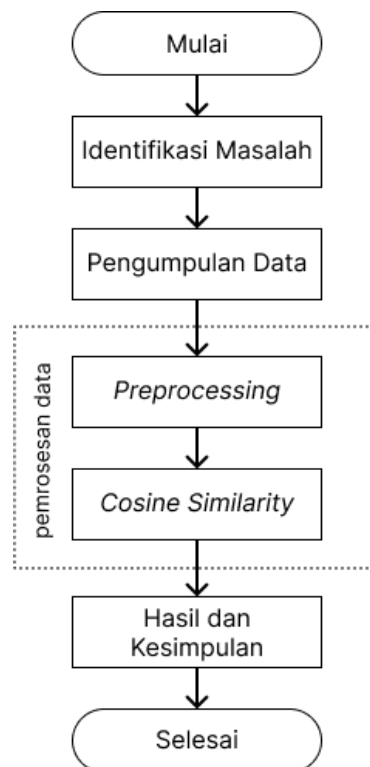
Keterangan:

- $A$  : Vektor A, yang akan dibandingkan kemiripannya
- $B$  : Vektor B, yang akan dibandingkan kemiripannya
- $\|A\|$  : Panjang vektor A
- $\|B\|$  : Panjang vektor B
- $i$  : Jumlah term dalam kalimat
- $n$  : Jumlah vektor

## 3. METODE

### 3.1 Tahapan Penelitian

Penelitian ini melalui beberapa alur, seperti yang tertera dalam flowchart dibawah ini.



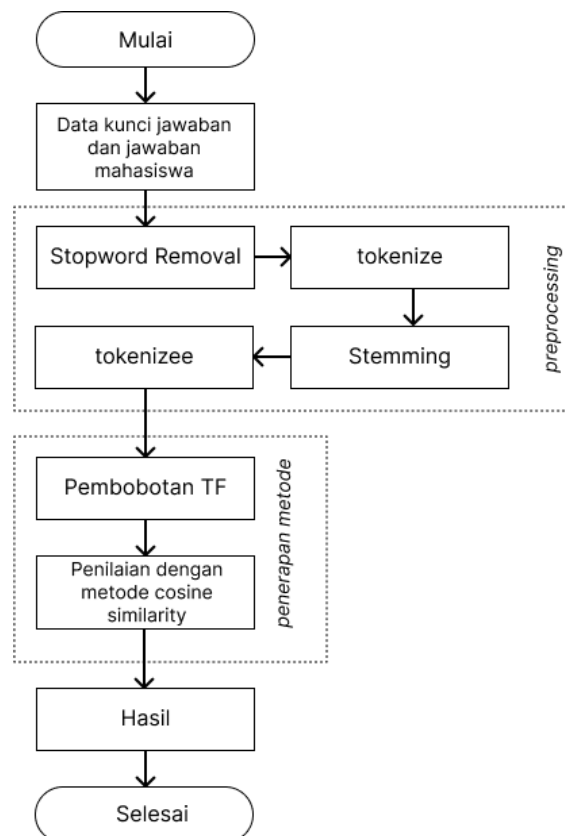
Gambar 1. Flowchart Penelitian

Dalam melakukan penelitian ini, penulis melalui beberapa alur. Tahapan penelitian ini dimulai dengan mengidentifikasi masalah dimana hal ini menjadi faktor utama dengan mengamati permasalahan yang terjadi yaitu banyaknya lembar jawab yang harus dinilai manual sehingga dapat menyebabkan kualitas dalam penilaian turun, ketersediaan waktu yang sangat padat sehingga soal *post-test* tidak dapat ternilai. Langkah berikutnya adalah pengumpulan data yang akan digunakan untuk melakukan penelitian implementasi metode cosine similarity. Data yang akan digunakan dalam penelitian ini adalah data soal dan jawaban pre-test post-test kegiatan perkaderan DAD (Darul Arqam Dasar) IMM Teknik.

Setelah data terkumpul, penelitian masuk pada tahap pemrosesan data, yang mana disini terdapat dua langkah dalam pemrosesan data yaitu tahap *preprocessing* dan penerapan metode cosine similarity. Pada tahap *preprocessing*, data yang sudah terkumpul akan diolah menggunakan metode yang sesuai untuk mengenali pola yang signifikan. Hasil dari penelitian ini disajikan dalam bab pembahasan yang mana memberikan gambaran tentang hasil penelitian yang dilakukan. Selanjutnya ialah tahap kesimpulan yang akan merangkum seluruh hasil penelitian dan memberikan pandangan dan saran terhadap penelitian selanjutnya sebagai landasan untuk perbaikan, pengembangan serta penelitian dimasa mendatang.

### 3.2 Pengolahan Data

*Cosine similarity* merupakan metode yang digunakan dalam berbagai bidang, salah satunya yaitu bidang pengolahan bahasa alami (*Natural Language Processing, NLP*) yang digunakan untuk mengukur kemiripan antara kalimat atau dokumen. Pada penelitian ini, tahapan yang dilalui dalam penerapan metode *cosine similarity* adalah sebagai berikut.



Gambar 2. Flowchart pengolahan data (preprocessing dan metode cosine similarity)

Setelah proses pengumpulan data, maka langkah yang dilakukan pertama kali adalah melakukan *preprocessing*. Pada proses *preprocessing* melalui beberapa tahapan seperti *stopword removal*, *tokenize*, dan *stemming*. Setelah itu, data akan lanjut diolah melalui proses pembobotan *Term Frequency* (TF) dimana proses ini menghitung jumlah kemunculan kata dalam tiap kalimat. Langkah selanjutnya adalah penerapan metode *cosine similarity* yang bertujuan untuk mengukur sejauh mana tingkat kesamaan antar kata pada kalimat dalam jawaban dan kunci jawaban. Setelah proses perhitungan *cosine similarity*, langkah selanjutnya adalah proses penilaian dari hasil yang didapatkan pada perhitungan sebelumnya. Hasil penilaian tersebut akan disajikan dan dianalisis pada bab hasil dan pembahasan. Hal tersebut akan memberikan hasil apakah metode yang digunakan sesuai atau tidak. Sehingga hasil penelitian ini menghasilkan analisis dalam mencapai tujuan penelitian.

## 4. HASIL DAN PEMBAHASAN

### 4.1 Tahap Pengumpulan Data

Penerapan *Cosine Similarity* dimulai dari persiapan data yang akan diolah, data tersebut merupakan data kunci jawaban dari soal pre-test peserta perkaderan IMM yaitu DAD (Darul Arqam Dasar) IMM Teknik. Berikut adalah tabel data soal pre-test beserta kunci jawaban. *Pre-test* dan *post-test* memiliki soal dan kunci jawaban yang sama.

Tabel 1. Soal dan kunci jawaban *pre-test* dan *post test*

No	Pertanyaan	Jawaban	Skor
1	Apa yang dimaksud dengan Islam?	Agama Islam adalah agama yang diwahyukan Allah kepada umat manusia melalui Nabi Muhammad yang bersumber pada Al-Quran dan Sunnah (Hadits) sebagai pedoman hidup.	5
2	Apa yang dimaksud dengan syahadat?	Syahadat merupakan kesaksian bahwa Allah tidak ada Tuhan yang wajib disembah selain Allah, dan Nabi Muhammad adalah utusan Allah. Syahadat merupakan rukun islam yang pertama	5
3	Apa yang kalian ketahui tentang Muhammadiyah?	Muhammadiyah merupakan organisasi islam yang didirikan oleh Ahmad Dahlan di Kauman, Yogyakarta tanggal 18 November 1912.	5
4	Sebutkan nilai dasar IMM	Keagamaan, kemahasiswaan, kemasyarakatan	5
5	Sebutkan 3 pendiri IMM	Djazman Al-Kindi, Rosyad Sholeh, Amien Rais, Abdul Hadi, Marzuki Usman, Rosyad Saleh	5
6	Apa tujuan IMM	Terbentuknya akademisi islam yang berakhlak mulia dalam rangka mencapai tujuan Muhammadiyah	5

Berikut data jawaban *pre-test* peserta perkaderan. Data yang digunakan adalah 6 jawaban dari soal *pre-test* dengan 5 mahasiswa yang telah mengerjakan soal tersebut pada pelaksanaan perkaderan. Berikut adalah tabel jawaban peserta yang akan diproses yang akan digunakan untuk penelitian penilaian otomatis.

Tabel 2. Data jawaban *pre-test* soal nomor 1

Peserta	Jawaban
Peserta1	Islam adalah agama yang diciptakan Allah berlandaskan Al-Quran dan Nabi Muhammad SAW yang dijadikan landasan/pedoman hidup manusia hingga akhir jaman.
Peserta2	Islam adalah agama terakhir yang diturunkan oleh Allah melalui Nabi Muhammad dan berkitab al-Quran
Peserta3	Islam adalah agama yang paling benar dimana ajarannya bukan hanya diperuntukkan satu umat tetapi juga umat yang lain.
Peserta4	Islam artinya suatu keadaan atau keyakinan bahwa ia berserah diri atas segala sesuatu kepada yang maha kuasa Allah
Peserta5	Agama yang menjadi pedoman manusia

Tabel 3. Data jawaban *pre-test* soal nomor 2

Peserta	Jawaban
Peserta1	Syahadat adalah kalimat yang meyakini bahwa Tiada Tuhan selain Allah dan Nabi Muhammad adalah utusan Allah
Peserta2	Syahadat salah satu rukun islam
Peserta3	Syahadat adalah kalimat tauhid yang mana menjadi kunci seseorang masuk surga dan masuk agama islam. Dengan syahadat kita yakin bahwa tidak ada Tuhan selain Allah dan Rasulullah adalah utusan Allah

Peserta4	Syahadat adalah kalimat yang berarti dan menyatakan bahwa Tiada Tuhan selain Allah dan Muhammad adalah utusannNya
Peserta5	Dua kalimat yang berisi kesaksian bahwa tiada tuhan selain Allah dan Muhammad adalah Nabi dan rasul kita

Tabel 4. Data jawaban *pre-test* soal nomor 3

Peserta	Jawaban
Peserta1	Muhammadiyah adalah pengikut Nabi muhammad pendirinya KH Ahmad Dahlan pada tanggal 18 November 1912
Peserta2	Muhammadiyah didirikan oleh Ahmad Dahlan di Yogyakarta pada tanggal 18 November 1912
Peserta3	Muhammadiyah didirikan di Yogyakarta, pendiri Kyai Haji Ahmad Dahlan, organisasi yang bergerak dalam bidang keagamaan, sosial budaya, dan politik
Peserta4	Muhammadiyah adalah sebuah organisasi dakwah islam yang didirikan di Yogyakarta oleh K.H Ahmad Dahlan
Peserta5	Organisasi keislaman, dibentuk dikauman Yogyakarta oleh Ahmad Dahlan

Tabel 5. Data jawaban *pre-test* nomor 4

Peserta	Jawaban
Peserta1	Kemahasiswaan, kemasyarakatan, religiusitas, intelektualitas, keagamaan
Peserta2	Keagamaan, Sosial, dan pendidikan
Peserta3	Menjunjung tinggi nilai keagamaan, menjunjung tinggi nilai pendidikan berdasar Al-Quran dan hadits, membentuk perkaderan Muhammadiyah
Peserta4	Dakwah islam
Peserta5	Pendidikan, sosial, keagamaan

Tabel 6. Data jawaban *pre-test* nomor 5

Peserta	Jawaban
Peserta1	Drs. Djazman, K.H Ahmad Dahlan, Siti Waliah
Peserta2	K H Ahmad Dahlan
Peserta3	KH Ahmad Dahlan, Amin Rais, Zaman Al Kindi
Peserta4	Belum Tau
Peserta5	Immawan, Immawati

Tabel 7. Data jawaban *pre-test* nomor 6

Peserta	Jawaban
Peserta1	Menciptakan Kader IMM yang berakhlak mulia untuk mencapai tujuan Muhammadiyah
Peserta2	Membentuk kader Muhammadiyah yang intelek
Peserta3	Mencetak perkaderan Muhammadiyah yang kreatif, inspiratif dan inovatif
Peserta4	Berdakwah juga berkomunikasi pada budaya, masyarakat, dan juga pendidikan
Peserta5	Pendidikan, sosial, keagamaan



## 4.2. Preprocessing

Tahap selanjutnya setelah pengumpulan data adalah melakukan *preprocessing*. *Preprocessing* merupakan langkah yang dilakukan pada data teks mentah untuk dipersiapkan pada pemrosesan selanjutnya agar lebih mempermudah dalam pemrosesan data. Dalam penelitian ini menggunakan tahap *preprocessing lower casing dan stopwords removal, stemming, dan tokenizing*.

### 4.2.1. Lower Casing dan Stopword Removal, Stemming, dan Tokenizing

*Lowercasing* adalah proses merubah huruf kapital dalam dokumen menjadi huruf kecil[3]. *Stopword removal* adalah menghapus kata yang memiliki frekuensi kemunculan tinggi pada dokumen namun tidak memiliki nilai yang tinggi [14]. *Stemming* adalah merubah semua kata dalam dokumen menjadi kata dasar[14]. *Tokenizing* adalah proses pemisahan kalimat menjadi potongan kata yang biasa disebut token yang kemudian akan dianalisa [15]. Pada tahap ini data kunci jawaban dan jawaban peserta akan menghasilkan data seperti pada tabel berikut.

Tabel 8. Hasil *Preprocessing* soal 1

Peserta	Hasil <i>Preprocessing</i> soal 1
Kunci	agama islam adalah agama yang wahyu Allah kepada umat manusia melalui Nabi muhammad yang bersumber pada al Quran dan sunnah hadits bagai pedoman hidup
Peserta1	islam adalah agama yang cipta Allah landas al quran dan Nabi muhammad saw yang jadi landas pedoman hidup manusia hingga akhir jaman
Peserta2	islam adalah agama akhir yang turun oleh Allah lalu nabi muhammad dan kitab al Quran
Peserta3	islam adalah agama yang paling benar mana ajar bukan hanya untuk satu umat tetapi juga umat yang lain
Peserta4	islam arti suatu ada atau yakin bahwa ia serah diri atas segala sesuatu kepada yang maha kuasa Allah
Peserta5	agama yang jadi pedoman manusia

Tabel 9. Hasil *Preprocessing* soal 2

Peserta	Hasil <i>Preprocessing</i> soal 2
Kunci	syahadat rupa kalimat saksi yang yakin bahwa tidak ada tuhan yang wajib sembah selain Allah dan nabi muhammad adalah utus Allah syahadat rupa rukun islam yang pertama
Peserta1	syahadat adalah kalimat yang yakin bahwa tiada tuhan selain Allah dan nabi muhammad adalah utus Allah
Peserta2	syahadat salah satu rukun islam
Peserta3	syahadat adalah kalimat tauhid yang mana jadi kunci orang masuk surga dan masuk agama islam dengan syahadat kita yakin bahwa tidak ada tuhan selain Allah dan rasulullah adalah utus Allah
Peserta4	syahadat adalah kalimat yang arti dan nyata bahwa tiada tuhan selain Allah dan muhammad adalah utusannya
Peserta5	dua kalimat yang isi saksi bahwa tiada tuhan selain Allah dan muhammad adalah Nabi dan rasul kita

Tabel 10. Hasil *Preprocessing* soal 3

Peserta	Hasil <i>Preprocessing</i> soal 3
Kunci	Muhammadiyah rupa organisasi islam yang diri oleh kh ahmad dahlan di kaum yogyakarta tanggal 18 november 1912
Peserta1	muhammadiyah adalah ikut nabi muhammad diri kh ahmad dahlan pada tanggal 18 november 1912
Peserta2	muhammadiyah diri oleh ahmad dahlan di yogyakarta pada tanggal 18 november 1912
Peserta3	muhammadiyah diri di yogyakarta diri kyai haji ahmad dahlan organisasi yang gerak dalam bidang agama sosial budaya dan politik
Peserta4	muhammadiyah adalah buah organisasi dakwah islam yang diri di yogyakarta oleh k h ahmad dahlan
Peserta5	organisasi islam bentuk kaum yogyakarta oleh ahmad dahlan

Tabel 11. Hasil *Preprocessing* soal 4

Peserta	Hasil <i>Preprocessing</i> soal 4
Kunci	agama mahasiswa masyarakat
Peserta1	mahasiswa masyarakat religiusitas intelektualitas agama
Peserta2	agama sosial dan didik
Peserta3	junjung tinggi nilai agama junjung tinggi nilai didik dasar al quran dan hadits bentuk kader muhammadiyah
Peserta4	dakwah islam
Peserta5	didik sosial agama

Tabel 12. Hasil *Preprocessing* soal 5

Peserta	Hasil <i>Preprocessing</i> soal 5
Kunci	djazman al kindi rosyad sholeh amien rais Abdul Had Marzuki Usman Rosyad Saleh
Peserta1	drs djazman k h ahmad dahlan siti waliah
Peserta2	k h ahmad dahlan
Peserta3	kh ahmad dahlan amin rais zaman al kindi
Peserta4	belum tau
Peserta5	immawan immawati

Tabel 13. Hasil *Preprocessing* soal 6

Peserta	Hasil <i>Preprocessing</i> soal 6
Kunci	bentuk akademisi islam yang akhlak mulia dalam rangka capai tuju muhammadiyah
Peserta1	cipta kader imm yang akhlak mulia untuk capai tuju muhammadiyah
Peserta2	bentuk kader muhammadiyah yang intelek
Peserta3	cetak kader muhammadiyah yang kreatif inspratif dan inovatif
Peserta4	dakwah juga komunikasi pada budaya masyarakat dam juga didik
Peserta5	didik sosial agama

### 4.3. Hasil Perhitungan Cosine Similarity

Setelah melalui proses *preprocessing*, data akan dilakukan proses perhitungan *Term Frequency* (TF) dengan cara menggabungkan *term* yang terdapat pada kunci dan jawaban peserta yang kemudian dihitung frekuensi kemunculan setiap kata sebagai bobot masing-masing *term*. TF adalah frekuensi dari kemunculan sebuah *term* dalam dokumen [16]. Setelah dilakukan pembobotan, langkah selanjutnya adalah menerapkan metode *cosine similarity* ke data yang sudah siap diolah. *Cosine similarity* merupakan algoritma yang digunakan dalam mengukur tingkat persamaan antara dua vektor. Tingkat persamaan antar dua vektor atau dokumen dinyatakan sama apabila keduanya mendekati nilai 1 (satu), sebaliknya jika hasilnya bernilai 0 (nol) maka berarti kedua vektor tidak memiliki nilai kesamaan atau berlawanan [6]. Berikut merupakan hasil perhitungan cosine similarity dari masing-masing jawaban peserta dan kunci jawaban pada tiap soal *pre-test* dan *post-test*.

Tabel 14. Hasil Perhitungan *Cosine Similarity Pre-Test*

Nama	<i>Cosine Similarity</i> soal 1	<i>Cosine Similarity</i> soal 2	<i>Cosine Similarity</i> soal 3	<i>Cosine Similarity</i> soal 4	<i>Cosine Similarity</i> soal 5	<i>Cosine Similarity</i> soal 6
Peserta 1	0.630062	0.644503	0.583383	0.774597	0.091287	0.572078
Peserta 2	0.634335	0.286446	0.770154	0.288675	0.000000	0.404520
Peserta 3	0.402911	0.557007	0.476331	0.123091	0.273861	0.213201
Peserta 4	0.222718	0.501280	0.626224	0.000000	0.000000	0.000000
Peserta 5	0.507093	0.514303	0.600245	0.333333	0.000000	0.000000

Tabel 15. Hasil Perhitungan *Cosine Similarity Post Test*

Nama	<i>Cosine Similarity</i> soal 1	<i>Cosine Similarity</i> soal 2	<i>Cosine Similarity</i> soal 3	<i>Cosine Similarity</i> soal 4	<i>Cosine Similarity</i> soal 5	<i>Cosine Similarity</i> soal 6
Peserta 1	0.462910	0.736590	0.770154	1.0	0.292770	0.818182
Peserta 2	0.595854	0.372104	0.804400	0.0	0.577350	0.174078
Peserta 3	0.429198	0.280224	0.594089	0.0	0.298142	0.369274
Peserta 4	0.243975	0.289414	0.325396	0.0	0.149071	0.522233
Peserta 5	0.483494	0.393863	0.619930	0.0	0.527046	0.909091

#### 4.3.1. Persentase Kemiripan

Perhitungan persentase kemiripan data yang telah diproses bertujuan agar memudahkan dalam membaca data atau hasil perhitungan *cosine similarity* serta menentukan kemiripan antara 2 data. Selain itu juga berfungsi untuk pengenalan pola hasil perhitungan dan menganalisa data sejauh mana metode ini dapat diterapkan dalam penelitian dan mendapatkan hasil yang sesuai efektif dan sesuai dengan harapan. Berikut adalah tabel persentase kemiripan hasil perhitungan metode *cosine similarity*.

Tabel 16. Persentase Kemiripan Jawaban *Pre Test*

Nama	<i>Cosine Similarity</i> soal 1	<i>Cosine Similarity</i> soal 2	<i>Cosine Similarity</i> soal 3	<i>Cosine Similarity</i> soal 4	<i>Cosine Similarity</i> soal 5	<i>Cosine Similarity</i> soal 6
------	------------------------------------	------------------------------------	------------------------------------	------------------------------------	------------------------------------	------------------------------------

Peserta 1	63,01%	64,45%	58,34%	77,46%	9,13%	57,21%
Peserta 2	63,43%	28,64%	77,02%	28,87%	0,00%	40,45%
Peserta 3	40,29%	55,70%	47,63%	12,31%	27,39%	21,32%
Peserta 4	22,27%	50,13%	62,62%	0,00%	0,00%	0,00%
Peserta 5	50,71%	51,43%	60,02%	33,33%	0,00%	0,00%

Tabel 17. Persentase Kemiripan Jawaban *Post Test*

Nama	<i>Cosine Similarity</i> soal 1	<i>Cosine Similarity</i> soal 2	<i>Cosine Similarity</i> soal 3	<i>Cosine Similarity</i> soal 4	<i>Cosine Similarity</i> soal 5	<i>Cosine Similarity</i> soal 6
Peserta 1	46,29%	73,66%	77,02%	100%	29,28%	81,82%
Peserta 2	59,59%	37,21%	80,44%	0,0%	57,74%	17,41%
Peserta 3	42,92%	28,02%	59,41%	0,0%	29,81%	36,93%
Peserta 4	24,40%	28,94%	32,54%	0,0%	14,91%	52,22%
Peserta 5	48,35%	39,39%	61,99%	0,0%	52,70%	90,91%

Tabel diatas adalah hasil dari perhitungan *cosine similarity* yang dimulai dari proses *preprocessing* hingga perhitungan dari penerapan metode soal 1 sampai soal 6 berdasarkan data pada data kunci jawaban dan jawaban peserta.

## 5. Penilaian

Untuk proses selanjutnya adalah menghitung bobot dari masing-masing soal sampai dengan perhitungan nilai akhir yang didapatkan masing-masing peserta. Cara menghitung nilai ialah dengan mengalikan bobot soal yang telah ditentukan dengan hasil perhitungan *cosine similarity* kemudian dibagi 3 (tiga). Hasil perhitungan nilai soal hingga jumlah nilai keseluruhan *pre-test* dan *post-test* dapat dilihat pada tabel berikut.

Tabel 18. Nilai *Pre Test*

Nama	<i>nilai soal 1</i>	<i>nilai soal 2</i>	<i>nilai soal 3</i>	<i>nilai soal 4</i>	<i>nilai soal 5</i>	<i>nilai soal 6</i>	Total Nilai
Peserta 1	3.15	3.22	2.92	3.87	0.46	2.86	5.5
Peserta 2	3.17	1.43	3.85	1.44	0.00	2.02	4.0
Peserta 3	2.01	2.79	2.38	0.62	1.37	1.07	3.4
Peserta 4	1.11	2.51	3.13	0.00	0.00	0.00	2.2
Peserta 5	2.54	2.57	3.00	1.67	0.00	0.00	3.3

Tabel 19. Nilai *Post Test*

Nama	<i>nilai soal 1</i>	<i>nilai soal 2</i>	<i>nilai soal 3</i>	<i>nilai soal 4</i>	<i>nilai soal 5</i>	<i>nilai soal 6</i>	Total Nilai
Peserta 1	2.31	3.68	3.85	5.0	1.46	4.09	6.8
Peserta 2	2.98	1.86	4.02	0.0	2.89	0.87	4.2
Peserta 3	2.15	1.40	2.97	0.0	1.49	1.85	3.3
Peserta 4	1.22	1.45	1.63	0.0	0.75	2.61	2.6
Peserta 5	2.42	1.97	3.10	0.0	2.64	4.55	4.9

Pada tabel 14 dan tabel 15 merupakan hasil perhitungan penilaian metode *cosine similarity*. Hasil perhitungan *cosine similarity* pada soal *pre-test* menunjukkan hasil tertinggi

diperoleh peserta 1 pada soal 4 dan peserta 2 pada soal 3 dengan tingkat persamaan 77,46% atau mendekati 1. Sedangkan pada soal *post-test* terdapat peserta yang memiliki nilai persamaan tertinggi jawabannya mencapai persamaan angka 1 atau 100% sama yaitu peserta 1 pada soal 4, peserta 5 pada soal 6 dengan hasil 90,91%, dan peserta 1 pada soal 6 dengan hasil 81,82%. Sedangkan hasil 0,0 pada jawaban *pre-test* dan *post-test* beberapa peserta dikarenakan *term* antar kunci jawaban dan jawaban peserta tidak memiliki persamaan.

Dalam penilaian soal esai, bobot nilai pada soal menjadi dasar perhitungan untuk mendapatkan total nilai akhir peserta. Adapun nilai yang dihasilkan dari perhitungan tersebut berdasarkan pada jawaban peserta dari soal yang disediakan, yang mana tentunya masing-masing peserta memiliki jawaban yang polanya tidak sama. Berdasarkan perhitungan tersebut, metode *cosine similarity* cocok digunakan dalam perhitungan jawaban esai, namun metode *cosine similarity* tidak dapat mendeteksi persamaan kalimat. Sehingga, kata berbeda yang memiliki makna sama akan dihitung sebagai *term* yang berbeda.

## 6. KESIMPULAN

Penelitian ini telah mengimplementasikan metode *cosine similarity* untuk menghitung nilai jawaban *pre-test* dan *post-test* peserta perkaderan. Hasil perhitungan pada soal *pre-test* mendapatkan hasil tertinggi dengan tingkat persamaan 77,45% oleh peserta 1 pada soal 4 dan peserta 2 pada soal 3, dan *post-test* dengan hasil tertinggi 100% oleh peserta 1 pada soal 4. Sedangkan untuk hasil 0,0 yang diperoleh beberapa peserta dikarenakan *term* antar kunci jawaban dan jawaban peserta tidak memiliki persamaan. Tentunya soal esai memiliki tingkat kesulitan tersendiri dalam melakukan penilaian dikarenakan pola jawaban masing-masing tentunya berbeda. Berdasarkan penelitian ini, metode *cosine similarity* dapat digunakan dalam pemberian penilaian jawaban esai secara otomatis dilihat dari hasil yang telah diperoleh dalam penelitian ini.

## REFERENSI

- [1] F. A. Sandiah, "Kader, Perkaderan, Kaderisasi, dan Sumber Daya Insani di Muhammadiyah #1," *Muhammadiyah*, 2023. <https://muhammadiyah.or.id/kader-perkaderan-kaderisasi-dan-sumber-daya-insani-di-muhammadiyah/> (accessed Nov. 11, 2023).
- [2] I. Magdalena, M. Nurul Annisa, G. Ragin, and A. R. Ishaq, "Analisis Penggunaan Teknik Pre-Test Dan Post-Test Pada Mata Pelajaran Matematika Dalam Keberhasilan Evaluasi Pembelajaran Di Sdn Bojong 04," *J. Pendidik. dan Ilmu Sos.*, vol. 3, no. 2, pp. 150–165, 2021, [Online]. Available: <https://ejournal.stitpn.ac.id/index.php/nusantara>.
- [3] D. O. Sihombing, "Implementasi Natural Language Processing ( NLP ) dan Algoritma Cosine Similarity dalam Penilaian Ujian Esai Otomatis," vol. 4, pp. 396–406, 2022, doi: 10.30865/json.v4i2.5374.
- [4] T. Wahyuningsih, Henderi, and Winarno, "Text Mining an Automatic Short Answer Grading (ASAG), Comparison of Three Methods of Cosine Similarity, Jaccard Similarity and Dice's Coefficient," *J. Appl. Data Sci.*, vol. 2, no. 2, pp. 45–54, 2021, doi: 10.47738/jads.v2i2.31.
- [5] Anggun Fergina, Kamdan, and Moh. Iqbal Maulana, "Analisis Ujian Essay Menggunakan Metode Cosine Similarity," *Jurnal RESTIKOM : Riset Teknik Informatika dan Komputer*, vol. 4, no. 1. pp. 36–44, 2022, doi: 10.52005/restikom.v4i1.107.

- [6] K. Telaumbanua and L. Nababan, "Implementasi Metode Cosine Similarity Dalam Mendeteksi Kemiripan Dan Perbedaan Gambar Hasil Scan Berbasis Android," *Informatics Eng. Electron. Data*, vol. 1, no. 1, pp. 27–36, 2022, doi: 10.59840/ieed.v1i1.176.
- [7] I. Mawanta, T. S. Gunawan, and W. Wanayumini, "Uji Kemiripan Kalimat Judul Tugas Akhir dengan Metode Cosine Similarity dan Pembobotan TF-IDF," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 726, 2021, doi: 10.30865/mib.v5i2.2935.
- [8] E. L. Amalia, A. J. Jumadi, I. A. Mashudi, and D. W. Wibowo, "Analisis Metode Cosine Similarity Pada Aplikasi Ujian Online Otomatis (Studi Kasus JTI POLINEMA)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 2, p. 343, 2021, doi: 10.25126/jtiik.2021824356.
- [9] N. S. Atmaja, "Implementasi Metode Levensthein Distance Dan Cosine Similarity Untuk Deteksi Kemiripan Gambar," *RJOCS (Riau J. Comput. Sci.)*, vol. 8, no. 2, pp. 85–93, 2022, doi: 10.30606/rjocs.v8i2.1415.
- [10] A. W. Nila Andriani, "Implementasi Text Mining Klasifikasi Topik Tugas Akhir Mahasiswa Teknik Informatika Menggunakan Pembobotan TF-IDF dan Metode Cosine Similarity Berbasis Web," *Senamika*, no. September, pp. 130–137, 2021, [Online]. Available: <https://conference.upnvj.ac.id/index.php/senamika/article/view/1807%0Ahttps://conference.upnvj.ac.id/index.php/senamika/article/download/1807/1350>.
- [11] A. E. Budiman, "Analisis Pengaruh Teks Preprocessing Terhadap Deteksi Plagiarisme Pada Dokumen Tugas Akhir," vol. 6, pp. 475–488, 2020.
- [12] A. R. Lahitani, "Automated Essay Scoring menggunakan Cosine Similarity pada Penilaian Esai Multi Soal," *J. Kaji. Ilm.*, vol. 22, no. 2, pp. 107–118, 2022, doi: 10.31599/jki.v22i2.1121.
- [13] H. Herlambang, J. Suwita, B. Tiara, P. Studi, S. Informasi, and P. Sistem, "Analisa Dan Perancangan Sistem Pendeteksi Plagiarisme Skripsi Pada STMIK Insan Pembangunan Menggunakan Metode Cosine Similarity," vol. 9, no. 1, 2021.
- [14] M. W. Arbiantono, "Pengembangan Aplikasi Assessment Tool Menggunakan Metode Cosine Semantic Similarity untuk Automatic Scoring Jawaban Tes Uraian pada Mata Pelajaran Basis Data di SMKN 1 Surabaya," *Inf. Technol. Educ.*, vol. 5, no. 2, pp. 657–666, 2021.
- [15] A. Apriani, H. Zakiyudin, and K. Marzuki, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta," *J. Bumigora Inf. Technol.*, vol. 3, no. 1, pp. 19–27, 2021, doi: 10.30812/bite.v3i1.1110.
- [16] D. N. Lindang, A. Y. Muniar, A. Halid, M. Muhajirin, and A. Amiruddin, "Sistem Penentuan Kemiripan Antar Skripsi Menggunakan Metode Cosine Similarity Pada Perpustakaan," *Sntei*, pp. 321–324, 2022.