

**PERBANDINGAN DOKUMEN MODUL AJAR MATA PELAJARAN  
INFORMATIKA SEKOLAH MENENGAH ATAS (SMA)  
BERBASIS *SIMILARITY***

**THESIS**

**Oleh:  
SYAIFUDIN RAMADHANI  
NIM. 220605210001**



**PROGRAM STUDI MAGISTER INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2023**

**PERBANDINGAN DOKUMEN MODUL AJAR MATA PELAJARAN  
INFORMATIKA SEKOLAH MENENGAH ATAS (SMA)  
BERBASIS *SIMILARITY***

**THESIS**

**Diajukan kepada:  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
Untuk memenuhi Salah Satu Persyaratan dalam  
Memperoleh Gelar Magister Komputer (M.Kom)**

**Oleh:  
SYAIFUDIN RAMADHANI  
NIM. 220605210001**

**PROGRAM STUDI MAGISTER INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2023**

**PERBANDINGAN DOKUMEN MODUL AJAR MATA PELAJARAN  
INFORMATIKA SEKOLAH MENENGAH ATAS (SMA)  
BERBASIS *SIMILARITY***

**THESIS**

**Diajukan kepada:  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
Untuk memenuhi Salah Satu Persyaratan dalam  
Memperoleh Gelar Magister Komputer (M.Kom)**

**Oleh:  
SYAIFUDIN RAMADHANI  
NIM. 220605210001**

**PROGRAM STUDI MAGISTER INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2023**

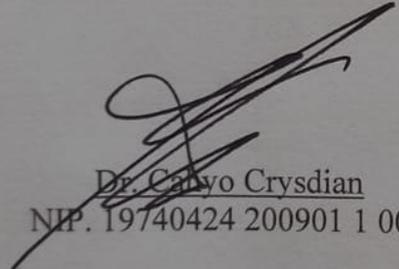
**PERBANDINGAN DOKUMEN MODUL AJAR MATA PELAJARAN  
INFORMATIKA SEKOLAH MENENGAH ATAS (SMA)  
BERBASIS *SIMILARITY***

**THESIS**

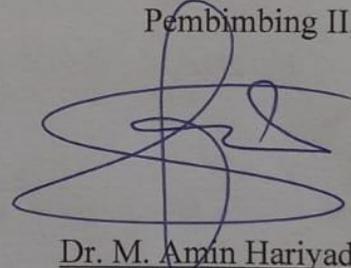
**Oleh:  
SYAIFUDIN RAMADHANI  
NIM. 220605210001**

Telah diperiksa dan disetujui untuk di uji:  
Tanggal 17 November 2023

Pembimbing I,

  
Dr. Cahyo Crysdiyan  
NIP. 19740424 200901 1 008

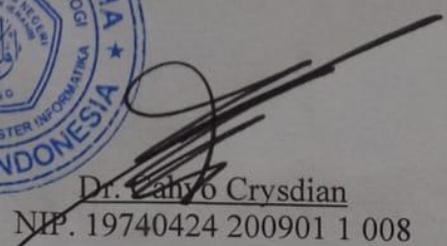
Pembimbing II,

  
Dr. M. Amin Hariyadi, M.T  
NIP. 19670018 200501 1 001

Mengetahui,

Ketua Program Studi Magister Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim



  
Dr. Cahyo Crysdiyan  
NIP. 19740424 200901 1 008

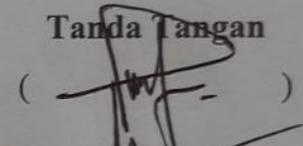
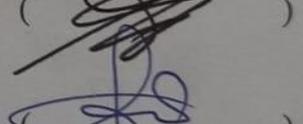
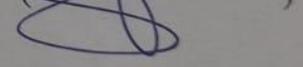
PERBANDINGAN DOKUMEN MODUL AJAR MATA PELAJARAN  
INFORMATIKA SEKOLAH MENENGAH ATAS (SMA)  
BERBASIS *SIMILARITY*

THESIS

Oleh:  
**SYAIFUDIN RAMADHANI**  
NIM. 220605210001

Telah Dipertahankan di Depan Dewan Penguji Thesis  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
Untuk Memperoleh Gelar Magister Komputer (M.Kom)  
Tanggal 30 November 2023

**Susunan Dewan Penguji**

		Tanda Tangan
Penguji Utama	: <u>Dr. Yunifa Miftachul Arif, M.T</u> NIP. 19830616 201101 1 004	(  )
Ketua Penguji	: <u>Dr. Irwan Budi Santoso, M.Kom</u> NIP. 19770103 201101 1 004	(  )
Sekretaris Penguji	: <u>Dr. Cahyo Crysdian</u> NIP. 19740424 200901 1 008	(  )
Anggota Penguji	: <u>Dr. M. Amin Hariyadi, M.T</u> NIP. 19670018 200501 1 001	(  )

Mengetahui dan Mengesahkan,  
Ketua Program Studi Magister Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim



  
Dr. Cahyo Crysdian  
19740424 200901 1 008

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Syaifudin Ramadhani

NIM : 220605210001

Program Studi : Magister Informatika

Fakultas : Sains dan Teknologi

Menyatakan dengan sebenarnya bahwa Thesis yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan Thesis ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 30 November 2023

Yang menyatakan,



Syaifudin Ramadhani

NIM. 220605210001

## **MOTTO**

*“Hal yang baru dan belum pernah anda alami atau pelajari adalah tantangan.  
Terus semangat belajar, mengeksplere diri serta tingkatkan kreativitas untuk hal  
baru yang menantang”*

## **PERSEMBAHAN**

Dengan mengucapkan syukur Alhamdulillah rabbil alamin, Thesis ini saya persembahkan untuk :

1. Seluruh keluarga tercinta (Istri, Anak pertama Akhdan Habibi Ramadhani dan Anak kedua Aleena Hamida Ramadhani), Orang Tua (Almarhum Ayahanda dan Ibunda tercinta, serta Kedua Mertua) yang selalu memberikan dukungan, doa dan semangat yang tiada henti.
2. Seluruh Civitas Akademika Universitas Islam Negeri Maulana Malik Ibrahim Malang untuk dapat menimba ilmu, teknologi serta agama.
3. Seluruh jajaran Tim Manajemen SMA Negeri 4 Malang yang memberikan dukungan moril, materiil dan kesempatan untuk dapat studi lanjut.
4. Tim Kurikulum SMA Negeri 4 Malang yang banyak mendukung untuk penyelesaian Thesis ini.
5. Seluruh pengurus Komite SMA Negeri 4 Malang yang memberikan kesempatan mendapatkan pembiayaan dalam studi lanjut.
6. Seluruh rekan guru mata pelajaran informatika SMA Negeri dan Swasta se-Kota Malang yang memberikan dukungan data dalam penyelesaian penelitian ini.
7. Seluruh rekan-rekan Asosiasi Mahasiswa Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang semua angkatan atas kerjasamanya selama ini.
8. Bapak, ibu, saudara dan rekan-rekan sekalian yang tidak bisa saya sebutkan satu persatu yang telah mendukung penyelesaian Thesis ini.

## KATA PENGANTAR

*Assalamu'alaikum Wr. Wb.*

Syukur *Alhamdulillah* penulis menghaturkan kehadiran Allah SWT yang telah melimpahkan Rahmat dan Hidayah-Nya, sehingga penulis dapat menyelesaikan studi di Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang sekaligus menyelesaikan Thesis ini dengan baik. Selanjutnya penulis menghaturkan ucapan terima kasih teriring do'a dan harapan kepada semua pihak yang telah membantu terselesaikannya Thesis ini. Ucapan terima kasih ini penulis sampaikan kepada:

1. Bapak Dr. Cahyo Crysdiyan dan Bapak Dr. M. Amin Hariyadi, M.T, selaku dosen pembimbing Thesis, yang telah banyak memberikan arahan dan pengalaman yang sangat berharga.
2. Segenap civitas akademika Program Studi Magister Informatika, terutama seluruh Bapak / Ibu dosen, terima kasih atas ilmu dan bimbingannya.
3. Keluarga tercinta yang senantiasa memberikan do'a dan semangat.
4. Semua rekan-rekan seperjuangan yang ikut mendukung dan membantu.
5. Penulis menyadari bahwa dalam penyusunan Thesis ini masih terdapat kekurangan dan penulis berharap semoga Thesis ini bisa memberikan manfaat kepada para pembaca khususnya bagi penulis secara pribadi.

*Wasalamu'alaikum Wr. Wb*

Malang, 30 November 2023

Penulis



3.2.2.6. Proses Perhitungan Pemenuhan Dokumen Modul Ajar Informatika dengan <i>Knowledge Area</i> .....	32
3.2.3. Implementasi Sistem .....	34
3.2.3.1. Tahap <i>Preprocessing Text</i> .....	36
3.2.3.1.1. <i>Preprocessing Text</i> dengan <i>Case Folding</i> .....	36
3.2.3.1.2. <i>Preprocessing Text</i> dengan <i>Filtering</i> .....	37
3.2.3.1.3. <i>Preprocessing Text</i> dengan <i>Stemming</i> .....	37
3.2.3.1.4. <i>Preprocessing Text</i> dengan <i>Tokenizing</i> .....	38
3.2.3.1.5. Tahap Similarity Dokumen .....	38
3.2.3.1.6. Tahap Perhitungan Pemenuhan Konten Dokumen Modul Ajar Informatika dengan Konten Sub <i>Knowledge Area</i> .....	39
3.2.4. Ujicoba .....	39
3.2.4.1. Skenario Pertama Pengujian Similarity Berbasis Term Dengan Gabungan Term .....	40
3.2.4.2. Skenario Kedua Pengujian Similarity Berbasis Term Dengan Kalimat .....	41
3.2.4.3. Skenario Ketiga Pengujian Similarity Berbasis Term Dengan Paragraf .....	43
3.2.5. Perbandingan Pengukuran Efektifitas Similarity Metode Cosine Similarity dan Word2Vec .....	44
<b>BAB IV METODE COSINE SIMILARITY</b> .....	47
4.1. Desain Metode .....	47
4.2. Ujicoba .....	48
4.2.1. Skenario Ujicoba Pertama .....	49
4.2.2. Skenario Ujicoba Kedua .....	52
4.2.3. Skenario Ujicoba Ketiga .....	55
<b>BAB V METODE WORD2VEC</b> .....	58
5.1. Desain Metode .....	58
5.2. Ujicoba .....	59
5.2.1. Skenario Ujicoba Pertama .....	59
5.2.2. Skenario Ujicoba Kedua .....	63
5.2.3. Skenario Ujicoba Ketiga .....	67
<b>BAB VI PEMBAHASAN</b> .....	71
6.1. Perbandingan Hasil Pemenuhan <i>Knowledge Area</i> .....	71
6.2. Perbandingan Pengukuran Akurasi Pemenuhan <i>Knowledge Area</i> .....	86
6.3. Analisis Similarity pada Dokumen dalam pandangan Islam .....	90
<b>BAB VII KESIMPULAN</b> .....	92
7.1. Kesimpulan .....	92
7.2. Saran .....	93
<b>DAFTAR PUSTAKA</b> .....	94

<b>LAMPIRAN.....</b>	<b>97</b>
----------------------	-----------

## DAFTAR GAMBAR

Gambar 2.1 : Kerangka Teori .....	14
Gambar 3.1 : Kerangka Konsep .....	19
Gambar 3.2 : Prosedur Penelitian.....	20
Gambar 3.3 : Contoh dokumen teks <i>Computer Science Curricula 2013</i> dan Dokumen Modul Ajar Mata Pelajaran Informatika jenjang SMA	21
Gambar 3.4 : Desain Sistem.....	22
Gambar 3.5 : Proses Teks <i>Preprocessing</i> .....	24
Gambar 3.6 : Proses <i>Case Folding</i> .....	25
Gambar 3.7 : Proses <i>Filtering</i> .....	26
Gambar 3.8 : Proses <i>Stemming</i> .....	26
Gambar 3.9 : Proses <i>Tokenizing</i> .....	27
Gambar 3.10 : Proses <i>similarity</i> dokumen menggunakan metode Cosine Similarity .....	28
Gambar 3.11 : Arsitektur CBOW pada Word2Vec .....	30
Gambar 3.12 : Proses <i>similarity</i> dokumen menggunakan metode Word2Vec .....	31
Gambar 3.13 : Tahap Pemenuhan Hasil <i>Similarity</i> dengan Knowledge Area .....	33
Gambar 3.14 : Desain Prototipe Sistem .....	34
Gambar 3.15 : Graphic <i>User Interface</i> (GUI) Implementasi Sistem Tab 1 .....	35
Gambar 3.16 : Graphic <i>User Interface</i> (GUI) Implementasi Sistem Tab 2 .....	35
Gambar 3.17 : Graphic <i>User Interface</i> (GUI) Implementasi Sistem Tab 3 .....	36
Gambar 3.18 : Tahapan Pengujian <i>Similarity</i> Berbasis <i>Term</i> dengan Gabungan Term .....	40
Gambar 3.19 : Tahapan Pengujian <i>Similarity</i> Berbasis <i>Term</i> dengan Gabungan Kalimat .....	42
Gambar 3.20 : Tahapan Pengujian <i>Similarity</i> Berbasis <i>Term</i> dengan Gabungan Paragraf.....	43
Gambar 3.21 : Tahapan Perbandingan Pengukuran Akurasi Pemenuhan <i>Knowledge Area</i> .....	45
Gambar 4.1 : Desain Metode Cosine Similarity .....	47
Gambar 4.2 : Tahap Pemenuhan <i>Knowledge Area</i> .....	48
Gambar 4.3 : Tahapan skenario uji coba pertama dengan metode Cosine <i>similarity</i> .....	49
Gambar 4.4 : Visualisasi pemenuhan <i>Knowledge Area</i> uji coba pertama .....	51
Gambar 4.5 : Tahapan skenario uji coba kedua .....	52
Gambar 4.6 : Visualisasi pemenuhan <i>Knowledge Area</i> uji coba kedua.....	54
Gambar 4.7 : Tahapan skenario uji coba ketiga.....	55
Gambar 4.8 : Visualisasi pemenuhan <i>Knowledge Area</i> uji coba ketiga .....	57
Gambar 5.1 : Desain metode Word2Vec .....	58

Gambar 5.2	: Tahapan skenario ujicoba pertama dengan metode Word2Vec....	59
Gambar 5.3	: Visualisasi pemenuhan <i>Knowledge Area</i> ujicoba pertama .....	62
Gambar 5.4	: Tahapan skenario ujicoba kedua dengan metode Word2Vec .....	63
Gambar 5.5	: Visualisasi pemenuhan <i>Knowledge Area</i> ujicoba kedua.....	66
Gambar 5.6	: Tahapan skenario ujicoba ketiga dengan metode Word2Vec .....	67
Gambar 5.7	: Visualisasi pemenuhan <i>Knowledge Area</i> ujicoba ketiga .....	70
Gambar 6.1	: Hasil rata-rata pemenuhan Knowledge Area AI-Algoritma dan Kompleksitas dari perbandingan elemen materi menggunakan metode Cosine Similarity dan Word2Vec .....	75
Gambar 6.2	: Hasil rata-rata pemenuhan Knowledge Area AR-Arsitektur dan Organisasi dari perbandingan elemen materi menggunakan metode Cosine Similarity dan Word2Vec .....	75
Gambar 6.3	: Hasil rata-rata pemenuhan Knowledge Area CN-Ilmu Komputasi dari perbandingan elemen materi menggunakan metode Cosine Similarity dan Word2Vec .....	76
Gambar 6.4	: Hasil rata-rata pemenuhan Knowledge Area SDF- Dasar-Dasar Pengembangan Perangkat Lunak dari perbandingan elemen materi menggunakan metode Cosine Similarity dan Word2Vec .....	77
Gambar 6.5	: Hasil rata-rata pemenuhan Knowledge Area Dasar-Dasar Sistem-SF dari perbandingan elemen materi menggunakan metode Cosine Similarity dan Word2Vec .....	77
Gambar 6.6	: Hasil rata-rata pemenuhan Knowledge Area GV-Grafik dan Visualisasi dari perbandingan elemen materi menggunakan metode Cosine Similarity dan Word2Vec .....	78
Gambar 6.7	: Hasil rata-rata pemenuhan Knowledge Area IAS-Jaringan dan Keamanan Informasi dari perbandingan elemen materi menggunakan metode Cosine Similarity dan Word2Vec .....	79
Gambar 6.8	: Hasil rata-rata pemenuhan Knowledge Area IM-Manajemen Informasi dari perbandingan elemen materi menggunakan metode Cosine Similarity dan Word2Vec .....	79
Gambar 6.9	: Hasil rata-rata pemenuhan Knowledge Area HCI-Interaksi Manusia dan Komputer dari perbandingan elemen materi menggunakan metode Cosine Similarity dan Word2Vec .....	80
Gambar 6.10	: Hasil rata-rata pemenuhan Knowledge Area IS-Sistem Cerdas dari perbandingan elemen materi menggunakan metode Cosine Similarity dan Word2Vec .....	81
Gambar 6.11	: Hasil rata-rata pemenuhan Knowledge Area NC-Jaringan dan Komunikasi dari perbandingan elemen materi menggunakan metode Cosine Similarity dan Word2Vec .....	81

Gambar 6.12 : Hasil rata-rata pemenuhan Knowledge Area OS-Sistem Operasi dari perbandingan elemen materi menggunakan metode Cosine Similarity dan Word2Vec.....	82
Gambar 6.13 : Hasil rata-rata pemenuhan Knowledge Area PD-Komputasi Paralel dan Terdistribusi dari perbandingan elemen materi menggunakan metode Cosine Similarity dan Word2Vec.....	82
Gambar 6.14 : Hasil rata-rata pemenuhan Knowledge Area PBD-Pembangunan Berbasis Platform dari perbandingan elemen materi menggunakan metode Cosine Similarity dan Word2Vec.....	83
Gambar 6.15 : Hasil rata-rata pemenuhan Knowledge Area PL-Bahasa Pemrograman dari perbandingan elemen materi menggunakan metode Cosine Similarity dan Word2Vec.....	84
Gambar 6.16 : Hasil rata-rata pemenuhan Knowledge Area SE-Rekayasa Perangkat Lunak dari perbandingan elemen materi menggunakan metode Cosine Similarity dan Word2Vec.....	84
Gambar 6.17 : Hasil rata-rata pemenuhan Knowledge Area SP-Masalah Sosial dan Praktik Profesi dari perbandingan elemen materi menggunakan metode Cosine Similarity dan Word2Vec.....	85
Gambar 6.18 : Hasil rata-rata pemenuhan Knowledge Area DS-Struktur Diskrit dari perbandingan elemen materi menggunakan metode Cosine Similarity dan Word2Vec.....	86

## DAFTAR TABEL

Tabel 2.1. Daftar Jurnal Penelitian.....	15
Tabel 3.1. Sampel Hasil Similarity Dokumen .....	32
Tabel 3.2. Sampel Hasil Pemenuhan Konten Dokumen Modul Ajar Informatika	34
Tabel 3.3. Ketentuan klasifikasi persentase pemenuhan <i>Knowledge Area</i> .....	44
Tabel 3.4. Ketentuan Confusion Matrix.....	45
Tabel 4.1. Hasil similarity skenario ujicoba pertama.....	50
Tabel 4.2. Hasil persentase pemenuhan <i>Knowledge Area</i> dengan ujicoba kedua	53
Tabel 4.3. Hasil persentase pemenuhan ujicoba ketiga.....	56
Tabel 5.1. Hasil persentase pemenuhan <i>Knowledge Area</i> ujicoba pertama dengan metode Word2Vec .....	60
Tabel 5.2. Hasil persentase pemenuhan <i>Knowledge Area</i> ujicoba kedua dengan metode Word2Vec .....	64
Tabel 5.3. Hasil persentase pemenuhan <i>Knowledge Area</i> ujicoba ketiga dengan metode Word2Vec .....	68
Tabel 6.1. Hasil persentase pemenuhan <i>Knowledge Area</i> menggunakan metode Cosine Similarity dan Word2Vec ujicoba pertama .....	71
Tabel 6.2. Hasil persentase pemenuhan <i>Knowledge Area</i> menggunakan metode Cosine Similarity dan Word2Vec ujicoba kedua .....	72
Tabel 6.3. Hasil persentase pemenuhan <i>Knowledge Area</i> menggunakan metode Cosine Similarity dan Word2Vec ujicoba ketiga .....	73
Tabel 6.4. Sampel Demografi Responden.....	86
Tabel 6.5. Sampel Ground Truth.....	87
Tabel 6.6. Sampel Hasil Uji Validasi Ground Truth pada KA. Algoritma dan Kompleksitas .....	88
Tabel 6.7. Hasil perhitungan akurasi metode Cosine-similarity dan Word2Vec..	89

## ABSTRAK

Ramadhani, Syaifudin. 2023. **Perbandingan Dokumen Modul Ajar Mata Pelajaran Informatika Sekolah Menengah Atas (SMA) Berbasis Similarity**. Thesis. Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Cahyo Crysdiyan.(II) Dr. Dr. M. Amin Hariyadi, M.T.

**Kata kunci:** *Text Mining, Similarity Document*, Kurikulum Informatika, Dokumen Kurikulum, *Cosine Similarity, Word2Vec*

Pemerintah telah merilis kebijakan Kurikulum Merdeka pada tahun 2022. Kebijakan tersebut mengubah penggunaan kurikulum sebelumnya yaitu kurikulum 2013. Dalam pelaksanaan kurikulum merdeka pada jenjang Sekolah Menengah Atas (SMA) yaitu pada Fase E, mata pelajaran informatika memiliki elemen materi sebanyak delapan elemen dasar yang dapat dikembangkan secara mandiri pada setiap sekolah. Dalam praktik pengembangan modul ajar oleh guru di sekolah belum ada evaluasi terukur dari pemerintah untuk mengukur pengembangan konten materi apakah sudah sesuai dengan kurikulum standar yang ada. Pada penelitian ini mengusulkan analisis *similarity* dokumen modul ajar mata pelajaran informatika dengan pembandingan dokumen *Computer Science Curricula 2013*. Dengan mengusulkan metode *Cosine Similarity* mendapatkan hasil rata-rata *similarity* sebesar 0.29606 sedangkan dengan menggunakan metode *Word2Vec* mendapatkan hasil skor *similarity* dengan rata-rata *similarity* sebesar 0.98449. Pada tahap pemenuhan *Knowledge Area* dari skor *similarity* yang didapatkan dengan metode *Cosine-similarity* pada seluruh *Knowledge Area* yang diujikan mendapatkan rata-rata pemenuhan sebesar 29.6%, sedangkan dengan metode *Word2Vec* pada seluruh *Knowledge Area* yang diujikan mendapatkan rata-rata pemenuhan sebesar 98.44%. Pada tahap pengukuran akurasi dengan metode *Cosine-similarity* mendapatkan rata-rata akurasi pada seluruh *Knowledge Area* yang diujikan sebesar 0.6525 dan dengan metode *Word2Vec* mendapatkan rata-rata akurasi pada seluruh *Knowledge Area* yang diujikan sebesar 0.5895.

## ABSTRACT

Ramadhani, Syaifudin. 2023. **Comparison of Similarity-Based High School (SMA) Informatics Subject Teaching Module Documents**. Theses. Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Promotor: (I) Dr. Cahyo Crys dian.(II) Dr. Dr. M. Amin Hariyadi, M.T.

**Key words:** Text Mining, Similarity Document, Computer Science Curriculum, Curriculum Document, Cosine-similarity, Word2Vec

The government has issued an Independent Curriculum policy in 2022. This policy changes the use of the previous curriculum, namely the 2013 curriculum. In implementing the independent curriculum at the Senior High School (SMA) level, namely in Phase E, the informatics subject has eight basic material elements that can be developed independently at each school. In the practice of open development modules by teachers in schools, there has been no measurable evaluation from the government to measure whether the content development material is in accordance with the existing standard curriculum. In this research, recommendations for similarity analysis of informatics teaching module documents are compared with the 2013 Computer Science Curriculum document comparison. With the addition of the Cosine Similarity method, we get an average similarity result of 0.29606, while using the Word2Vec method we get a similarity score result with an average similarity of 0.98449. At the Knowledge Area stage, the similarity scores obtained using the Cosine-similarity method for all Knowledge Areas tested obtained a satisfactory average of 29.6%, while using the Word2Vec method for all Knowledge Areas tested obtained a satisfactory average of 98.44%. At the effectiveness stage of measurement using the Cosine-similarity method, the average accuracy in all Knowledge Areas tested was 0.6525 and with the Word2Vec method, the average accuracy in all Knowledge Areas tested was 0.5895.

## الملخص

رمضاني، شيف الدين. 2023. مقارنة وثائق وحدة التدريس لمواد المعلوماتية في المدرسة الثانوية على أساس التشابه (سيميلاريتي). الأطروحة. برنامج دراسة ماجستير المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف: (1) الدكتور جاهيو كريسديان (2) الدكتور محمد أمين هاريادي، الماجستير.

الكلمات الدالة: منهج دراسة المعلوماتية، وثيقة منهج الدراسة, *Text Mining, Similarity Document, Cosine Similarity, Word2Vec*

أصدرت الحكومة منهج دراسة "مردكا" في عام 2022. وتغير هذا المنهج استخدام المنهج السابق وهو منهج دراسة "2013". عند تنفيذ منهج "مردكا" على مستوى المدرسة الثانوية في مرحلة "ه"، يحتوي موضوع المعلوماتية على ثمانية عناصر مادية أساسية يمكن تطويرها بشكل مستقل في كل مدرسة. في ممارسة تطوير وحدات التدريس من قبل المعلمين في المدارس، لم يكن هناك تقييم قابل للقياس من قبل الحكومة لقياس ما إذا كان تطوير محتوى المواد يتوافق مع المنهج القياسي الحالي. في هذا البحث يقترح الباحث تحليل "سيميلاريتي" لوثائق وحدة تدريس المعلوماتية من خلال مقارنة وثائق "كمبيوتر ساينس كريكلا 2013". ومن خلال اقتراح طريقة تشابه جيب التمام (Cosine Similarity) حصل على نتيجة سيميلاريتي متوسطة قدره 0.29606. وأما استخدام طريقة "Word2Vec" فيؤدي إلى الحصول على درجة سيميلاريتي بمتوسط سيميلاريتي قدره 0.98449. في مرحلة استيفاء مجال المعرفة (Knowledge Area)، حصلت درجة سيميلاريتي التي تم الحصول عليها من طريقة "كوسين-سيميلاريتي" لجميع مجالات المعرفة التي تم اختبارها على متوسطة استيفاء قدرها 29.6%. وأما طريقة "Word2Vec" لجميع مجالات المعرفة التي تم اختبارها فحصلت على متوسطة استيفاء قدرها 98.44%. وفي مرحلة قياس الدقة، حصلت طريقة "كوسين-سيميلاريتي" لجميع المجالات المعرفة التي تم اختبارها على دقة متوسطة قدرها 0.6525. وأما طريقة "Word2Vec" لجميع مجالات المعرفة التي تم اختبارها فحصلت على دقة متوسطة قدرها 0.5895.

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Pada awal semester tahun pelajaran 2022/2023, Kementerian Pendidikan dan Kebudayaan Republik Indonesia mengeluarkan kebijakan Kepala BSKAP Nomor. 034/H/KR/2022 tentang Satuan Pendidikan Pelaksana Implementasi Kurikulum Merdeka pada Tahun Pelajaran 2022/2023 kepada seluruh lembaga pendidikan untuk dapat melaksanakan kurikulum merdeka. Istilah dalam pelaksanaan kurikulum itu adalah Implementasi Kurikulum Merdeka (IKM). Seluruh lembaga pendidikan di target dapat memfasilitasi menggunakan kurikulum baru tersebut dapat dilaksanakan pada bulan Juli tahun 2023. Pada Mata Pelajaran Informatika khususnya pada jenjang Sekolah Menengah Atas (SMA) kurikulum baru ini menjadi tantangan bagi guru mata pelajaran informatika untuk dapat mengembangkan kurikulum tersebut dapat terlaksana pada satuan pendidikan masing-masing. Bimbingan teknis atau pelatihan secara terjadwal dan komprehensif tidak diberikan oleh pemerintah. Pemerintah hanya membekali satu platform dengan nama Platform Merdeka Mengajar (PMM) agar para guru dapat berlatih secara mandiri, berkolaborasi serta *sharing* dengan para guru lain dan mengadopsi segala bentuk praktik baik pada platform tersebut. Tantangan ini menjadi satu tantangan yang nyata untuk menguji kompetensi para guru dalam mengembangkan kurikulum merdeka pada mata pelajaran masing-masing. Dalam konsep Islam berkaitan dengan evaluasi dan penelitian untuk membuktikan kebenaran dan keadilan dalam konteks *similarity* dokumen sebagaimana tertuang dalam Firman Allah SWT QS. Al- Hujurat Ayat 6:

نَدِيمِينَ فَعَلْتُمْ مَا عَلَىٰ فِتْنَصِبِحُوا بِجَهَالَةٍ قَوْمًا تُصِيبُوا أَنْ فَنَبِيئُونَا بِنَبَا فَاسِقٌ جَاءَكُمْ إِنْ آمَنُوا الَّذِينَ يَأْتِيهَا

*"Hai orang-orang yang beriman, jika datang kepadamu seorang fasik membawa suatu berita, maka selidikilah dengan teliti agar kamu tidak menimpakan suatu musibah kepada suatu kaum tanpa mengetahui keadaannya yang menyebabkan kamu menyesal atas perbuatanmu itu." (Qs. Al- Hujurat : 6)*

Dalam tafsir Al-Jalalain (2008) memaknai ayat yang terkandung dalam QS. Al-Hujurat Ayat 6 menurut suatu *qiraat* dibaca *Fatatsabbatuu* berasal dari lafal *Ats-Tsabaat*, artinya telitilah terlebih dahulu kebenarannya (agar kalian tidak menimpakan musibah kepada suatu kaum) menjadi *Maf'ul* dari lafal *Fatabayyanuu*, yakni dikhawatirkan hal tersebut akan menimpa musibah kepada suatu kaum (tanpa mengetahui keadaannya) menjadi hal atau kata keterangan keadaan dari *Fa'il*, yakni tanpa sepengetahuannya (yang menyebabkan kalian) membuat kalian (atas perbuatan kalian itu) yakni berbuat kekeliruan terhadap kaum tersebut (menyesal).

Dalam konteks kemiripan dokumen, tafsir ini memberikan pengertian bahwa sebelum menilai atau menyebarkan informasi tentang kemiripan dokumen, diperlukan penyelidikan dan pengecekan yang cermat untuk memastikan kebenaran dan keakuratan informasi tersebut. Hal ini penting untuk menghindari kesalahan penilaian, penyebaran informasi yang tidak benar, atau penyalahgunaan informasi.

Komunitas Guru Informatika SMA Negeri se-Kota Malang mengembangkan kurikulum ini dengan membentuk Musyawarah Guru Mata Pelajaran (MGMP) Informatika yang dalam forum pertemuan tersebut para guru informatika bersama-sama mengembangkan kurikulum yang dapat di akomodir ke satuan pendidikan masing-masing. Pada penerapannya, implementasi kurikulum merdeka pada mata pelajaran informatika yang diberikan oleh pemerintah terdapat elemen kompetensi pada mata pelajaran informatika yang tersusun dalam elemen kompetensi informatika pada Fase E dan Fase F. Fase E dipergunakan pada jenjang Kelas X, sedangkan Fase F dipergunakan pada jenjang Kelas XI dan XII. Pada pengembangan kurikulum elemen kompetensi pada masing-masing fase tersebut di diskusikan bersama dan disepakati dalam forum yang nantinya dapat dipergunakan oleh seluruh SMA Negeri se-Kota Malang.

Dalam sebuah penelitian yang ditulis oleh Gokul et al. (2017) mengemukakan pula teknik dalam mengidentifikasi parafrase pada bahasa melayu dengan pendekatan teks mining menggunakan metode *Cosine Similarity* dan *Jaccard* dengan performa *similarity* terbaik menggunakan metode *Cosine Similarity* sebesar 86%, sedangkan Ramadhanti dan Mariyah (2019) melakukan penelitian serupa untuk mengetahui *similarity* suatu dokumen artikel berbahasa indonesia dari

Wikipedia dengan menghitung kesamaan kosinus dari vektor kata menggunakan metode *Word2Vec* dengan performa *similarity* sebesar 91%. Dari dua penelitian sebelumnya, penelitian ini dilakukan untuk menguji kedua metode tersebut kemudian membandingkan hasil optimasi *similarity* serta mengevaluasi faktor, aspek dan metode yang paling efektif untuk merekomendasikan kepada pihak terkait dalam evaluasi optimasi kurikulum informatika kedepan lebih baik.

### **1.2. Pernyataan Masalah**

1. Apa saja faktor yang mempengaruhi *similarity* dokumen berbasis term dasar, gabungan term, kalimat dan paragraf pada dokumen modul ajar mata pelajaran informatika ?
2. Metode apa yang paling akurat dalam melakukan *similarity* dokumen modul ajar mata pelajaran informatika dengan kurikulum informatika yang berlaku ?

### **1.3. Tujuan Penelitian**

1. Untuk menganalisis faktor yang mempengaruhi *similarity* dokumen berbasis *term* dasar, gabungan *term*, kalimat dan paragraf pada dokumen modul ajar mata pelajaran informatika.
2. Untuk mengukur akurasi metode *similarity* dokumen modul ajar mata pelajaran informatika dalam pengembangan perangkat kurikulum informatika.

### **1.4. Batasan Masalah**

Dokumen Kurikulum Informatika yang digunakan sebagai pembanding *similarity* adalah dokumen kurikulum informatika pada jenjang Sekolah Menengah Atas (SMA) Negeri se- Kota Malang. Dokumen ini didapatkan dari dokumen yang dikembangkan oleh masing-masing guru informatika di sekolah pada jenjang Sekolah Menengah Atas (SMA) Negeri dan Swasta se- Kota Malang. Dokumen tersebut bernama RPP (Rencana Pelaksanaan Pembelajaran) atau Modul Ajar berbasis teks yang berisi pengembangan materi informatika berdasarkan kurikulum yang berlaku.

Dokumen Kurikulum yang berlaku yang digunakan dalam penelitian ini adalah dokumen Computer Science Curricula 2013. Dokumen ini didapatkan dari *Association of Computing Machinery* (ACM) berbentuk teks yang berisi pengembangan standar kurikulum informatika yang diberlakukan internasional pada link [https://www.acm.org/binaries/content/assets/education/cs2013\\_web\\_final.pdf](https://www.acm.org/binaries/content/assets/education/cs2013_web_final.pdf).

### **1.5. Manfaat Penelitian**

Hasil penelitian yang diperoleh dari *similarity* dokumen modul ajar mata pelajaran informatika dengan dokumen kurikulum informatika yang berlaku antara lain :

1. Bagi guru informatika: dapat mengetahui konten materi yang sudah dikembangkan dalam perangkat kurikulumnya apakah sudah sesuai dengan dokumen *computer science curricula* 2013.
2. Bagi satuan pendidikan (bidang akademik): dapat menjadi tolok ukur apakah konten materi kompetensi pada dokumen yang sudah dikembangkan guru informatika di sekolah sudah mencakup semua konten materi kompetensi pada dokumen *computer science curricula* 2013.
3. Bagi tim pengembang kurikulum informatika dinas pendidikan provinsi atau BSKAP Kementrian pusat: Sistem ini dapat menjadi tolok ukur kemampuan pengembangan modul ajar mata pelajaran informatika pada satuan pendidikan dan guru pada mapel informatika.

## BAB II STUDI PUSTAKA

### 2.1. *Similarity* Dokumen

Dalam penelitian yang ditulis oleh Tardan et al. (2013) disebutkan bahwa dalam melakukan pengukuran teks pada dokumen berbahasa indonesia yang telah di ringkas dalam kontek teks mining menggunakan pendekatan analisis semantik. Input data yang digunakan pada penelitian ini adalah dokumen berbahasa indonesia. Metode yang dilakukan dalam dalam penelitian ini antara lain menggunakan metode statistik dan analisis semantic yaitu metode statistik dengan *Word Frequency*, analisis semantik dengan *Euclidean*, analisis semantik dengan *Cosine Similarity* dan analisis semantik dengan *Jaccard*. Parameter pengukuran yang digunakan dalam penelitian ini terdapat tiga parameter yaitu SM (*Subjectivity Measurement*), CR (*Compression Rate*), PT (*Processing Time*). Hasil eksperimen dengan menggunakan keempat metode menunjukkan perbedaan yang signifikan dalam hal performa yang dihasilkan. Untuk metode statistik dengan *Word Frequency* dihasilkan SM sebesar 75,96%, CR sebesar 76,53% dan PT sebesar 00m01s. Sedangkan metode analisis semantik dengan *Euclidean* dihasilkan SM sebesar 85%, CR sebesar 53,27% dan PT sebesar 03m54s. Metode *Cosine Similarity* dihasilkan SM sebesar 83,46%, CR sebesar 46,66% dan PT sebesar 03m42s. Dan metode *Jaccard* dihasilkan SM sebesar 85,18%, CR sebesar 47,93% dan PT sebesar 00m03s. Dari hasil eksperimen pada parameter SM diperoleh metode terbaik menggunakan metode *Euclidean*, sedangkan pada parameter CR diperoleh metode terbaik menggunakan metode *Word Frequency* dan pada parameter PT diperoleh metode terbaik menggunakan metode *Word Frequency*.

Pada penelitian yang ditulis oleh Fuddoli et al. (2013) mengembangkan riset tentang bagaimana memproses kemiripan kata kunci suatu dokumen berita online berbahasa indonesia dengan melakukan identifikasi topik. Metode yang digunakan dalam penelitian ini menggunakan metode algoritma *Bracewell's* dan algoritma *Top-N Keywords Selection*. Pada tahap eksperimen dilakukan menggunakan tiga skenario pengukuran performa yaitu akurasi, waktu komputasi dan evaluasi

manusia. Pada eksperimen pengukuran akurasi, dataset yang di ujicoba adalah artikel <http://www.kompas.com> yang diambil mulai tahun 2011 sampai 2012 dengan data *training* sebanyak 979 item, dan data *testing* sebanyak 455 item dengan mengambil artikel dari sembilan kategori pada portal kompas ditemukan sebanyak 559 topik. Hasil pada eksperimen ini dihasilkan tingkat akurasi diatas 90%. Pada eksperimen kedua menggunakan skenario pengukuran waktu komputasi, hasil dari eksperimen ini didapatkan tingkat akurasi terbaik sebesar 72% dengan *top-N* sebanyak 20 dihasilkan waktu komputasi sebesar 2,81 detik. Pada eksperimen terakhir dengan menggunakan skenario pengukuran evaluasi manusia, dataset yang digunakan adalah ahli atau pakar 5 bahasa untuk mengevaluasi artikel tersebut. Dari ujicoba ini didapatkan hasil kecocokan dengan menggunakan pakar sebesar 71% dan menggunakan algoritma sebesar 84%. Dari ketiga skenario tersebut diatas dengan menggunakan algoritma *Bracewell's* didapatkan tingkat akurasi sebesar 95,26% pada dokumen offline dan 95,22% pada dokumen online dengan kecocokan sebesar 84% pada tahap evaluasi pakar dan rata-rata waktu komputasi sebesar 2,96 detik.

Penelitian lain dilakukan oleh Lahitani et al. (2016) mengembangkan teknik lain dalam mengolah dokumen esai teks berbahasa indonesia untuk memudahkan dalam memberikan skor atas tes esai yang dilakukan oleh pengajar pada suatu universitas. Metode tersebut menggunakan *Automated Essay Scoring* (AES) dengan dasar algoritma *Support Vector Regression* (SVR) serta metode *Cosine Similarity* untuk mengukur *similarity* dokumen yang di deteksi. Dengan dataset dari dokumen esai berbahasa indonesia yang di ujicoba dan di tes sebanyak 10 dokumen esai (dokumen 1 sampai dengan dokumen 10), dihasilkan level perangkingan *similarity* pada dokumen terbaik (memiliki tingkat *similarity* rendah) pada dokumen ke 3 sebesar 0,77% dan dokumen ke 10 memiliki tingkat *similarity* paling tinggi yaitu 39%.

Pada penelitian yang dilakukan oleh Gokul et al. (2017) mengemukakan pula teknik dalam mengidentifikasi parafrase pada bahasa melayu dengan pendekatan teks mining menggunakan metode *Cosine Similarity* dan Jaccard. Dataset yang digunakan sebanyak 2 set, yang pertama 1000 pasang kalimat parafrase dan 1500

pasang kalimat bukan parafrase dalam bahasa melayu, dan set yang kedua sebanyak 1000 pasang kalimat parafrase, 1000 pasang kalimat semi parafrase dan 1500 pasang kalimat bukan parafrase dalam bahasa melayu. Dari masing-masing dataset diambil sebanyak 900 pasang untuk di ujicobakan pada *Task 1* dan 1400 pasang di ujicobakan pada *Task 2*. Hasil ekperimen yang dilakukan dihasilkan performa *similarity* sebesar 86%. Pada *Task 1* untuk pengukuran performa presisi sebesar 0.87, recall sebesar 0.84, akurasi 86%, f1 score sebesar 0.84. Sedangkan pada *Task 2* untuk pengukuran performa presisi sebesar 0.81, recall sebesar 0.72, akurasi 78%, f1 score sebesar 0.76.

Pendekatan teks mining untuk menganalisis performa pengukuran *similarity* pada dokumen novel (Sohangir dan Wang, 2017) juga dilakukan dengan menggunakan metode *Cosine Similarity*, *ISC Similarity* dan *Gaussian Similarity*. Dengan dataset antara lain kumpulan dari sekitar 550 abstrak laporan teknis yang diterbitkan dari 1991 hingga 2007 dalam jurnal ilmu komputer di *University of Rochester*; Kumpulan data DBLP berisi judul makalah 20 tahun terakhir yang diterbitkan oleh 552 orang peneliti aktif dari sembilan bidang penelitian yang berbeda yaitu basis data, penambahan data, rekayasa perangkat lunak, teori komputer, visi komputer, sistem operasi, mesin pembelajaran, jaringan, dan pemrosesan bahasa alami; kumpulan dokumen yang muncul di kantor berita Reuters di 1987; Kumpulan data WebKB berisi 8280 dokumen yang merupakan halaman web dari berbagai departemen ilmu komputer perguruan tinggi. Dokumen-dokumen ini dibagi menjadi tujuh kelompok: mahasiswa, dosen, staf, kursus, proyek, departemen, dan lainnya. Empat yang paling populer kategori dari tujuh kategori ini dipilih dan dijadikan set WebKB4. Keempat kategori tersebut adalah mahasiswa, dosen, kursus dan proyek; dan Kumpulan data 20 Newsgroup adalah kumpulan dari sekitar 20 newsgroup yang berbeda. Dari dataset tersebut di proses menggunakan tiga metode *similarity* yaitu *Cosine Similarity*, *ISC Similarity* dan *Gaussian Similarity*. Sedangkan untuk mengukur performas kinerja metode menggunakan tiga pendekatan metode *KNN*, *Naive Bayes* dan *SVM*. Dari penelitian tersebut dihasilkan performa akurasi terbaik pada metode *Naive Bayes* dengan *Cosine Similarity* sebesar 86.33% dan AUC sebesar 88.92%.

Pendekatan lain dalam penelitian (Slamet et al., 2018) untuk melakukan teks mining peringkasan teks otomatis menggunakan *Vector Space Model*. Subyek yang diteliti adalah dokumen karya ilmiah atau artikel dengan metode tersebut dilakukan peringkasan teks otomatis untuk menghasilkan suatu ringkasan seperti abstrak. Data yang digunakan untuk penilaian adalah artikel yang diambil dari jurnal informatika yang ditulis dalam bahasa Indonesia. Jurnal memiliki kriteria tertentu seperti penghilangan header dan footer, judul, sub-judul, dan tabel dan gambar. Jenis file jurnal adalah PDF. Data test diambil sebanyak lima dokumen (D1 sampai dengan D5). Hasil eksperimen yang dilakukan dihasilkan kemiripan terbanyak terdapat pada dokumen D1 dan D4 dengan empat kesamaan sedangkan yang paling sedikit terletak pada dokumen D2 dimana tidak ada kesamaan sama sekali. Dengan demikian, tingkat akurasi dari sistem, khususnya dalam pembuatan abstrak otomatis, masih kalah dibandingkan dengan yang manual. Dengan metode *Vector Space Model* diinterpretasikan berdasarkan akurasi tingkat kemiripan yang hanya mencapai 40% pada setiap dokumen yang di tes.

Studi penelitian lain berkaitan dengan deteksi kemiripan konten suatu dokumen penugasan siswa (Soyusiawaty dan Zakaria, 2018) juga dilakukan menggunakan metode *Rochio*. Dengan dataset dokumen penugasan siswa sebanyak sepuluh dokumen (D1 sampai dengan D10) dalam ujicobanya dihasilkan skor *Rochio* terbaik pada dokumen ke 9 sebesar 4.25 dengan presentasi sebesar 0.00%. Dengan tingkat skor *Rochio* yang rendah, maka dipastikan bahwa dokumen tersebut memiliki tingkat kemiripan plagiasi dengan dokumen lain sangat rendah atau unik. Dari hasil skor *Rochio* yang didapatkan di hitung rata-rata error dan akurasinya didapatkan tingkat akurasi sebesar 74.6% dan *error similarity* sebesar 25.4%.

Pada penelitian berkenaan dengan teks mining untuk mengidentifikasi tes dalam bentuk esai yang dilakukan oleh Amalia et al. (2019), juga diteliti kembali oleh Amalia et al., 2019 dengan pengembangan konten yang sedikit berbeda pada platform e-learning dan dengan metode yang berbeda. Metode yang digunakan adalah *Latent Semantic Analysis (LSA)*. Metode ini digunakan untuk mengukur relevansi teks esai yang dijawab siswa dengan kunci yang dibuat oleh guru. Skenario ujicoba dilakukan sebanyak tiga skenario. Pada skenario 1 untuk

menghitung tingkat kesamaan 100%, atau jawaban siswa mirip dengan kunci jawaban, sedangkan pada skenario 2 untuk Kemiripan sinonim, atau istilah jawaban siswa tidak mirip dengan kunci jawaban, tetapi istilah-istilah ini memiliki arti yang sama. Dan pada skenario 3 untuk tingkat kemiripan 0%, atau jawaban siswa berbeda dengan kunci jawaban. Hasil dari ujicoba dihasilkan akurasi sistem ini lebih baik bila dibandingkan penilaian manual guru adalah 83,3%.

Ramadhanti dan Mariyah, (2019) melakukan penelitian serupa untuk mengetahui *similarity* suatu dokumen artikel berbahasa indonesia dari Wikipedia dengan menghitung kesamaan kosinus dari vektor kata. Penelitian ini mengusulkan dua teknik untuk menghitung kesamaan yang merupakan perbandingan simultan dan parsial. Metode yang dilakukan untuk mendukung kedua teknik tersebut menggunakan model *Word2vec*. Pada tahap simulasi digunakan dua puluh dokumen (Doc 1 sampai dengan Doc 20). Hasil dari masing-masing simulasi didapatkan rata-rata *similarity* dokumen sebesar 91%.

Penelitian yang dilakukan oleh Faisal et al., (2020), pada studi kasus plagiarism dokumen abstrak, peneliti menggunakan dua metode dalam membandingkan hasil *similarity*-nya menggunakan algoritma Winnowing dan Manber. Hasil akurasi yang didapatkan dengan rata-rata kemiripan dokumen sebesar 90.56% dengan algoritma Manber, 94% dengan algoritma Winnowing dan 91.22% dengan algoritma Winnowing triword.

Pada penelitian yang dilakukan oleh Rosid et al., (2020), pada studi kasus klasifikasi dokumen pengaduan siswa, peneliti menggunakan lima puluh (50) dokumen berbahasa indonesia. Pokok riset yang dilakukan peneliti pada aspek *preprocessing text* dengan menggunakan algoritma *Porter Stemmer* dan Sastrawi. Hasil dari penelitian tersebut menunjukkan bahwa pada kategori stemmer *Exact Match* menghasilkan 82% untuk *Porter stemmer* dan 92% untuk Sastrawi, sedangkan pada kategori *Unchange* 4% untuk *Porter stemmer* dan 6% untuk Sastrawi, pada kategori *Under Stemming* sebesar 4% untuk *Porter stemmer* dan 2% untuk sastrawi, dan pada kategori *Overstemming* sebesar 10% untuk *Porter stemmer* dan 0% untuk Sastrawi. Berdasarkan waktu pengerjaannya, dengan menggunakan *library* sastrawi adalah lebih cepat dalam melakukan proses

*stemming* yaitu rata-rata 0,6 detik sedangkan menggunakan algoritma porter membutuhkan waktu rata-rata 241,6 detik. Sehingga dari penelitian ini dengan menggunakan algoritma *library* sastrawi lebih unggul dalam melakukan *stemming* dokumen berbahasa indonesia dengan tingkat akurasi sebesar 92%.

Studi kasus lain pada penerapan text mining dokumen (Madyatmadja et al., 2022), peneliti melakukan riset dengan memanfaatkan *citizen report* dari dua media sosial yang bersumber Pemerintah Kota Tangerang, Indonesia yaitu melalui fitur komentar atau inbox yang berisi pengaduan masyarakat. Fitur komentar yang ada berisi 9865 kejadian dan memiliki 320 kategori. Kemudian kumpulan data yang dikumpulkan dan diklasifikasikan menjadi empat (4) kategori utama yaitu bencana, infrastruktur, sosial, dan urusan bangsa. Metode yang digunakan dalam melakukan text mining ini menggunakan *DT*, *k-NN*, *SVM*, *Random Forest* dan *Ensemble*. Hasil pengujian yang dilakukan untuk presisi terbaik pada *framework* Cor-Tp + Con-TP menggunakan metode *SVM-Polynomial* sebesar 77,01%.

Pada penelitian yang ditulis oleh Diartono et al., (2022) yang meneliti tentang *similarity* dokumen teks berbahasa indonesia berbasis kata majemuk. Dimana didalam dokumen dilakukan *similarity* untuk menghasilkan tingkat kemiripan dokumen. Metode yang diusulkan pada penelitian ini adalah *Jaccard Similarity*. Input data yang digunakan adalah sebanyak 15 dokumen teks berbahasa indonesia. Dalam penelitian ini ditemukan kata majemuk pada 10 dokumen. Sedangkan dokumen yang tidak memiliki kata majemuk sebanyak 5 dokumen. Hasil pengukuran *similarity* pada dokumen tersebut menghasilkan tingkat *similarity* yang tinggi namun dalam prosesnya tidak ditemukan kata majemuk sebanyak 33.3%. Kemudian hasil pengukuran *similarity* yang memiliki tingkat lebih rendah karena tidak memiliki kata majemuk sebanyak 66.3%.

## **2.2. Standar Kurikulum Informatika**

Pada sektor pendidikan sangat memegang peranan penting dalam pengembangan kompetensi. Dalam proses pendidikan itu di implementasikan dalam bentuk proses belajar mengajar dengan tujuan mulia untuk mencerdaskan generasi muda (Mubai et al., 2020; Qolbi & Hamami, 2019). Dalam pengembangan

untuk mendukung proses pendidikan dibutuhkan perencanaan dan aturan yang baku dalam implementasinya (Astawa, 2019). Aturan yang baku tersebut dinamakan Kurikulum yang berlaku. Dalam pengembangan Kurikulum Informatika pada jenjang Sekolah Menengah Atas (SMA) sudah dilaksanakan sejak Kurikulum Satuan Pendidikan (KTSP) diberlakukan pada tahun 2006. Dimana Konten Kurikulum Informatika yang disajikan memuat kompoetensi dasar dalam ilmu komputer atau informatika.

Seiring berkembangnya waktu, konten kurikulum informatika dikembangkan lebih spesifik pada tahun 2022 yaitu dengan dikembangkannya Platform Merdeka Belajar (PMM), dimana pada era tersebut kurikulum informatika dikembangkan dalam bentuk elemen-elemen kompetensi informatika antara lain: Keterampilan Generik, Berpikir Komputasional, Teknologi Informasi dan Komunikasi, Sistem Komputer, Jaringan Komputer dan Internet, Analisis Data, Algoritma dan Pemrograman, dan Dampak Sosial Informatika. Dari delapan elemen kompetensi pada kurikulum informatika tersebut, pedoman dan standarnya ditujukan untuk penyelarasan dalam perancangannya, pembuatannya dan penggunaan kurikulum informatika yang berlaku tersebut, dimana pada pengembangan kurikulum yang standar digunakan untuk memberikan pengarah dan meminimalisir pengembangan yang tidak sesuai dengan kurikulum yang berlaku. Dalam metode dalam mengevaluasi pembelajaran menggunakan dua tahap antara lain analisis kebijakan pemerintah dan evaluasi pembelajaran berbasis praktik pada sistem model pembelajaran (C. Crysdiyan, 2022).

Pada standar kurikulum informatika yang berlaku di tingkat internasional, pengembangan standar kurikulum tersebut salah satunya dilakukan oleh *Association For Computing Machinery* (ACM). Asosiasi ini fokus dalam pengembangan konten ilmu komputer salah satunya kurikulum informatika yang digunakan sebagai dasar atau acuan dalam pengembangan kurikulum informatika pada jenjang Sekolah Menengah Atas (SMA) di Indonesia. Konten kurikulum yang disajikan dalam bentuk *Knowledge Area* (KA). KA tersebut antara lain: *AL-Algorithms and Complexity*, *SP-Social Issues and Professional Practice*, *SF-Systems Fundamentals*, *SE-Software Engineering*, *SDF-Software Development*

*Fundamentals, PL-Programming Languages, PD-Parallel and Distributed Computing, PBD-Platform-based Development, OS-Operating Systems, NC-Networking and Communication, IS-Intelligent Systems, IM-Information Management, IAS-Information Assurance and Security, HCI-Human-Computer Interaction, GV-Graphics and Visualization, DS-Discrete Structures, CN-Computational Science, dan AR-Architecture and Organization.*

### 2.3. Ekstraksi Fitur

Ekstraksi fitur yang dilakukan pada penelitian ini antara lain adalah tahap *preprocessing* teks. Dimana pada tahap ini dilakukan tahapan antara lain *case folding*, *filtering*, *stemming*, dan *tokenizing* (Gokul et al. (2017)). *Case folding* digunakan untuk mengubah dokumenteks menjadi teks huruf kecil (*lowercase*). Sedangkan *filtering* digunakan untuk menghapus kata tidak penting (*stop word*). Untuk proses *stemming* digunakan untuk menghapus awalan dan akhiran kata menjadi term baku atau kata dasar. Sedangkan tahap *tokenizing* digunakan untuk memisahkan kalimat menjadi potongan kata.

### 2.4. Dokumen TF-IDF

Model perhitungan dengan metode TF-IDF digunakan untuk merepresentasikan teks sebagai vektor yang terdiri dari n istilah kata berbobot yang muncul dalam teks seperti rumus persamaan (1-3) berikut: (F. Chen et al., (2022)

$$TF - IDF(w_i) = tf(w_i) \times idf(w_i) = tf(w_i) \times \log\left(\frac{N}{df(w_i)}\right), \quad (2.1)$$

$$tf(w_i) = \frac{n_{ij}}{\sum_{k=1}^m n_{kj}}, \quad (2.2)$$

$$idf(w_i) = |\{j : w_i \in d_j\}| \quad (2.3)$$

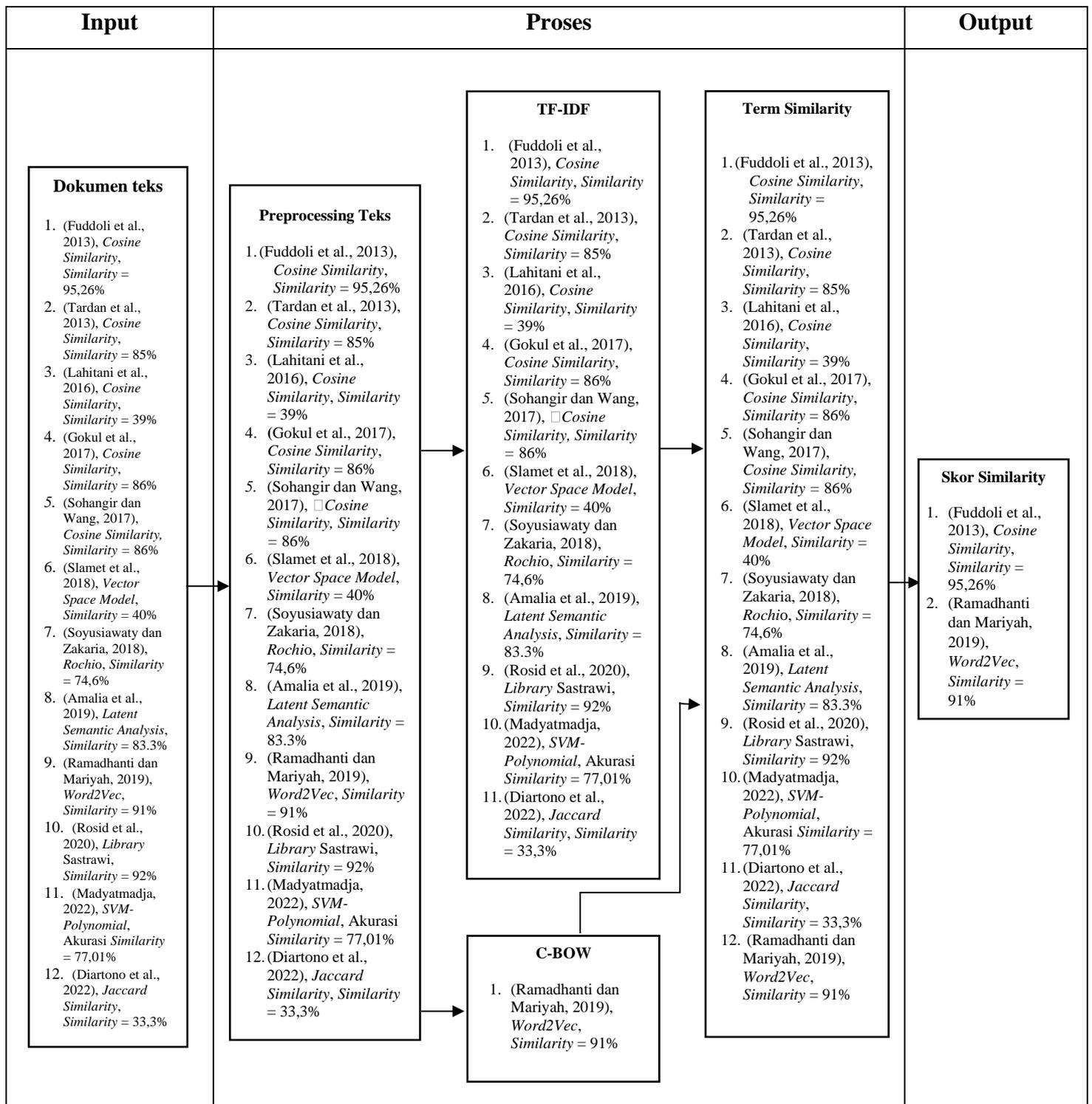
Dimana  $tf(w_i)$  adalah frekuensi kemunculan  $w_i$  dalam teks  $j$ , dan  $N$  adalah jumlah seluruh teks dalam kumpulan teks  $\{d_j\}$ .  $df(w_i)$  menunjukkan berapa banyak teks yang menunjukkan istilah  $w_i$  dalam kumpulan teks.  $n_{ij}$  adalah frekuensi kemunculan istilah ke- $i$  yang muncul dalam dokumen ke- $j$ .  $n_{kj}$  adalah frekuensi kemunculan suku ke- $k$  yang muncul pada dokumen ke- $j$ .  $|\{j : w_i \in d_j\}|$  TF-IDF

adalah jumlah dokumen termasuk term ke- $i$ . Nilai setiap istilah dalam setiap teks dapat diperoleh dengan menganalisis setiap istilah dalam kumpulan teks, dan model vektor setiap teks dibangun dengan nilai istilah TF-IDF. Sehingga, kemiripan teks dapat ditentukan dengan menghitung cosine similarity atau Word2Vec. Kemunculan kata dalam TF-IDF dapat didefinisikan sebagai berikut:

1. Semakin sering sebuah kata muncul dalam sebuah teks, semakin relevan kata tersebut dengan topik teks tersebut.
2. Semakin sering suatu istilah muncul di beberapa teks dalam kumpulan teks, semakin buruk istilah tersebut.

## **2.5. Kerangka Teori**

Pada sub bab ini disajikan kerangka teori dari hasil studi literatur dengan cara mengkategorikan parameter atau variabel sebagai input data kerangka teori dan mencantumkan metode apa yang digunakan dan apa keluaran dari hasil kerangka teori tersebut. Kerangka teori dapat dilihat pada Gambar 2.1.



Gambar 2.1. Kerangka Teori

Gambar 2.1 menjelaskan bahwa data masukan yang digunakan adalah dalam dokumen dalam bentuk kata, kalimat dan paragraf. Input data yang digunakan adalah dokumen *computer science curricula 2013* dan dokumen kurikulum informatika jenjang SMA. Dari dokumen *computer science curricula 2013* adalah dokumen yang diproses sebagai sumber dokumen pembandingan dan dokumen kurikulum informatika jenjang SMA adalah dokumen yang akan diujicobakan pada penelitian ini.

Tahapan berikutnya adalah menjabarkan dan memilih metode-metode terbaik yang telah didapatkan dari studi literatur dari penelitian yang sudah ada sebelumnya dalam melakukan *similarity* dokumen. Beberapa penelitian atau jurnal yang dijadikan acuan dalam penelitian ini dapat di lihat pada Tabel 2.1 berikut:

Tabel 2.1. Daftar Jurnal Penelitian

No.	Nama Penulis dan Tahun	Metode	Judul	Hasil Penelitian
1	Fuddoli <i>et al.</i> (2013)	<ul style="list-style-type: none"> <li>• <i>Cosine Similarity</i></li> <li>• <i>Top N Keyword Selection</i></li> </ul>	<i>Keywords Similarity Based Topic Identification for Indonesian News Documents</i>	Skor Similarity 95,26%
2	Tardan <i>et al.</i> (2013)	<ul style="list-style-type: none"> <li>• <i>Euclidean Distance</i></li> <li>• <i>Word Frequency</i></li> <li>• <i>Jaccard</i></li> <li>• <i>Cosine Similarity</i></li> </ul>	<i>Automatic text summarization based on semantic analysis approach for documents in Indonesian language</i>	Skor Similarity 85%
3	Lahitani <i>et al.</i> (2016)	<ul style="list-style-type: none"> <li>• <i>Cosine Similarity</i></li> <li>• <i>Automated Essay Scoring (AES)</i></li> <li>• <i>Support Vector Regression (SVR)</i></li> </ul>	<i>Cosine similarity to determine similarity measure: Study case in online essay assessment</i>	Skor Similarity 39%
4	Gokul <i>et al.</i> (2017)	<ul style="list-style-type: none"> <li>• <i>Cosine Similarity</i></li> <li>• <i>Jaccard</i></li> </ul>	<i>Sentence similarity detection in Malayalam language using cosine similarity</i>	Skor Similarity 86%
5	Sohangir dan Wang, (2017)	<ul style="list-style-type: none"> <li>• <i>Cosine Similarity</i></li> <li>• <i>ISC Similarity</i></li> <li>• <i>Gaussian Similarity</i></li> </ul>	<i>Improved sqrt-cosine similarity measurement</i>	<ul style="list-style-type: none"> <li>• Skor Similarity 86.33%</li> <li>• AUC 88.92%</li> </ul>

No.	Nama Penulis dan Tahun	Metode	Judul	Hasil Penelitian
6	Slamet <i>et al.</i> (2018)	<i>Vector Space Model</i>	<i>Automated Text Summarization for Indonesian Article Using Vector Space Model</i>	Skor Similarity 40%
7	Soyusiawaty dan Zakaria, (2018)	<i>Rochio</i>	<i>Similarity Detection of Student Assignments Using Rocchio Method</i>	Skor Similarity 74.6%
8	Amalia <i>et al.</i> (2019)	<i>Latent Semantic Analysis (LSA)</i>	<i>Automated Bahasa Indonesia essay evaluation with latent semantic analysis</i>	Skor Similarity 83.3%
9	Ramadhanti dan Mariyah, (2019)	<i>Word2Vec</i>	<i>Document Similarity Detection Using Indonesian Language Word2vec Model</i>	Skor Similarity 91%
10	Rosid et al., (2020)	<ul style="list-style-type: none"> <li>• Porter Stemmer</li> <li>• Sastrawi</li> </ul>	<i>Improving Text Preprocessing For Student Complaint Document Classification Using Sastrawi</i>	Skor Similarity 92%
11	Madyatmadja et al., 2022	<ul style="list-style-type: none"> <li>• DT</li> <li>• k-NN</li> <li>• SVM</li> <li>• Random Forest</li> <li>• Ensemble</li> </ul>	<i>Contextual Text Analytics Framework for Citizen Report Classification: A Case Study Using the Indonesian Language</i>	Akurasi 77,01%
12	Diartono et al., 2022	<i>Jaccard Similarity</i>	Peningkatan Hasil Sistem Temu Kembali Informasi Berbasis Kata Majemuk Menggunakan Jaccard Similarity	Skor Similarity 33.3%

Pada tahap studi literatur, parameter performa yang dihasilkan dari masing-masing literatur adalah hasil akurasinya. Pertama, dengan membandingkan performa akurasi dari masing-masing studi literatur sebelumnya yang dipilih dengan berbagai metode yang diusulkan. Pada penelitian Fuddoli et al. (2013), dengan mengusulkan metode algoritma *Bracewell's* didapatkan skor *similarity* sebesar 95,26%. Pada penelitian yang dilakukan oleh Tardan et al. (2013) dengan

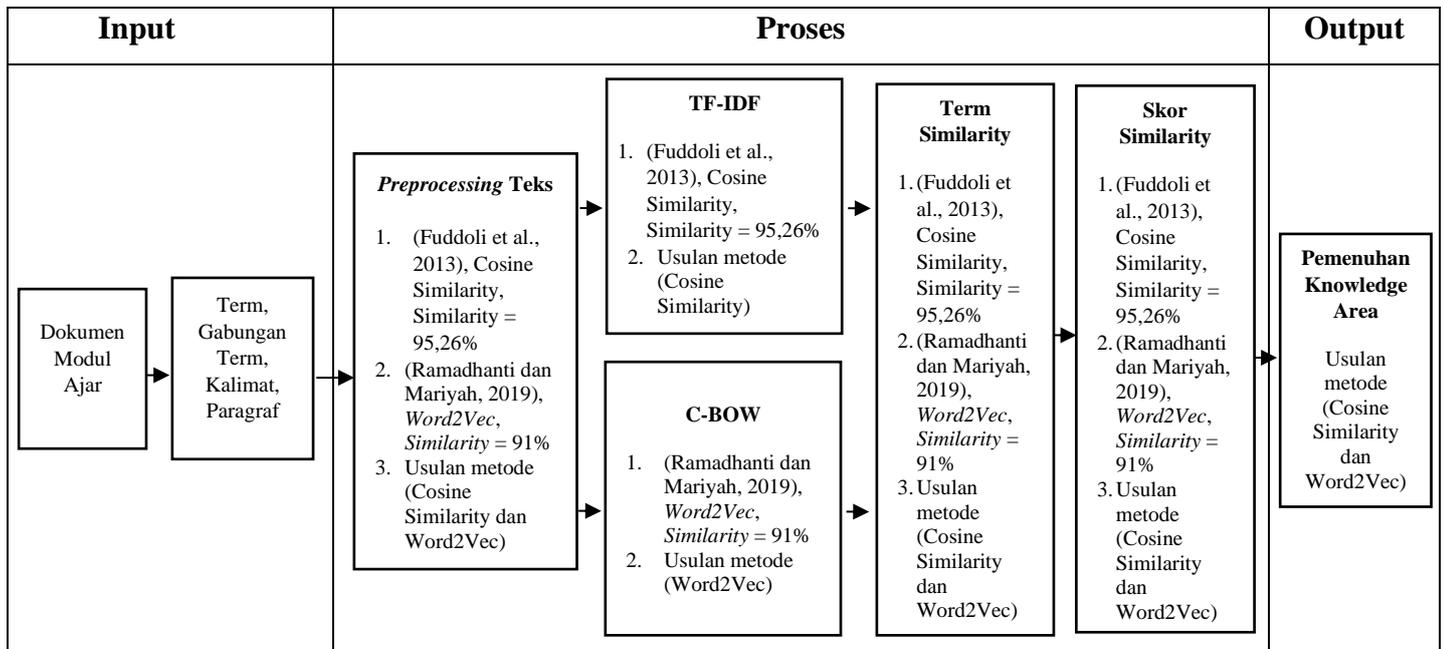
mengusulkan metode Euclidean didapatkan skor *similarity* sebesar 85%. Sedangkan pada penelitian Lahitani et al. (2016) dalam melakukan cek kemiripan teks menggunakan metode *Cosine Similarity* didapatkan skor *similarity* sebesar 39%. Pada penelitian yang dilakukan oleh Gokul et al. (2017) menggunakan metode *Cosine Similarity* didapatkan skor *similarity* sebesar 86%. Kemudian berbeda dengan penelitian yang dilakukan oleh Sohangir dan Wang (2017) dengan metode yang diusulkan menggunakan metode *Cosine Similarity* didapatkan hasil uji coba skor *similarity* sebesar 86,33%. Sedangkan pada penelitian yang dilakukan oleh Slamet et al. (2018) menggunakan metode *Vector Space Model* didapatkan skor *similarity* sebesar 40%. Untuk penelitian yang dilakukan oleh Soyusiawaty dan Zakaria (2018) dengan menggunakan metode *Rochio*, hasil uji coba yang dilakukan menghasilkan skor *similarity* sebesar 74,6%. Dan pada penelitian yang dilakukan oleh Amalia et al. (2019) dengan metode *Latent Semantic Analysis* dihasilkan skor *similarity* sebesar 83.3%. Pada penelitian yang dilakukan oleh Ramadhanti dan Mariyah (2019) dengan mengusulkan metode *Word2Vec* didapatkan skor *similarity* sebesar 91%. Berbeda dengan penelitian yang dilakukan oleh Rosid et al. (2020) dengan menggunakan *library Sastrawi* didapatkan skor *similarity* sebesar 92% dan pada studi literatur yang dikemukakan oleh Madyatmadja (2022) dengan menggunakan metode *SVM-Polynomial* didapatkan akurasi *similarity* sebesar 77,01%. pada studi literatur penelitian Diartono et al., (2022) yang meneliti tentang *similarity* dokumen teks berbahasa indonesia berbasis kata majemuk dengan metode *Jaccard Similarity* dihasilkan skor *similarity* sebesar 33,3%.

Dari studi literatur dalam mengukur *similarity* dokumen yang dilakukan diambil dua metode dengan performa terbaik. Hasil perbandingan performa *similarity* terbaik yang dapat dilakukan untuk mendukung penelitian ini adalah pada penelitian Gokul et al. (2017) menggunakan metode *Cosine Similarity* dengan skor *similarity* sebesar 86% dan pada penelitian Sohangir dan Wang (2017) menggunakan metode *Word2Vec Similarity* dengan skor *similarity* sebesar 91% dan untuk menganalisis perbandingan *similarity*, penulis mengusulkan dua metode untuk penelitian, yaitu metode algoritma *Cosine Similarity* dan *Word2Vec Similarity*. Dengan melakukan penelitian menggunakan kedua metode tersebut

diharapkan dapat menghasilkan hasil terbaik untuk mengoptimasi dalam evaluasi pengembangan dokumen modul ajar mata pelajaran informatika jenjang SMA di Kota Malang.

## BAB III DESAIN PENELITIAN

### 3.1. Kerangka Konsep



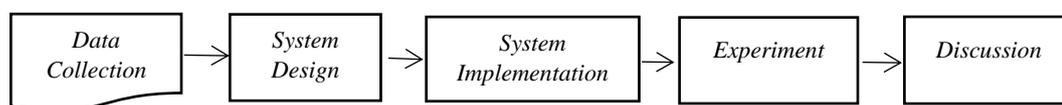
Gambar 3.1. Kerangka Konsep

Gambar 3.1 menjelaskan kerangka konsep yang dijalankan pada penelitian ini, dimana dokumen input adalah dokumen teks modul ajar informatika. Kemudian langkah selanjutnya adakah input dokumen dibagi menjadi tiga bagian yaitu term dasar, gabungan term, gabungan kalimat dan gabungan paragraf. Langkah selanjutnya pada tahap ini penelitian mengusulkan dua metode yang dapat dilakukan yaitu metode Cosine similarity dan *Word2Vec*. Langkah yang dilakukan pada tahap pertama adalah dengan melakukan *proprocessing* teks. Langkah berikutnya metode cosine melakukan perhitungan frekuensi term dengan menggunakan TF-IDF sedangkan metode *Word2Vec* melakukan proses C-BOW, kemudian melakukan proses perhitungan TF-IDF. Setelah kedua selesai menghitung TF-IDF, langkah selanjutnya adalah melakukan perhitungan term similarity sehingga didapatkan skor similarity dari masing-masing metode. Dan

tahap akhir setelah itu adalah dengan melakukan proses pemenuhan *Knowledge Area*.

### 3.2. Prosedur Penelitian

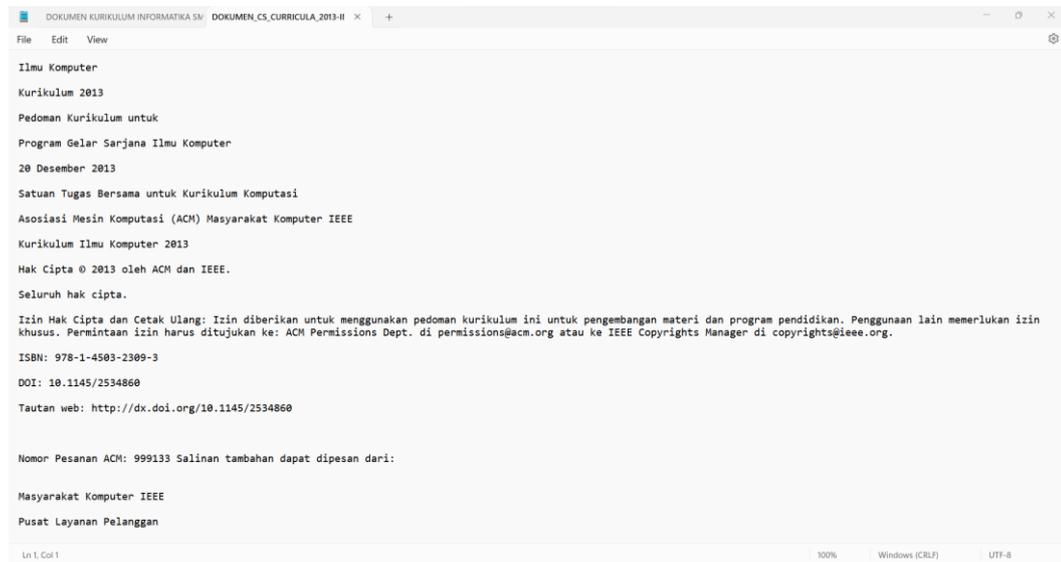
Berdasarkan hasil studi literatur dan mendeskripsikannya dalam kerangka teori, langkah selanjutnya adalah melakukan desain penelitian. Pada tahap penelitian desain terdapat beberapa tahapan proses diantaranya pengumpulan data, desain sistem, eksperimen, dan pembahasan. Tahapan-tahapan tersebut akan diperlihatkan pada Gambar 3.2.



Gambar 3.2. Prosedur Penelitian

#### 3.2.1. Pengumpulan Data

Pada tahap pengumpulan data, data yang diambil adalah data primer dalam bentuk dokumen dalam bentuk teks *Computer Science Curricula 2013* dan dokumen dalam bentuk teks Kurikulum Informatika Jenjang SMA. Untuk dokumen *Computer Science Curricula 2013* diambil dari laman Association for Computing Machinery (ACM) pada url [https://www.acm.org/binaries/content/assets/education/cs2013\\_web\\_final.pdf](https://www.acm.org/binaries/content/assets/education/cs2013_web_final.pdf). kemudian dokumen tersebut di terjemahkan menjadi teks berbahasa indonesia. Sedangkan dokumen kurikulum informatika jenjang SMA diambil dari dokumen yang dikembangkan oleh guru informatika pada SMA Negeri se-Kota Malang yang tergabung dalam Musyawarah Guru Mata Pelajaran (MGMP) Informatika SMA Negeri Kota Malang. Tahapan pengumpulan data dijelaskan pada contoh dokumen berbentuk teks yang diambil dari dokumen *Computer Science Curricula 2013* dan dokumen modul ajar mata pelajaran informatika jenjang SMA dapat dilihat pada Gambar 3.3.

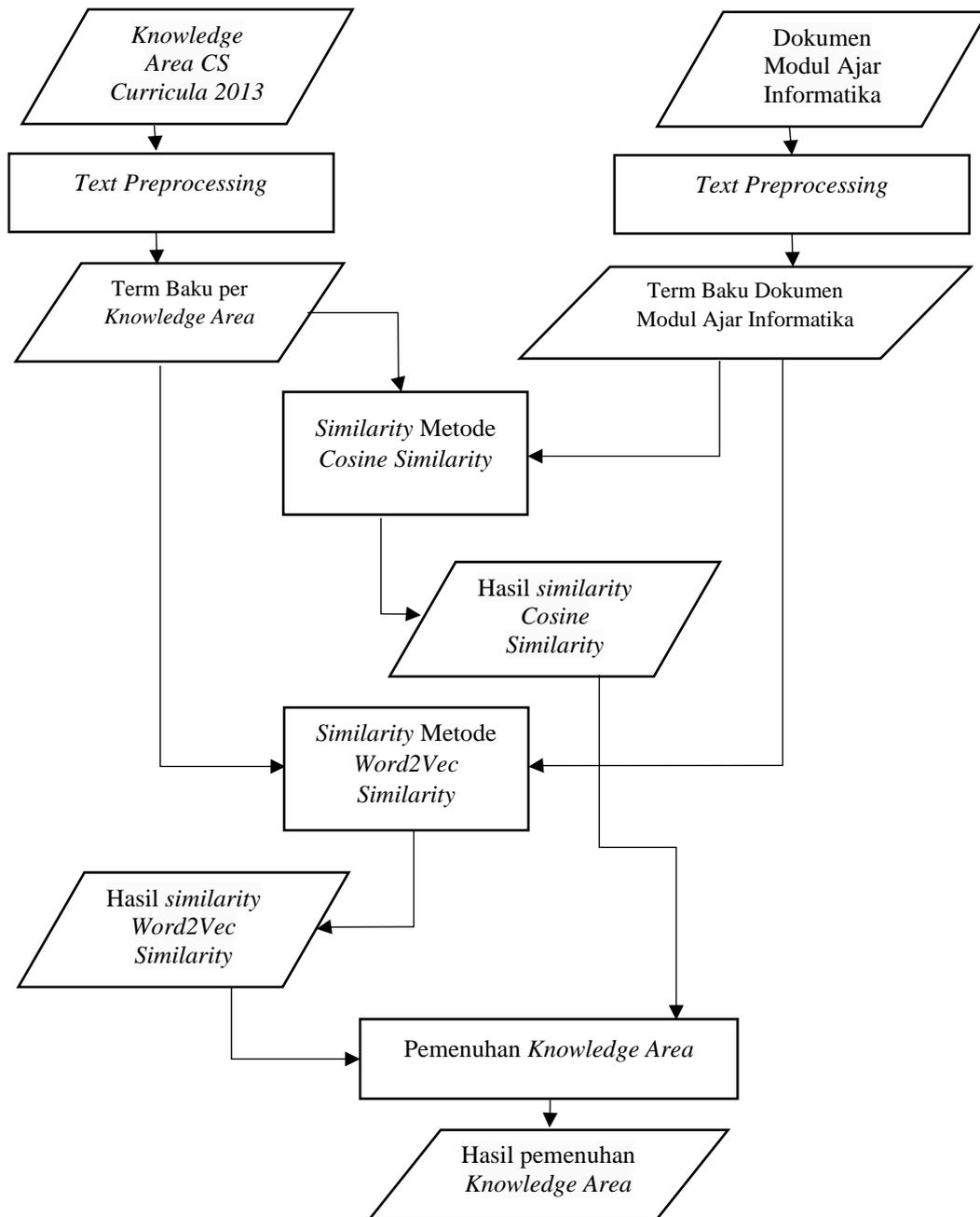


Gambar 3.3. Contoh dokumen teks Computer Science Curricula 2013 dan Dokumen Modul Ajar Mata Pelajaran Informatika jenjang SMA

Gambar 3.3 menjelaskan dokumen *Computer Science Curricula 2013* yang berbentuk pdf di terjemahkan menjadi berbahasa indonesia dan sedangkan dokumen modul ajar mata pelajaran informatika yang didapatkan sudah berbentuk teks. Dari kedua sumber dokumen tersebut di kumpulkan dan disimpan pada aplikasi notepad.

### 3.2.2. Desain Sistem

Pada tahap perancangan sistem, sistem yang akan dirancang pada penelitian ini terdiri dari tiga tahap yaitu *preprocessing text*, proses *similarity* dan proses penghitungan pemenuhan dokumen modul ajar mata pelajaran informatika SMA dengan *Knowledge Area*. Tahapan dalam perancangan sistem akan dijelaskan pada Gambar 3.4.



Gambar 3.4. Desan Sistem

### 3.2.2.1. Input Dokumen *Computer Science Curricula 2013*

Pada subbab ini, input data dokumen kesatu berupa dokumen teks berbahasa Inggris yang sudah diterjemahkan menjadi teks bahasa Indonesia terdapat delapan belas *Knowledge Area* antara lain *AL-Algorithms and Complexity*, *SP-Social Issues and Professional Practice*, *SF-Systems Fundamentals*, *SE-Software*

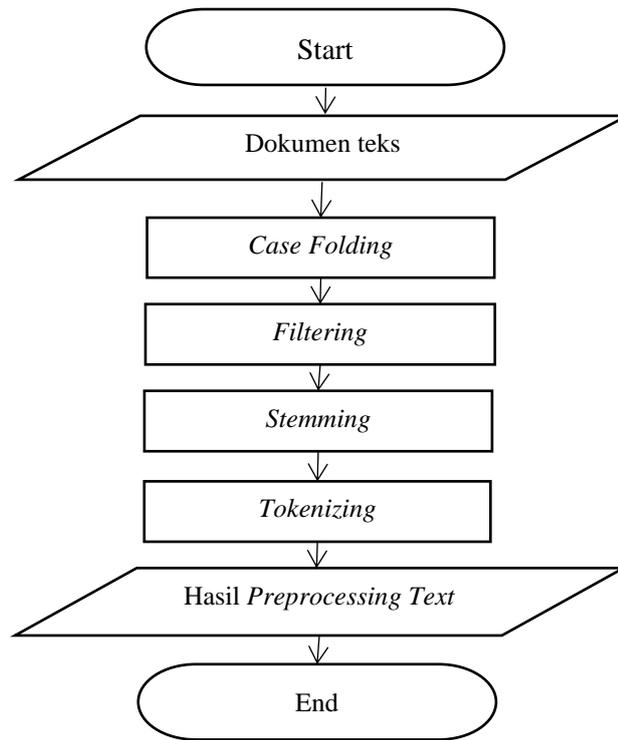
*Engineering, SDF-Software Development Fundamentals, PL-Programming Languages, PD-Parallel and Distributed Computing, PBD-Platform-based Development, OS-Operating Systems, NC-Networking and Communication, IS-Intelligent Systems, IM-Information Management, IAS-Information Assurance and Security, HCI-Human-Computer Interaction, GV-Graphics and Visualization, DS-Discrete Structures, CN-Computational Science, dan AR-Architecture and Organization.*

### **3.2.2.2. Input Dokumen Modul Ajar Mata Pelajaran Informatika**

Pada subbab ini, input data dokumen kedua berupa dokumen teks berbahasa indonesia. Dokumen modul ajar tersebut diperoleh dari guru mata pelajaran informatika SMA Negeri dan Awasta se-Kota Malang. Dokumen tersebut adalah dokumen modul ajar informatika yang telah dikembangkan oleh masing-masing guru informatika pada kegiatan Musyawarah Guru Mata Pelajaran (MGMP) Informatika SMA Negeri dan Swasta se-Kota Malang.

### **3.2.2.3. Proses Teks *Preprocessing***

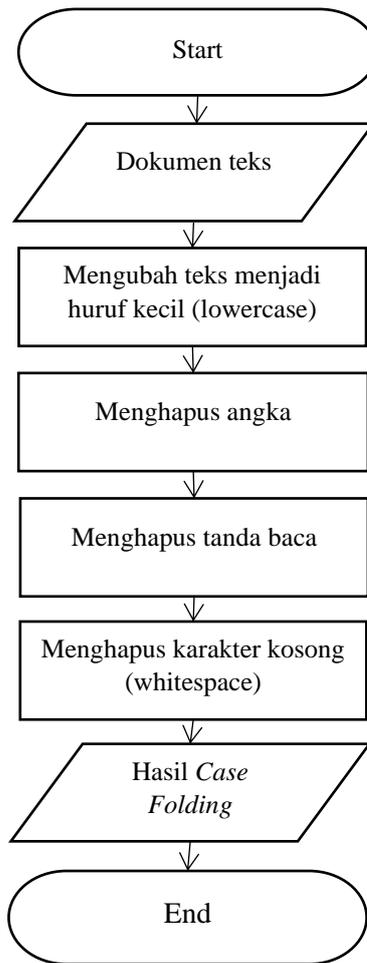
Pada subbab ini membahas tentang tahapan teks processing yang dilakukan dari input data dokumen kesatu dan dokumen kedua. Tahapan tersebut antara lain : proses *case folding*, proses *tokenizing*, proses *filtering* dan proses *stemming*. Tujuan pada tahap preprocessing ini adalah untuk menormalisasi sumber teks yang akan diolah pada proses *similarity* dokumen, sehingga dokumen yang akan diproses pada proses *similarity* dapat di minimalisir bentuk term dasarnya. Pada proses Teks Preprocessing dapat dilihat pada Gambar 3.5.



Gambar 3.5. Proses Teks *Preprocessing*

#### A. Tahap *Case Folding*

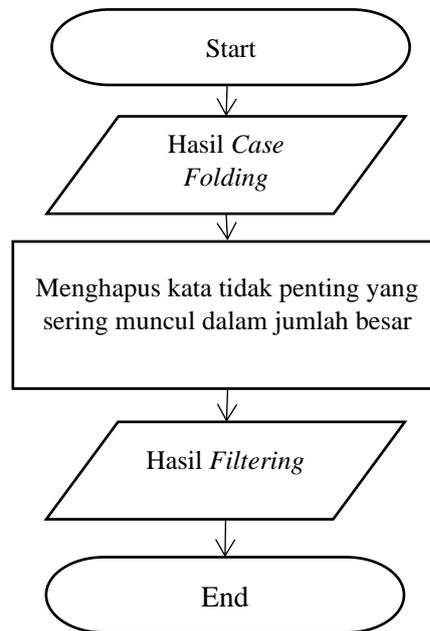
Pada tahap *case folding* beberapa tahap yang dilakukan antara lain mengubah teks menjadi huruf kecil (*lowercase*) yang dapat dilihat pada Gambar 3.6. Pada tahap *case folding*, tahapan yang dilakukan antara lain yaitu mengubah teks menjadi huruf kecil (*lower case*) kemudian dilakukan proses menghapus angka, menghapus tanda baca dan yang terakhir dilakukan proses menghapus karakter kosong (*whitespace*).



Gambar 3.6. Proses *Case Folding*

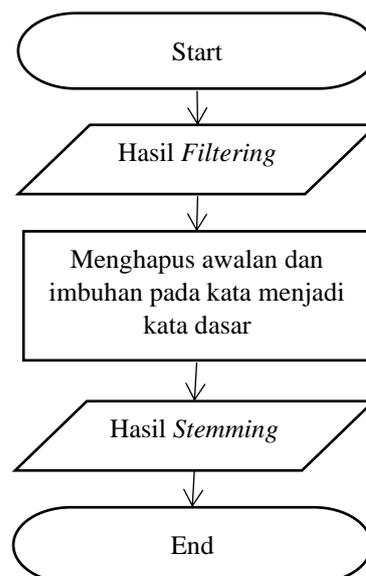
## B. Tahap *Filtering*

Pada tahap *filtering* yang dilakukan dapat dilihat pada Gambar 3.7. Pada tahap *filtering*, tahapan yang dilakukan yaitu dengan menghapus kata tidak penting yang sering muncul dalam jumlah besar (*stop word*).

Gambar 3.7. Proses *Filtering*

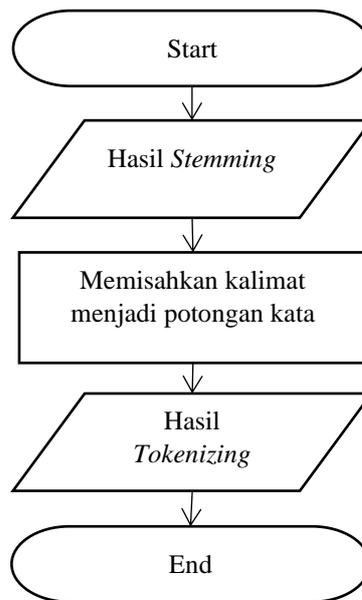
### C. Tahap *Stemming*

Pada tahap *stemming* yang dilakukan dapat dilihat pada Gambar 3.8. Pada tahap *stemming*, tahapan yang dilakukan yaitu dengan menghapus awalan dan imbuhan kata menjadi term baku atau kata dasar.

Gambar 3.8. Proses *Stemming*

#### D. Tahap *Tokenizing*

Pada tahap *tokenizing* yang dilakukan dapat dilihat pada Gambar 3.9. Pada tahap *tokenizing*, tahapan yang dilakukan yaitu dengan memisahkan kalimat menjadi potongan kata.



Gambar 3.9. Proses *Tokenizing*

#### 3.2.2.4. Proses *Similarity* Dokumen

Pada tahap *similarity* dokumen dari hasil *preprocessing text*, sistem dapat melakukan proses *similarity* antara lain: dengan menggunakan metode *Cosine Similarity* dan *Word2Vec* yang akan dijelaskan pada subbab 3.2.2.4.1 dan 3.2.2.4.2.

##### 3.2.2.4.1. Proses *Similarity* Dokumen Menggunakan Metode *Cosine Similarity*

Metode *Cosine similarity* adalah metode yang digunakan untuk mengukur jarak pada data dalam ukuran vektor. Dimana data tersebut dapat berisi ratusan bahkan ribuan term atau kata yang memiliki kemunculan yang beragam. Rumus yang digunakan pada metode ini sebagai berikut.

$$Sim(x, y) = \frac{x \cdot y^T}{\|x\| \|y\|} \quad (3.1)$$

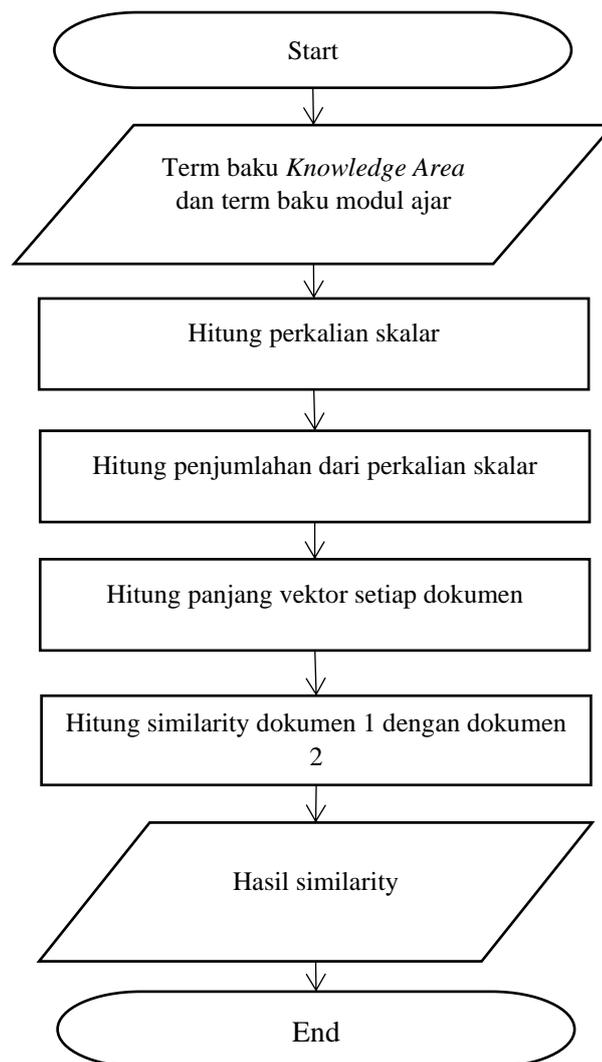
Dimana,

$x$  = vektor  $x$

$y$  = vektor  $y$

$y^T$  = tranpose dari vektor  $y$

Pada proses *similarity* dokumen menggunakan metode *Cosine Similarity* akan dijelaskan pada gambar 3.10.



Gambar 3.10. Proses *similarity* dokumen menggunakan metode *Cosine Similarity*

Gambar 3.10 menjelaskan proses *similarity* dokumen dilakukan dengan tahapan menghitung hasil perkalian skalar antara dokumen modul ajar mata pelajaran informatika (Dok 1) dan dokumen *knowledge area* (Dok 2) kemudian

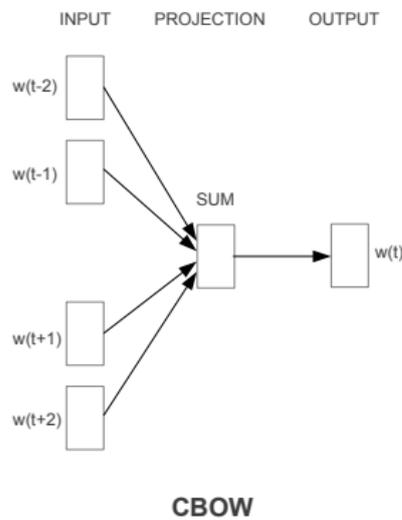
menghitung penjumlahan dari hasil perkalian skalar kedua dokumen, tahap berikutnya dilakukan perhitungan panjang vektor setiap dokumen dengan kuadratkan bobot setiap term dalam setiap dokumen, jumlahkan nilai kuadrat dan terakhir akarkan, Hitung kemiripan dokumen 1 dengan dokumen 2 dengan rumus yang disesuaikan dengan variabel sebenarnya pada dokumen.

$$\text{Cos(Dok2, Dok1)} = \frac{\text{panjang\_vektor\_Dok 2}}{\text{hitung\_panjang\_Dok 2} * \text{hitung\_panjang\_Dok 1}} \quad (3.2)$$

Dan tahap terakhir Urutkan hasil perhitungan kemiripan Dok 1 dan Dok 2. Hasil dari proses *similarity* dokumen ini dalam bentuk numerik dengan rentang 0 sampai 1.

#### **3.2.2.4.2. Proses *Similarity* Dokumen menggunakan Metode Word2Vec *Similarity***

Metode *Word2Vec similarity* adalah sebuah teknik yang digunakan untuk mewakili setiap kata dalam suatu konteks sebagai vektor dengan dimensi N. Dalam merepresentasikan sebuah kata, *Word2Vec* menggunakan jaringan saraf (*neural network*) untuk menghitung kesamaan kontekstual dan semantik dari setiap kata yang diwakili dalam bentuk vektor *one-hot encoded*.



Gambar 3.11. Arsitektur *CBOW* pada *Word2Vec*

(Sumber : Faisal Rahutomo dkk., 2019 dan Philips Nogo Raharjo dkk., 2022)

Gambar 3.11 menjelaskan tentang arsitektur *CBOW*. *CBOW* (*Continuous Bag-of-Words*) adalah arsitektur model *Word2Vec* yang digunakan untuk mempelajari representasi kata berdasarkan konteks sekitarnya. Tujuan dari *CBOW* adalah untuk memprediksi kata target berdasarkan kata-kata konteks sekitarnya. Sekumpulan kata konteks digunakan sebagai masukan, dan kata target yang akan diprediksi menjadi keluaran. Model mengalami proses pembelajaran dengan mengoptimalkan bobot vektor kata, untuk menghasilkan representasi vektor yang mendekati kata target berdasarkan konteksnya. Model arsitektur *CBOW* menggunakan input layer dengan jumlah neuron yang sama dengan jumlah kata dalam konteks. Setelah itu, terdapat lapisan tersembunyi dengan dimensi yang lebih kecil dibandingkan dengan jumlah neuron pada lapisan input. Lapisan tersembunyi ini bertanggung jawab untuk mempelajari representasi kata yang lebih terfokus. Selanjutnya, lapisan keluaran memiliki neuron yang sama dengan jumlah kata dalam kosa kata, dan menggunakan fungsi aktivasi softmax untuk menghasilkan distribusi probabilitas kata-kata yang mungkin menjadi target.

Rumus *CBOW* (*Continuous Bag-of-Words*) dapat dijelaskan sebagai berikut.

$$Output(t_{i,n}) = \sum_i^n (W_{ti,n} \times V_{ti,n}) + \dots + (W'_{t,n} \times H'_{t,n}) \quad (3.3)$$

Dimana,

*Output t* = Nilai Output target

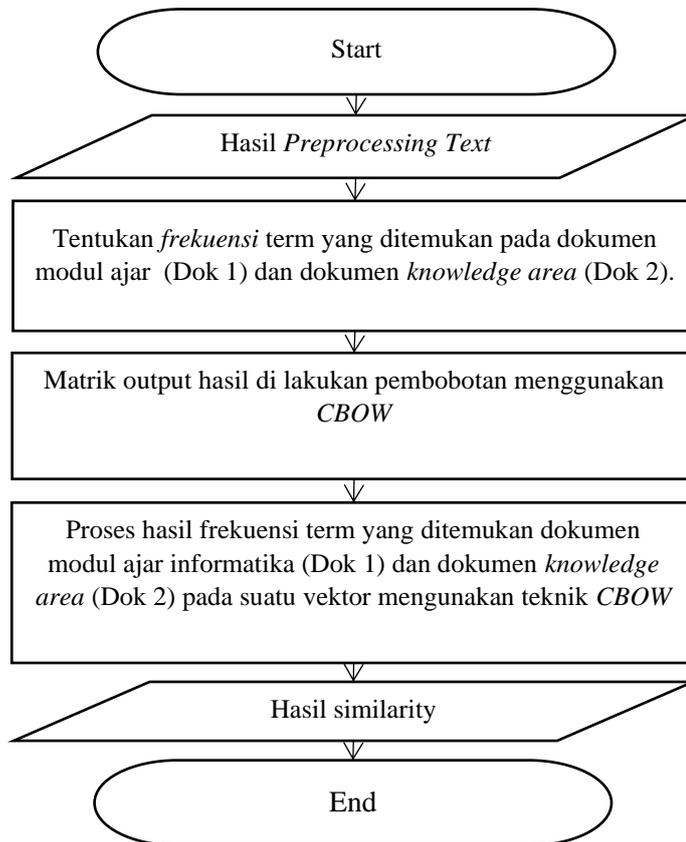
*W<sub>ti,n</sub>* = Nilai Output dari input layer

*V<sub>ti,n</sub>* = Nilai matrik input layer pada masing-masing term

*W'<sub>ti,n</sub>* = Nilai Output dari hidden layer

*H'<sub>ti,n</sub>* = Nilai matrik hidden layer pada masing-masing term

Pada proses *similarity* dokumen menggunakan metode *Word2Vec* akan dijelaskan pada Gambar 3.12.



Gambar 3.12. Proses *similarity* dokumen menggunakan metode *Word2Vec*

Gambar 3.12 menjelaskan proses *similarity* dokumen menggunakan metode *Word2Vec* tahap pertama yang dilakukan adalah dari hasil term dan jumlah term yang ditemukan disusun dengan model vektor dokumen informatika dan

dokumen knowledge area kemudian diproses training menggunakan teknik *CBOW*. Kemudian langkah selanjutnya adalah menghitung similarity dari hasil *CBOW* ini dihasilkan tingkat *similarity* dokumen dalam bentuk numerik 0 sampai dengan 1.

### 3.2.2.5. Hasil *Similarity* Dokumen

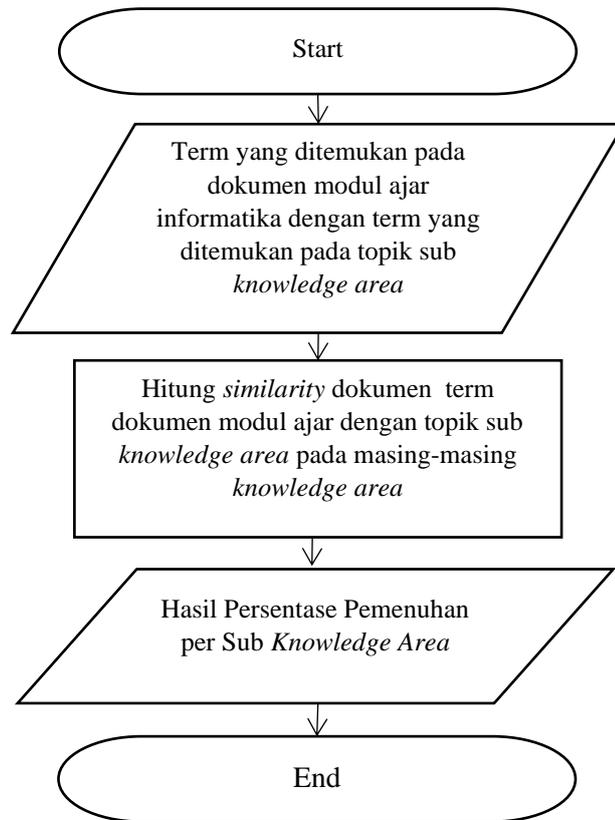
Hasil pada proses *similarity* adalah nilai *similarity* dari kedua dokumen yang dibandingkan. Hasil *similarity* pada rentang 0 sampai 1 dengan nilai output desimal. Sampel hasil *similarity* dapat dilihat pada Tabel 3.1.

Tabel 3.1 Sampel Hasil *similarity* Dokumen

Dokumen Teks Asli <i>Computer Science Curricula 2013 (Dokumen 1)</i>	Dokumen Teks Asli Modul Ajar Informatika (Dokumen 2)	Sampel Hasil <i>similarity</i>
Untuk hampir setiap ilmuwan komputer dan pengembang perangkat lunak, pemahaman tentang bagaimana manusia berinteraksi dengan mesin sangatlah penting. Meskipun topik ini mungkin tercakup dalam kursus grafik sarjana standar, topik ini juga dapat tercakup dalam kursus pengantar ilmu komputer dan pemrograman.	Guru Informatika mendidik siswa tentang bagaimana sistem komputasi berfungsi, mulai sains informatika dan matematika serta logika yang melatar-belakangi bagaimana perangkat keras berfungsi sendiri-sendiri, atau saling berkomunikasi bahkan membentuk sebuah jaringan komputer, dan bagaimana perangkat lunak dibangun dengan kemampuan pemrograman.	<b><i>Cosine similarity:</i></b> <i>Similarity</i> dokumen antara dokumen 1 dan dokumen 2 = <b>0.27</b>  <b><i>Word2vec:</i></b> <i>Similarity</i> dokumen antara dokumen 1 dan dokumen 2 = <b>0.5</b>

### 3.2.2.6. Proses Perhitungan Pemenuhan Dokumen Modul Ajar Informatika dengan *Knowledge Area*

Pada tahap ini merupakan tahap akhir untuk menghitung pemenuhan dokumen kurikulum informatika pada *knowledge area*. Pada tahapan ini dokumen *knowledge area* yang di perhitungkan adalah pada bagian topik sub *knowledge areanya*. Dimana setiap *knowledge area* memiliki beberapa sub kompetensi. Untuk menghitung pemenuhan pada masing-masing sub *knowledge area* tersebut dapat dilihat pada Gambar 3.13.



Gambar 3.13. Tahap Pemenuhan Hasil *similarity* dengan Knowledge Area

Gambar 3.13 menjelaskan proses untuk menghitung pemenuhan konten dokumen kurikulum informatika terhadap *knowledge area* dari hasil proses *similarity* dokumen yaitu dengan menghitung *similarity* dokumen dari hasil term yang ditemukan dari dokumen informatika dengan topik sub *knowledge area* pada masing-masing KA. Pada tahap ini, proses menghitung *similarity* dokumen dari hasil pengelompokan Dokumen Informatika dengan Topik Sub Knowledge Area pada masing-masing Knowledge Area dilakukan dengan menghitung *similarity* masing-masing term-term. Hasil dari menghitung *similarity* ini adalah nilai pemenuhan konten pada masing-masing Sub topik *Knowledge Area*. Pada hasil akhir perhitungan pemenuhan konten dokumen kurikulum informatika dengan Knowledge area didapatkan persentase pemenuhan konten pada masing-masing Sub *Knowledge Area* seperti Tabel 3.2.

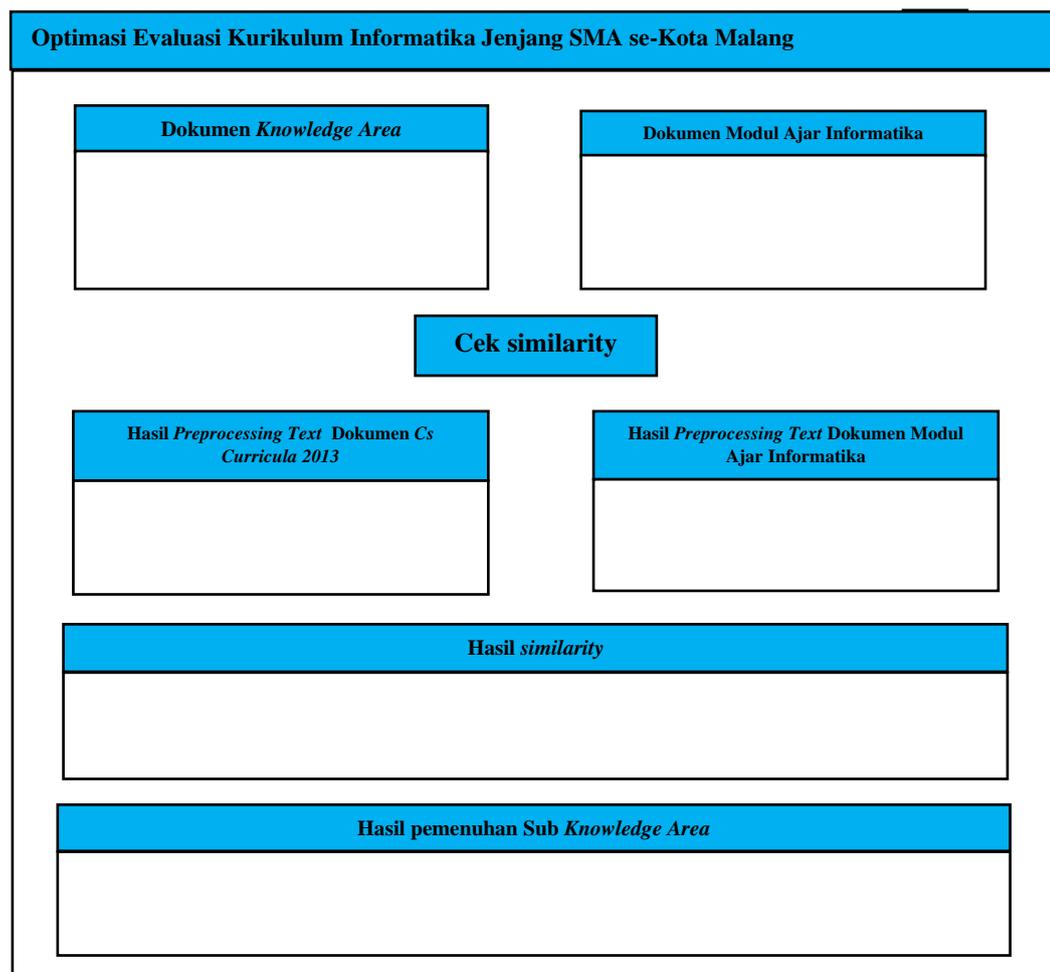
Tabel 3.2. Sampel Hasil Pemenuhan Konten Dokumen Modul Ajar Informatika

Sub KA1	Sub KA2	Sub KA3	Sub KA4	Sub KA5	Sub KA6	Sub KA7	Sub KA8	Sub KA9	Sub KA10
...%	...%	...%	...%	...%	...%	...%	...%	...%	...%
Sub KA11/	Sub KA12		Sub KA14	Sub KA15	Sub KA16	Sub KA17	Sub KA18		
...%	...%	...%	...%	...%	...%	...%	...%		

### 3.2.3. Implementasi Sistem

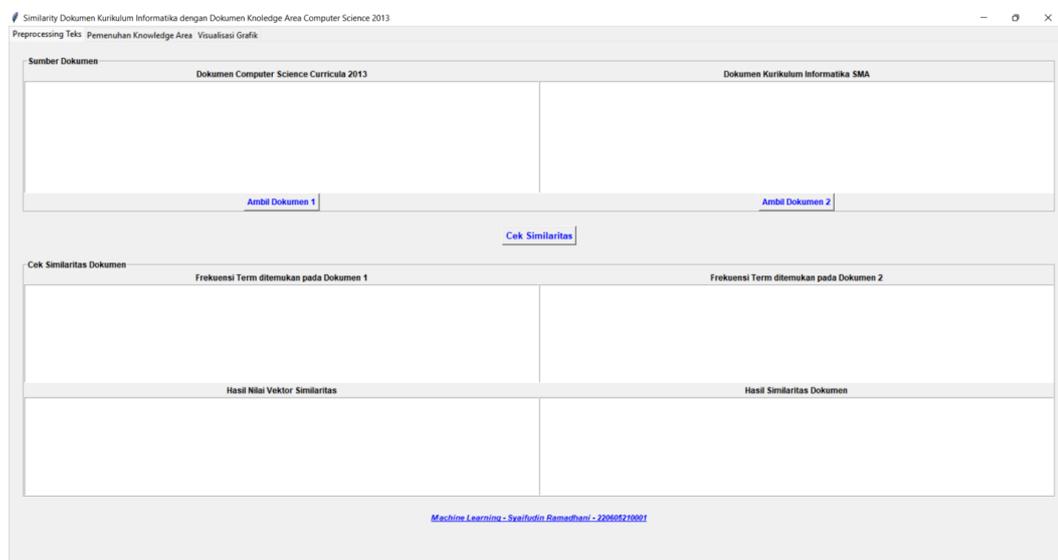
Pada tahap implementasi sistem, implementasi yang dilakukan untuk merancang sistem tahapan-tahapannya, antara lain:

1. Tahap pertama adalah membuat prototipe sistem. Pada tahap ini dilakukan perancangan desain prototipe yang akan diimplementasikan ke dalam sistem yang dapat di lihat pada Gambar 3.14.

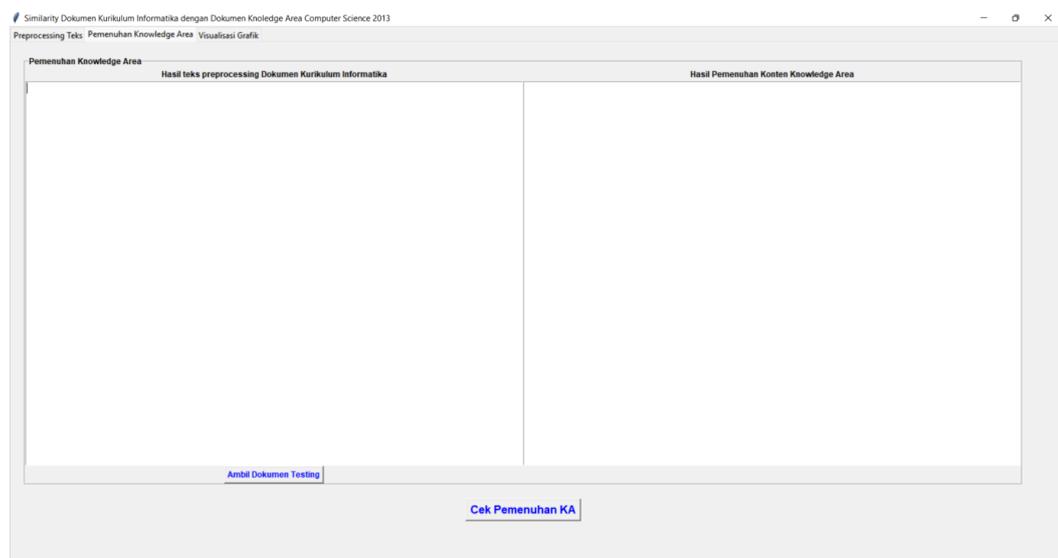


Gambar 3.14. Desain Prototipe Sistem

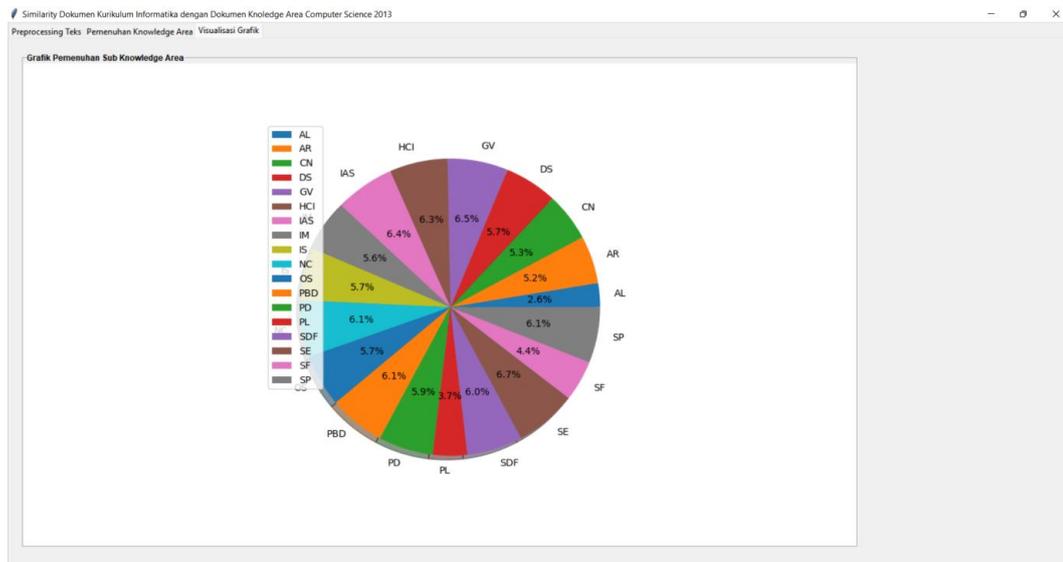
2. Pada tahap kedua adalah tahap merancang pseudo code pada setiap tahapan dalam implementasi sistem antara lain: Proses *preprocessing text*, proses perhitungan *similarity* dan proses perhitungan persentase pemenuhan *knowledge area*.
3. Pada tahap ketiga adalah mengimplementasikan dalam bentuk Graphic User Interface (GUI) menggunakan bahasa pemrograman Python yang dapat dilihat pada Gambar 3.15, Gambar 3.16 dan Gambar 3.17.



Gambar 3.15. *Graphic User Interface (GUI) Implementasi Sistem Tab 1*



Gambar 3.16. *Graphic User Interface (GUI) Implementasi Sistem Tab 2*



Gambar 3.17. *Graphic User Interface* (GUI) Implementasi Sistem Tab 3

### 3.2.3.1. Tahap *Preprocessing Text*

Pada tahap *preprocessing text*, dari data yang telah diperoleh dan sudah diolah menjadi teks, langkah selanjutnya adalah dengan melakukan empat tahap *preprocessing text* yaitu proses *case folding*, *tokenizing*, *filtering* dan *stemming*.

#### 3.2.3.1.1. *Preprocessing Text* dengan *Case Folding*

Pada proses *case folding* adalah tahap yang dilakukan antara lain sebagai berikut:

1. Mengubah teks menjadi huruf kecil (*lower case*)
2. Menghapus angka
3. Menghapus tanda baca
4. Menghapus karakter kosong (*whitespace*)

*Pseudo code* proses *case folding* ditunjukkan sebagai berikut:

<b>Case Folding (mengubah teks menjadi huruf kecil)</b>	
1	<b>Begin</b>
2	Input = inputString //input teks
3	outputString = "" //output teks = kosong
4	<b>For</b> i to inputString: //perulangan awal teks sampai akhir teks
5	<b>If</b> karakter is <b>uppercase</b> : // pengecekan bila karakter = Huruf Besar
6	outputString += karakter to <b>lowercase</b> //ubah karakter menjadi huruf kecil

7	<b>Else</b>
8	outputString += karakter //output = karakter hasil mengubah teks menjadi huruf kecil
9	<b>End For</b>
10	<b>Return</b> outputString //kembalikan output
11	<b>End</b>

### 3.2.3.1.2. Preprocessing Text dengan Filtering

Pada proses *filtering*, tahap yang dilakukan adalah dengan menghapus kata tidak penting yang sering muncul dalam jumlah besar (*stop word*). *Pseudo code* proses *filtering* ditunjukkan sebagai berikut:

<b>Filtering (menghapus kata tidak penting yang sering muncul dalam jumlah besar)</b>	
1	<b>Begin</b>
2	Input = inputString //input teks
3	outputString = "" //output teks = kosong
4	<b>For i to</b> inputString: //perulangan awal teks sampai akhir teks
5	word = word.strip(string.punctuation + string.digits //menghapus karakter non huruf
6	word1 = word.replace('yang','') //menghapus kata yang
7	word2 = word1.replace('dan','') //menghapus kata dan
8	word3 = word2.replace('dari','') //menghapus kata dari
9	Word4 = word3.replace('di','') //menghapus kata di
10	outputString = word4 //memberi value output dari variabel word4
11	<b>End For</b>
12	<b>Return</b> outputString //kembalikan output
13	<b>End</b>

### 3.2.3.1.3. Preprocessing Text dengan Stemming

Pada proses *stemming*, tahap yang dilakukan adalah dengan menghapus imbuhan pada kata menjadi kata dasar. *Pseudo code* proses *stemming* teks ditunjukkan sebagai berikut:

<b>Stemming (menghapus imbuhan pada kata menjadi kata dasar)</b>	
1	<b>Begin</b>
2	Input = inputString //input teks
3	outputString = "" //output teks = kosong
4	<b>For i to</b> inputString: //perulangan awal teks sampai akhir teks
5	word = word.stem(Input) //menghapus imbuhan pada kata berbahasa indonesia
6	outputString = word //memberi value output dari variabel word
7	<b>End For</b>
8	<b>Return</b> outputString //kembalikan output
9	<b>End</b>

### 3.2.3.1.4. Preprocessing Text dengan Tokenizing

Pada proses *tokenizing* adalah tahap yang dilakukan dengan memisahkan kalimat menjadi potongan kata. *Pseudo code* proses *tokenizing* ditunjukkan sebagai berikut:

<b>Tokenizing (memisahkan kalimat menjadi potongan kata)</b>	
1	<b>Begin</b>
2	Input = inputString //input teks
3	outputString = "" //output teks = kosong
4	words = Input.split() //memisahkan kata-kata dengan spasi
5	outputString = words //memberikan value outpur dari variabel words
6	<b>End</b>

### 3.2.3.1.5. Tahap Similarity Dokumen

Pada tahap *similarity* dokumen terdapat dua metode yang dilakukan yaitu metode *Cosine Similarity* dan *Word2Vec*. Dari data yang telah diproses pada tahap *preprocessing text*, langkah selanjutnya adalah dengan melakukan proses *similarity* dokumen menggunakan metode *Cosine Similarity* dan *Word2Vec* yang dapat dilihat sebagai berikut:

<b>Similarity Dokumen dengan metode Cosine Similarity</b>	
1	<b>Begin</b>
2	InputDok1 = inputString //input teks dokumen knowledge area
3	InputDok2 = inputString //input teks dokumen kurikulum informatika
4	dok = [InputDok1,InputDok2] //memasukkan 2 dokumen dalam index
5	for i, doc in enumerate(dok): //membuat looping untuk mengecek input dokumen1 dan dokumen 2
6	print(f"Similarity Dokumen {i+1}:") //cetak ke layar hasil similarity masing-masing dokumen
7	for j in range(len(cosine_similarities)): //membuat looping untuk mengecek similarity dokumen
8	print(f"Dokumen {j+1} - Similarity: {cosine_similarities[i][j]}") //cetak ke layar hasil similaritynya
9	print()
10	<b>End</b>

<b>Similarity Dokumen dengan metode Word2Vec</b>	
1	<b>Begin</b>
2	InputDok1 = inputString //input teks dokumen knowledge area
3	InputDok2 = inputString //input teks dokumen kurikulum informatika
4	model = Word2Vec(sentences, min_count=1)

5	<code>similarity = model.wv.similarity('InputDok1', 'InputDok2') # Menghitung similaritas antara dua dokumen</code>
6	<code>print("Similarity:", similarity)</code>
10	<b>End</b>

### 3.2.3.1.6. Tahap Perhitungan Pemenuhan Konten Dokumen Modul Ajar Informatika dengan Konten Sub *Knowledge Area*

Pada tahap perhitungan pemenuhan konten dokumen modul ajar informatika terhadap konten sub *knowledge area* dapat dijelaskan sebagai berikut:

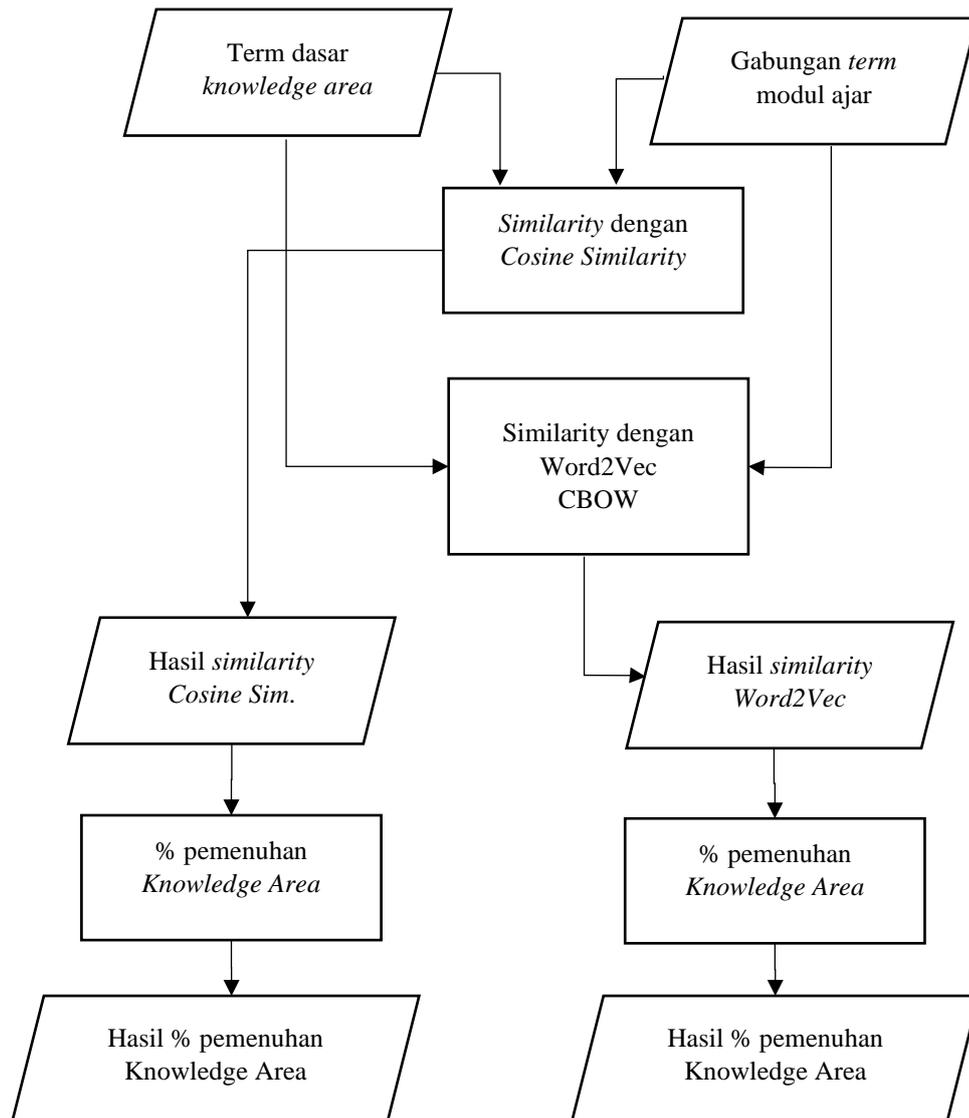
<b>Pemenuhan Konten Dokumen Modul Ajar Informatika dengan Sub <i>Knowledge Area</i></b>	
1	<b>Begin</b>
2	<code>InputSubKA = inputString //input teks sub knowledge area</code>
3	<code>InputDokInformatika = inputString //input teks dokumen kurikulum informatika</code>
4	<code>dok = [InputSubKA,InputDokInformatika] //memasukkan 2 dokumen dalam index</code>
5	<code>for i, doc in enumerate(dok): //membuat looping untuk mengecek input Sub Knowledge Area dan dok Kurikulum Informatika</code>
6	<code>print(f"Similarity Dokumen {i+1} :") //cetak ke layar hasil similarity masing- masing dokumen</code>
7	<code>for j in range(len(cosine_similarities)): //membuat looping untuk mengecek similarity dokumen dengan Cosim</code>
	<code>for j in range(len(word2vec)): //membuat looping untuk mengecek similarity dokumen Word2Vec</code>
8	<code>print(f"Dokumen {j+1} - Similarity: {cosine_similarities[i][j]},{word2vec[i][j]} ") //cetak ke layar hasil similaritynya</code>
9	<b>End</b>

### 3.2.4. Ujicoba

Pada tahap ujicoba dalam melakukan similarity dokumen modul ajar informatika dengan dokumen *knowledge area* yaitu dengan membandingkan *similarity term* dengan gabungan *term*, *term* dengan kalimat dan *term* dengan paragraph yang akan dijelaskan lebih lanjut pada sub bab berikutnya.

### 3.2.4.1. Skenario Pertama Pengujian *Similarity* Berbasis *Term* Dengan Gabungan *Term*

Pada tahap pengujian *similarity* antara *term* dengan gabungan *term* pada kedua dokumen yang diujikan *similarity*-nya, langkah yang dilakukan dapat dilihat pada Gambar 3.18.



Gambar 3.18. Tahapan pengujian similarity berbasis *term* dengan gabungan *term*

Gambar 3.18 menjelaskan tahap khusus yang membedakan adalah data input dokumen yang digunakan adalah dokumen teks yang sudah dipecah menjadi *term* dasar dan gabungan *term*. Kemudian data tersebut dilakukan pengukuran

*similarity* menggunakan metode *Cosine Similarity* dan *Word2Vec*. Setelah didapatkan skor *similarity* pada masing-masing metode, tahap selanjutnya adalah menghitung persentase pemenuhan *Knowledge Area* dengan menggunakan rumus *Euclidean Distance* berikut.

$$d(i, j) = \sqrt{(X_{i1} + X_{j1})^2 + (X_{i2} + X_{j2})^2 + \dots + (X_{ip} + X_{jp})^2} \quad (3.4)$$

Dimana  $i$  dan  $j$  adalah dua obyek data yang memiliki  $p$  atribut bernilai numerik yang dinyatakan sebagai  $i = (X_{i1}, X_{i2}, \dots, X_{ip})$  dan  $j = (X_{j1}, X_{j2}, \dots, X_{jp})$ .

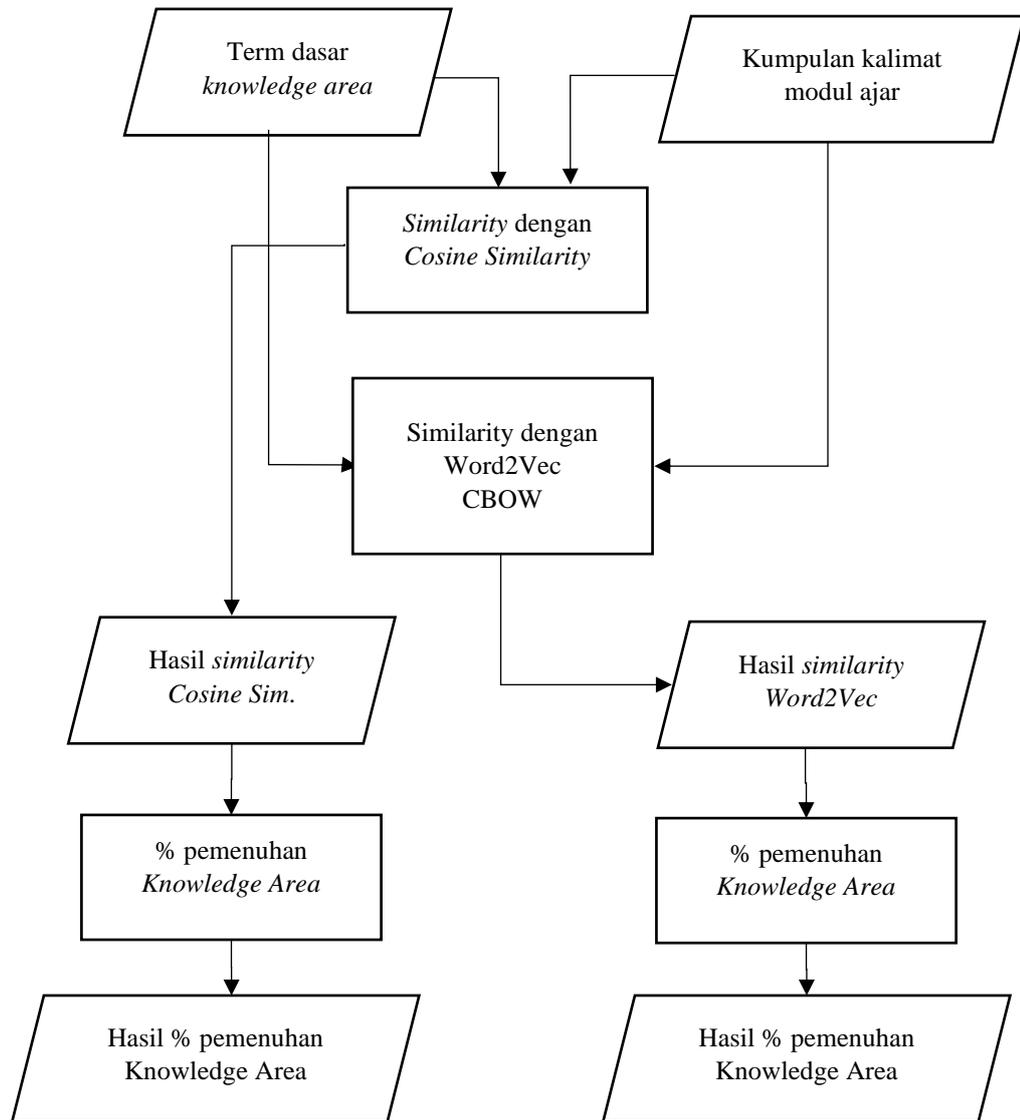
Langkah selanjutnya adalah melakukan persentase pemenuhan *Knowledge Area* dengan menggunakan rumus berikut.

$$\% \text{ pemenuhan KA} = d(i, j) * 100\% \quad (3.5)$$

Dimana,  $d(i, j)$  adalah hasil mengubah skor *similarity* dari *cosine similarity* dan *Word2Vec* menjadi bilangan positif (numerik).

#### **3.2.4.2. Skenario Kedua Pengujian *Similarity* Berbasis *Term* Dengan Kalimat**

Pada tahap pengujian *similarity* antara *term* dengan kalimat pada kedua dokumen yang diujikan *similarity*-nya, langkah yang dilakukan dapat dilihat pada Gambar 3.19.

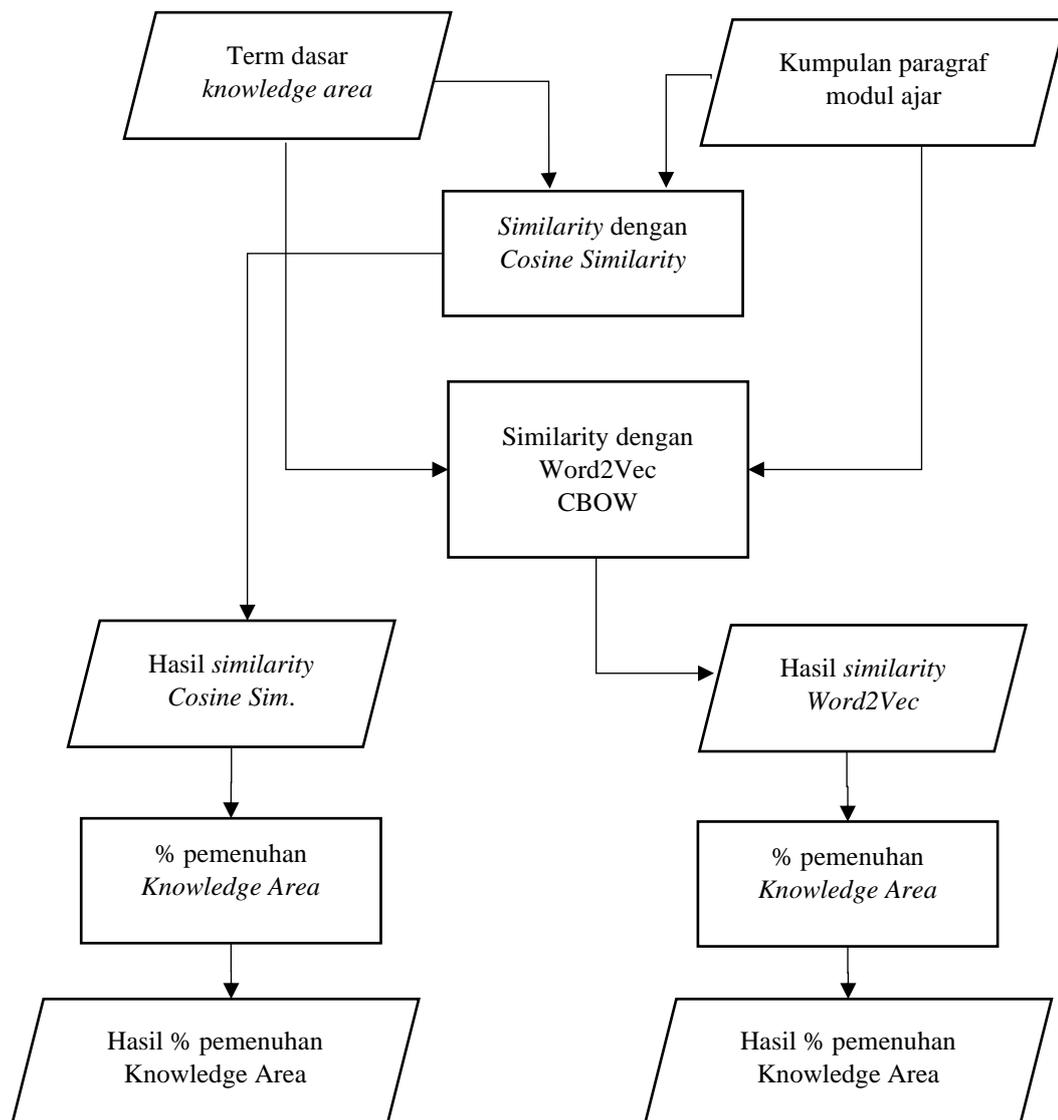


Gambar 3.19. Tahapan pengujian *similarity* berbasis *term* dengan kalimat

Pada Gambar 3.19 menjelaskan tahap khusus untuk input dokumen. Yang membedakan adalah data input dokumen yang digunakan adalah dokumen teks yang sudah dipecah menjadi *term* dasar dan kalimat. Kemudian data tersebut dilakukan pengukuran *similarity* menggunakan metode *Cosine Similarity* dan *Word2Vec*. Setelah didapatkan skor *similarity* pada masing-masing metode, tahap selanjutnya adalah menghitung persentase pemenuhan *Knowledge Area* dengan menggunakan rumus *Euclidean Distance* pada persamaan (3.4) dan persamaan (3.5) pada sub bab 3.2.4.1.

### 3.2.4.3. Skenario Ketiga Pengujian *Similarity* Berbasis *Term* Dengan Paragraf

Pada tahap pengujian *similarity* antara *term* dengan paragraf pada kedua dokumen yang diujikan *similarity*-nya, langkah yang dilakukan dapat dilihat pada Gambar 3.20.



Gambar 3.20. Tahapan pengujian *similarity* berbasis *term* dengan *gabungan term*

Gambar 3.20 menjelaskan tentang tahap khusus yang membedakan data input dokumen yang digunakan adalah dokumen teks yang sudah dipecah menjadi

*term* dasar dan paragraf. Kemudian data tersebut dilakukan pengukuran *similarity similarity* menggunakan metode *Cosine Similarity* dan *Word2Vec*. Setelah didapatkan skor *similarity* pada masing-masing metode, tahap selanjutnya adalah menghitung persentase pemenuhan *Knowledge Area* dengan menggunakan rumus *Euclidean Distance* pada persamaan (3.4) dan persamaan (3.5) pada sub bab 3.2.4.1.

### 3.2.5. Perbandingan Pengukuran Efektifitas *Similarity* Metode *Cosine Similarity* dan *Word2Vec*.

Dari ketiga skenario pengujian tersebut dalam bab sebelumnya, hasil *similarity* dari masing-masing skenario pengujian dilakukan analisis hasil pengukuran untuk menguji efektifitas metode. Menurut penulisan Kamber dan Han (2011), *confusion matrix* digunakan sebagai sebuah alat untuk menganalisis performa suatu classifier atau algoritma dalam kasus tertentu. Terdapat empat klasifikasi yang ada dalam *confusion matrix*, yaitu *True Positive*, *False Positive*, *False Negative*, dan *True Negative*. Klasifikasi-klasifikasi tersebut kemudian disingkat menjadi *TP*, *FP*, *FN*, dan *TN* secara berurut. Penjelasan ketentuan skor *similarity* untuk menghitung *ground thruth* dapat dilihat pada Tabel 3.3.

Tabel 3.3. Ketentuan klasifikasi persentase pemenuhan *Knowledge Area*

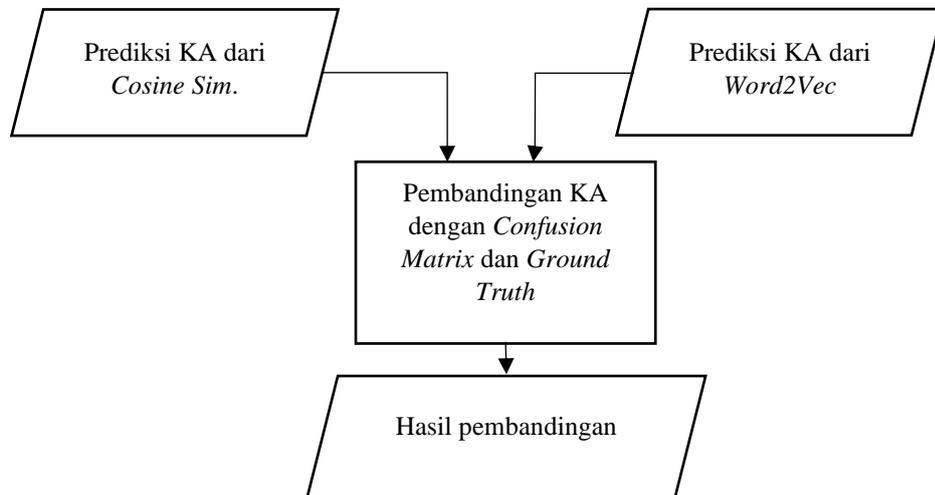
No.	Pemenuhan <i>Knowledge Area</i> (KA)	Ketentuan
1	KA <i>Expert</i> = KA Prediksi	Klasifikasi = Mirip
2	KA <i>Expert</i> $\neq$ KA Prediksi	Klasifikasi = Tidak Mirip

Berikutnya untuk menentukan *confusion matrix* dari klasifikasi yang sudah ditentukan pada Tabel 3.3 dapat dilihat pada Tabel 3.4.

Tabel 3.4. Ketentuan *Confusion Matrix*

No.	Variabel	Ketentuan
1	<i>True Positive (TP)</i>	<i>True Positive (TP)</i> di mana KA <i>Expert</i> dan KA Prediksi diklasifikasikan mirip, <i>ground truth</i> = 1
2	<i>False Positive (FP)</i>	<i>False Positive (FP)</i> di mana KA <i>Expert</i> diklasifikasikan mirip dan KA Prediksi diklasifikasikan tidak mirip, <i>ground truth</i> = 0
3	<i>False Negative (FN)</i>	<i>False Negative (FN)</i> di mana KA <i>Expert</i> diklasifikasikan tidak mirip dan KA Prediksi diklasifikasikan mirip, <i>ground truth</i> = 1
4	<i>True Negative (TN)</i>	<i>True Negative (TN)</i> di mana KA <i>Expert</i> dan KA Prediksi diklasifikasikan tidak mirip, <i>ground truth</i> = 0

Tahapan dari proses pengukuran efektifitas *similarity* dengan metode *Cosine Similarity* dan *Word2Vec* dapat dilihat pada Gambar 3.21.



Gambar 3.21. Tahapan pembandingan pengukuran efektifitas pemenuhan *Knowledge Area*

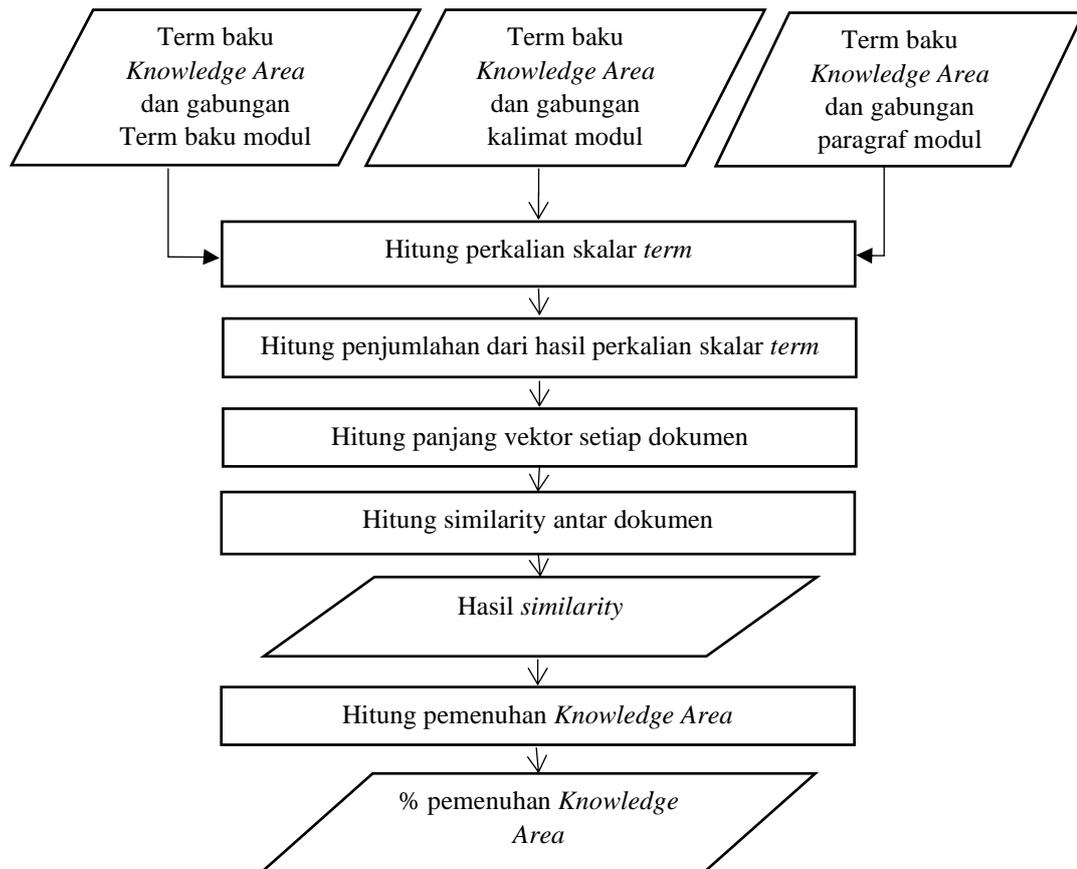
Gambar 3.21 menjelaskan hasil prediksi *Knowledge Area* dari kedua metode sudah dapat dihasilkan, maka langkah selanjutnya adalah melakukan pengukuran efektifitas hasil dari kedua metode dengan melakukan pembandingan hasil prediksi pemenuhan *Knowledge Area* dengan *confusion matrix* dan *ground truth*. Dari hasil

pembandingan tersebut akan dianalisis manakah metode yang efektif dalam melakukan similarity dokumen.

## BAB IV METODE *COSINE SIMILARITY*

### 4.1. Desain Metode

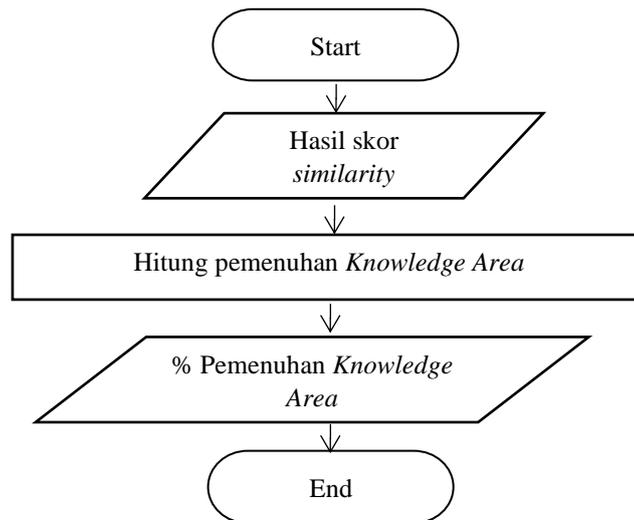
Pada tahap implementasi menggunakan metode *cosine similarity* dapat dilihat pada Gambar 4.1.



Gambar 4.1. Desain metode *Cosine Similarity*

Gambar 4.1 menjelaskan tahapan keseluruhan metodologi yang dilakukan dengan menggunakan tiga skenario input dokumen, antara lain input term baku *Knowledge Area* dengan gabungan term baku modul ajar, input term baku *Knowledge Area* dengan gabungan kalimat modul ajar dan input term baku *Knowledge Area* dengan gabungan paragraf modul ajar. Langkah selanjutnya adalah dengan menghitung perkalian skalar antara dokumen 1 (*Knowledge Area*) dengan dokumen 2 (modul ajar). Langkah selanjutnya adalah menghitung panjang

vektor setiap dokumen. Setelah didapatkan nilai panjang vektor, maka langkah selanjutnya adalah menghitung *similarity* dokumen dengan menggunakan rumus *cosine similarity*. Setelah skor *similarity* didapatkan, maka tahap selanjutnya adalah menghitung pemenuhan *Knowledge Area* pada dokumen yang dapat dilihat pada Gambar 4.2.



Gambar 4.2. Tahap Pemenuhan *Knowledge Area*

Gambar 4.2 menjelaskan dari hasil skor *similarity* yang didapatkan, maka tahap selanjutnya adalah menghitung pemenuhan *Knowledge Area* dengan menggunakan rumus berikut.

$$d(Dok1, Dok2) = \sqrt{(x_{Dok1,1} - x_{Dok2,1})^2 + (x_{Dok1,2} - x_{Dok2,2})^2 + \dots + (x_{Dok1,p} - x_{Dok2,p})^2} \quad (4.1)$$

Setelah nilai  $d(Dok1, Dok2)$  didapatkan, maka hasil tersebut di jadikan persentase dengan rumus berikut.

$$\% \text{ pemenuhan Knowledge Area} = d(Dok1, Dok2) * 100\% \quad (4.2)$$

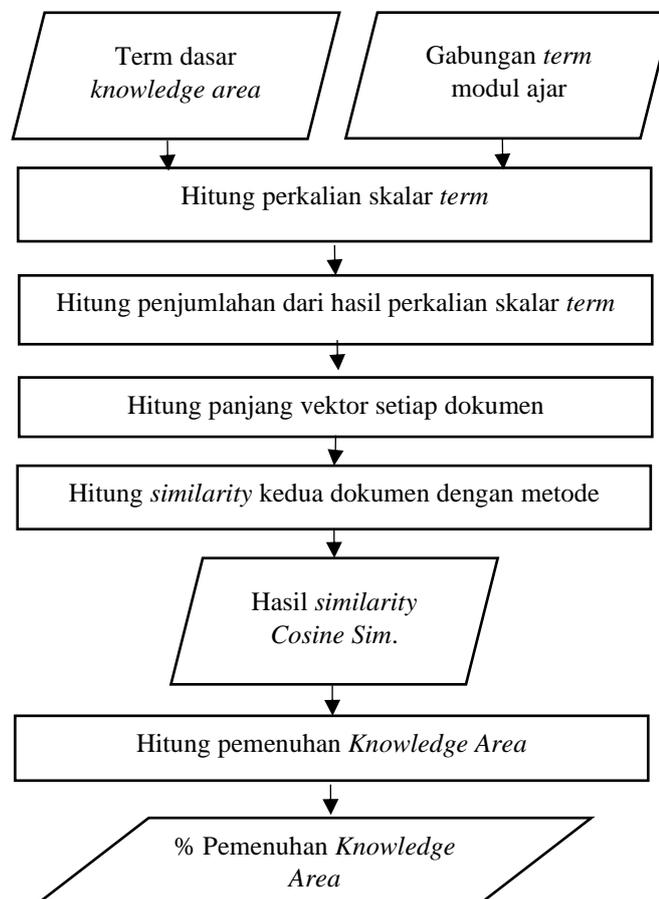
#### 4.2. Ujicoba

Pada tahap ujicoba metode *Cosine similarity* dilakukan sebanyak tiga skenario. Skenario pertama adalah dengan membandingkan *similarity term* dasar *Knowledge Area* dengan gabungan *term* modul ajar. Sedangkan pada skenario

kedua adalah dengan membandingkan *similarity term* dasar *Knowledge Area* dengan kumpulan kalimat modul ajar. Pada skenario ketiga yang dilakukan adalah dengan membandingkan *similarity term* dasar *Knowledge Area* dengan kumpulan paragraf pada modul ajar. Ketiga skenario tersebut dapat dilihat pada bab 4.2.1, 4.2.2 dan 4.2.3.

#### 4.2.1. Skenario Ujicoba Pertama

Pada tahapan skenario ujicoba pertama, langkah yang dilakukan adalah sebagai berikut.



Gambar 4.3. Tahapan skenario ujicoba pertama dengan metode Cosine similarity

Gambar 4.3 menjelaskan skenario ujicoba pertama, input yang diberikan adalah *term* dasar dokumen *Knowledge Area* dengan gabungan term dokumen modul ajar. Langkah selanjutnya adalah dengan menghitung perkalian skalar antara dokumen 1 (*Knowledge Area*) dengan dokumen 2 (modul ajar). Kemudian

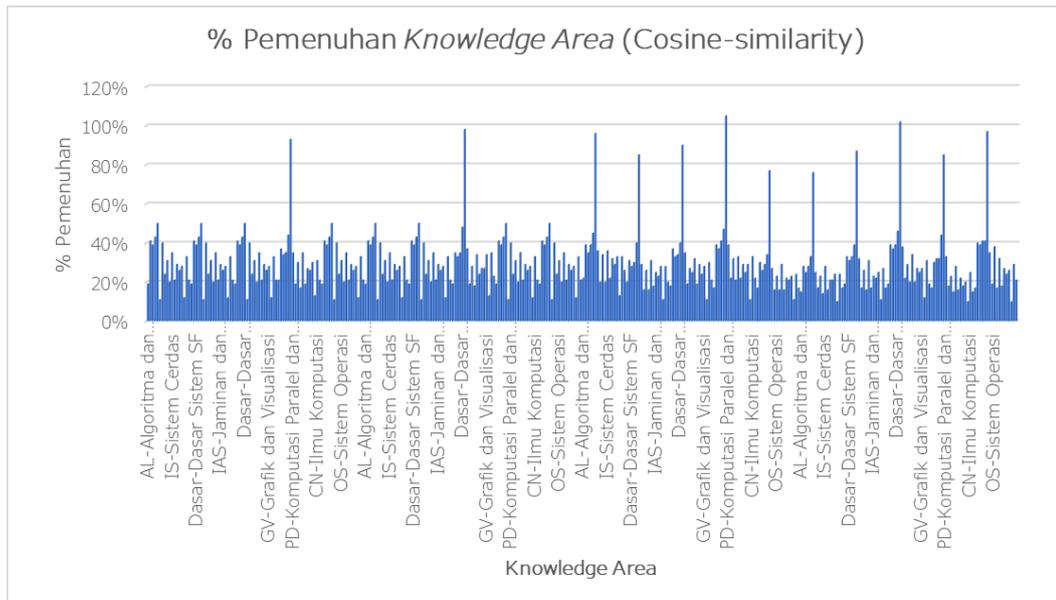
menghitung panjang vektor setiap dokumen. Setelah didapatkan nilai panjang vektor, maka langkah selanjutnya adalah menghitung *similarity* dokumen dengan menggunakan rumus cosine-similarity yang tertuang pada bab 3.2, sub bab 3.2.2.4.1. Setelah hasil *similarity* didapatkan, maka tahap selanjutnya adalah menghitung pemenuhan *Knowledge Area* dengan menggunakan rumus yang tertuang pada bab 4, sub bab 4.1. Hasil skenario ujicoba pertama dapat dilihat pada Tabel 4.1.

Tabel 4.1. Hasil *similarity* skenario ujicoba pertama

No.	Nama Sekolah	Elemen Materi	Nama Dokumen	% pemenuhan <i>Knowledge Area</i> ( <i>Cosine-similarity</i> )
1	SMAN 1 Malang	EM1 - Berpikir Komputasional	AL-Algorithm dan Kompleksitas	19%
2	SMAN 1 Malang	EM2 - Teknologi Infirmasi dan Komunikasi	AL-Algorithm dan Kompleksitas	41%
3	SMAN 1 Malang	EM3 - Sistem Komputer	AL-Algorithm dan Kompleksitas	41%
4	SMAN 1 Malang	EM4 - Jaringan Komputer dan Internet	AL-Algorithm dan Kompleksitas	29%
5	SMAN 1 Malang	EM5 - Analisis Data	AL-Algorithm dan Kompleksitas	21%
6	SMAN 1 Malang	EM6 - Algoritma dan Pemrograman	AL-Algorithm dan Kompleksitas	50%
7	SMAN 1 Malang	EM7 - Dampak Sosial Informatika	AL-Algorithm dan Kompleksitas	23%
8	SMAN 1 Malang	EM8 - Praktik Lintas Bidang	AL-Algorithm dan Kompleksitas	33%
9	SMAN 1 Malang	EM1 - Berpikir Komputasional	SP-Masalah Sosial dan Praktek Profesi	41%
10	SMAN 1 Malang	EM2 - Teknologi Infirmasi dan Komunikasi	SP-Masalah Sosial dan Praktek Profesi	22%
11	SMAN 1 Malang	EM3 - Sistem Komputer	SP-Masalah Sosial dan Praktek Profesi	20%
...	...	...	...	...
3021	SMA Negeri 114 Jakarta Utara	EM5 - Analisis Data	AR-Arsitektur dan Organisasi	35%
3022	SMA Negeri 114 Jakarta Utara	EM6 - Algoritma dan Pemrograman	AR-Arsitektur dan Organisasi	30%
3023	SMA Negeri 114 Jakarta Utara	EM7 - Dampak Sosial Informatika	AR-Arsitektur dan Organisasi	28%
3024	SMA Negeri 114 Jakarta Utara	EM8 - Praktik Lintas Bidang	AR-Arsitektur dan Organisasi	24%

Tabel 4.1 menjelaskan bahwa hasil pemenuhan *Knowledge Area* sebanyak 3024 dokumen yang diujikan, maka dihasilkan pemenuhan yang bervariasi dari masing-masing dokumen yang di ujicobakan sebanyak 18 *Knowledge Area* pada kurikulum *Computer Science Curricula 2013* dengan perbedaan signifikan antar perbandingan dokumen elemen materi modul ajar informatika dengan konten *Knowledge Area*. Pada hasil pemenuhan memiliki tren

*similarity* yang rendah dengan nilai dibawah 0.5. Namun terdapat satu dokumen pada elemen materi Algoritma dan Pemrograman memiliki pemenuhan *Knowledge Area* sebesar 50%. Visualiasasi dari hasil pemenuhan *Knowledge Area* dapat dilihat pada Gambar 4.4.

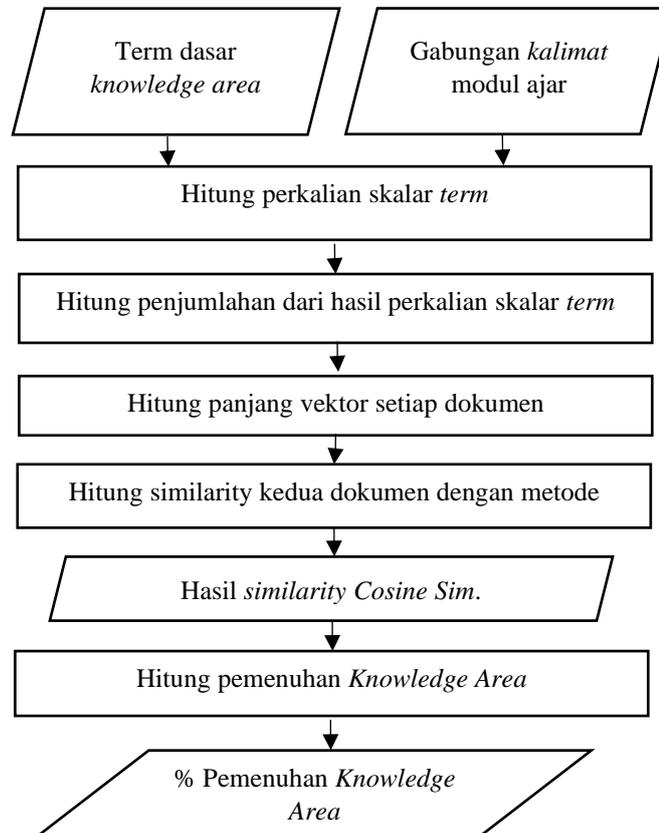


Gambar 4.4. Visualisasi pemenuhan *Knowledge Area* ujicoba pertama

Gambar 4.4 menjelaskan pada pemenuhan *Knowledge Area* dengan skor diatas 50% terdapat pada *Knowledge Area* PD-Komputasi Paralel dan Terdistribusi, SF-Dasar-Dasar Sistem, AL-Algoritma dan Kompleksitas, Dasar- SDF-Dasar Pengembangan Perangkat Lunak, CN-Ilmu Komputasi, dan OS-Sistem Operasi. Secara umum hasil pemenuhan *Knowledge Area* dengan membandingkan elemen materi modul ajar informatika sebanyak delapan dokumen didapatkan skor pemenuhan rata-rata dibawah 50%. Hal ini dapat di analisa berdasarkan konten isi materi pada masing-masing elemen materi yang diukur *similarity*-nya dan skor pemenuhannya masih belum memenuhi standar kurikulum yang berlaku, sehingga dapat dianalisis bahwa pada ujicoba pertama tren skor pemenuhan *Knowledge Area* yang sudah diujikan masih rendah.

#### 4.2.2. Skenario Ujicoba Kedua

Pada tahapan skenario ujicoba kedua, langkah yang dilakukan adalah sebagai berikut.



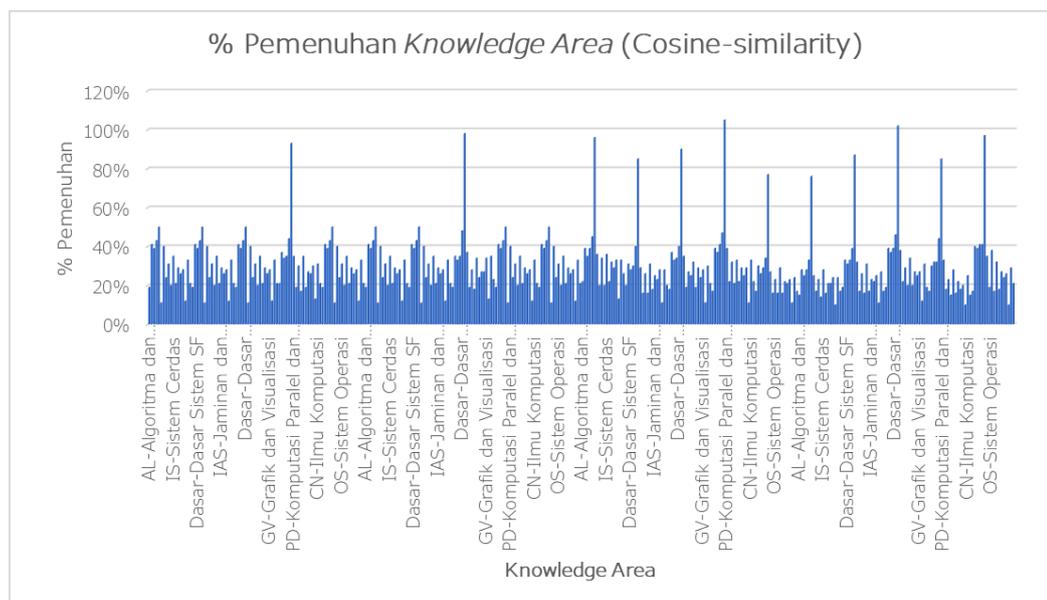
Gambar 4.5. Tahapan skenario ujicoba kedua

Gambar 4.5 menjelaskan skenario ujicoba kedua, input yang diberikan adalah term dasar dokumen Knowledge Area dengan gabungan kalimat dokumen modul ajar. Langkah selanjutnya adalah dengan menghitung perkalian skalar antara dokumen 1 (*Knowledge Area*) dengan dokumen 2 (modul ajar). Kemudian menghitung panjang vektor setiap dokumen. Setelah didapatkan nilai panjang vektor, maka langkah selanjutnya adalah menghitung similarity dokumen dengan menggunakan rumus cosine-similarity. Hasil skenario ujicoba kedua dapat dilihat pada Tabel 4.2.

Tabel 4.2. Hasil persentase pemenuhan Knowledge Area dengan ujicoba kedua

No.	Nama Sekolah	Elemen Materi	Nama Dokumen	% pemenuhan Knowledge Area (Cosine-similarity)
1	SMAN 1 Malang	EM1 - Berpikir Komputasional	AL-Algorithm dan Kompleksitas	20%
2	SMAN 1 Malang	EM2 - Teknologi Infiriasi dan Komunikasi	AL-Algorithm dan Kompleksitas	42%
3	SMAN 1 Malang	EM3 - Sistem Komputer	AL-Algorithm dan Kompleksitas	40%
4	SMAN 1 Malang	EM4 - Jaringan Komputer dan Internet	AL-Algorithm dan Kompleksitas	31%
5	SMAN 1 Malang	EM5 - Analisis Data	AL-Algorithm dan Kompleksitas	26%
6	SMAN 1 Malang	EM6 - Algoritma dan Pemrograman	AL-Algorithm dan Kompleksitas	51%
7	SMAN 1 Malang	EM7 - Dampak Sosial Informatika	AL-Algorithm dan Kompleksitas	25%
8	SMAN 1 Malang	EM8 - Praktik Lintas Bidang	AL-Algorithm dan Kompleksitas	36%
9	SMAN 1 Malang	EM1 - Berpikir Komputasional	SP-Masalah Sosial dan Praktek Profesi	47%
10	SMAN 1 Malang	EM2 - Teknologi Infiriasi dan Komunikasi	SP-Masalah Sosial dan Praktek Profesi	22%
11	SMAN 1 Malang	EM3 - Sistem Komputer	SP-Masalah Sosial dan Praktek Profesi	25%
...	...	...	...	...
3021	SMA Negeri 114 Jakarta Utara	EM5 - Analisis Data	AR-Arsitektur dan Organisasi	35%
3022	SMA Negeri 114 Jakarta Utara	EM6 - Algoritma dan Pemrograman	AR-Arsitektur dan Organisasi	33%
3023	SMA Negeri 114 Jakarta Utara	EM7 - Dampak Sosial Informatika	AR-Arsitektur dan Organisasi	28%
3024	SMA Negeri 114 Jakarta Utara	EM8 - Praktik Lintas Bidang	AR-Arsitektur dan Organisasi	25%

Tabel 4.2 menjelaskan bahwa hasil pemenuhan *Knowledge Area* sebanyak 3024 dokumen yang diujikan, maka dihasilkan pemenuhan yang bervariasi dari masing-masing dokumen yang diujicobakan sebanyak 18 Knowledge Area. Dengan menggunakan ujicoba kedua didapatkan hasil yang berbeda signifikan dengan ujicoba pertama, sehingga dapat dikatakan delta dari ujicoba pertama dan kedua tidak ada perbedaan yang jauh. Dengan perbedaan signifikan antar perbandingan dokumen elemen materi modul ajar informatika dengan konten *Knowledge Area* mendapatkan hasil pemenuhan memiliki tren *similarity* yang rendah dengan nilai dibawah 0.5. Terdapat satu dokumen pada elemen materi Algoritma dan Pemrograman memiliki pemenuhan *Knowledge Area* sebesar 50%. Visualiasasi dari hasil pemenuhan *Knowledge Area* dapat dilihat pada Gambar 4.6 berikut.

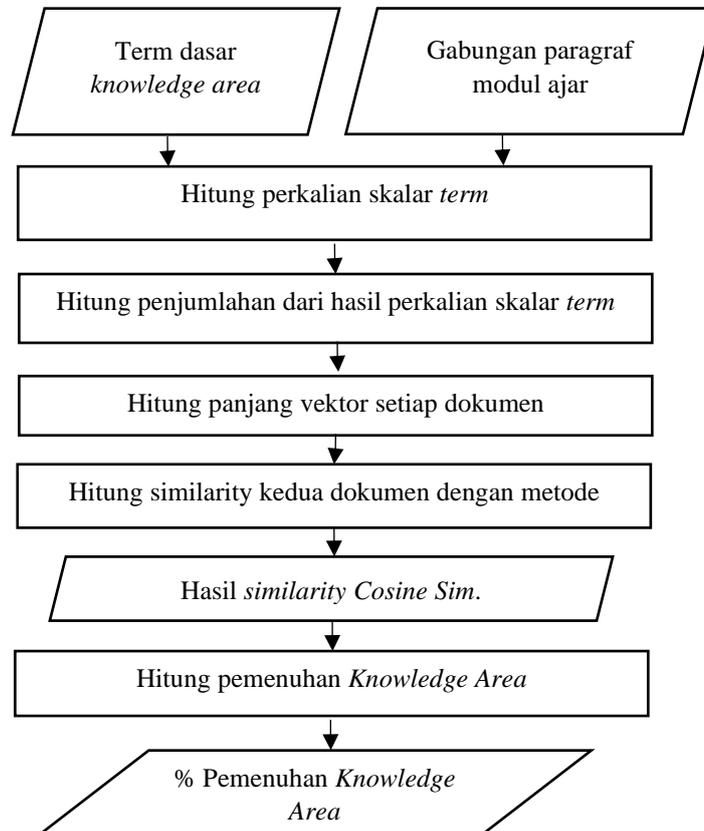


Gambar 4.6. Visualisasi pemenuhan *Knowledge Area* ujicoba kedua

Gambar 4.6 menjelaskan bahwa hasil ujicoba kedua dalam pemenuhan *Knowledge Area* dengan skor diatas 50% terdapat pada *Knowledge Area* PD-Komputasi Paralel dan Terdistribusi, SF-Dasar-Dasar Sistem, AL-Algorithm dan Kompleksitas, Dasar- SDF-Dasar Pengembangan Perangkat Lunak, CN-Ilmu Komputasi, dan OS-Sistem Operasi. Secara umum hasil pemenuhan *Knowledge Area* dengan membandingkan elemen materi modul ajar informatika sebanyak delapan dokumen didapatkan skor pemenuhan rata-rata dibawah 50%. Hal ini dapat di analisa berdasarkan konten isi materi pada masing-masing elemen materi yang diukur *similarity*-nya dan skor pemenuhannya masih belum memenuhi standar kurikulum yang berlaku, sehingga dapat dianalisis bahwa pada ujicoba kedua tren skor pemenuhan *Knowledge Area* yang sudah diujikan juga tergolong masih rendah.

### 4.2.3. Skenario Ujicoba Ketiga

Pada tahapan skenario ujicoba ketiga, langkah yang dilakukan adalah sebagai berikut.



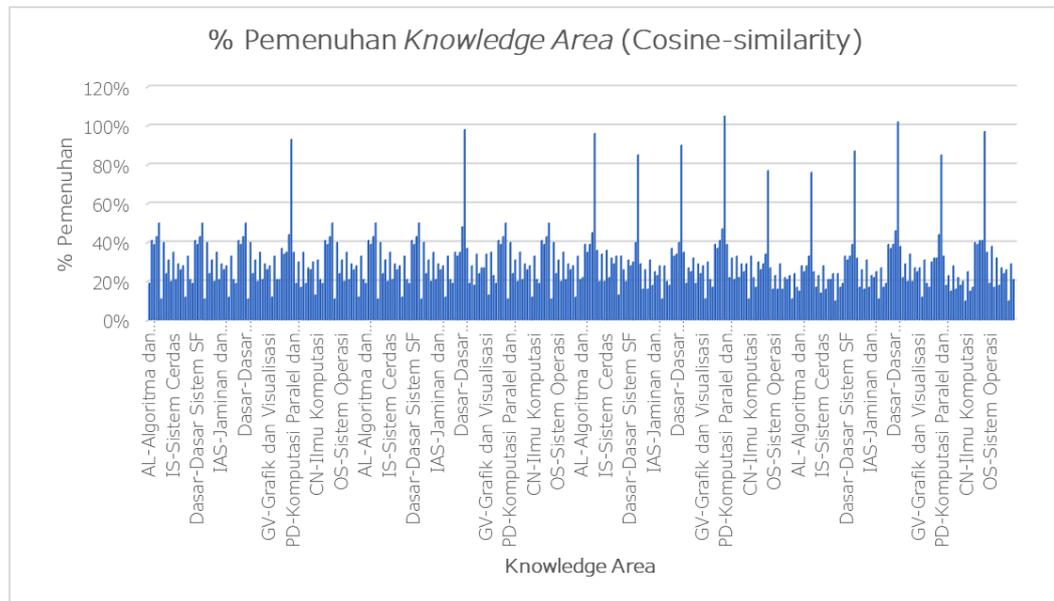
Gambar 4.7. Tahapan skenario ujicoba ketiga

Gambar 4.7 menjelaskan skenario ujicoba ketiga, input yang diberikan adalah *term* dasar dokumen *Knowledge Area* dengan gabungan paragraf dokumen modul ajar. Langkah selanjutnya adalah dengan menghitung perkalian skalar antara dokumen 1 (*Knowledge Area*) dengan dokumen 2 (modul ajar). Kemudian menghitung panjang vektor setiap dokumen. Setelah didapatkan nilai panjang vektor, maka langkah selanjutnya adalah menghitung *similarity* dokumen dengan menggunakan rumus cosine-similarity dan dari hasil skor *similarity* dilakukan pemenuhan *Knowledge Area*. Hasil skenario ujicoba ketiga dapat dilihat pada Tabel 4.3.

Tabel 4.3. Hasil persentase pemenuhan ujicoba ketiga

No.	Nama Sekolah	Elemen Materi	Nama Dokumen	% pemenuhan Knowledge Area (Cosine-similarity)
1	SMAN 1 Malang	EM1 - Berpikir Komputasional	AL-Algoritma dan Kompleksitas	18%
2	SMAN 1 Malang	EM2 - Teknologi Infirmasi dan Komunikasi	AL-Algoritma dan Kompleksitas	40%
3	SMAN 1 Malang	EM3 - Sistem Komputer	AL-Algoritma dan Kompleksitas	42%
4	SMAN 1 Malang	EM4 - Jaringan Komputer dan Internet	AL-Algoritma dan Kompleksitas	30%
5	SMAN 1 Malang	EM5 - Analisis Data	AL-Algoritma dan Kompleksitas	25%
6	SMAN 1 Malang	EM6 - Algoritma dan Pemrograman	AL-Algoritma dan Kompleksitas	52%
7	SMAN 1 Malang	EM7 - Dampak Sosial Informatika	AL-Algoritma dan Kompleksitas	26%
8	SMAN 1 Malang	EM8 - Praktik Lintas Bidang	AL-Algoritma dan Kompleksitas	35%
9	SMAN 1 Malang	EM1 - Berpikir Komputasional	SP-Masalah Sosial dan Praktek Profesi	45%
10	SMAN 1 Malang	EM2 - Teknologi Infirmasi dan Komunikasi	SP-Masalah Sosial dan Praktek Profesi	25%
11	SMAN 1 Malang	EM3 - Sistem Komputer	SP-Masalah Sosial dan Praktek Profesi	22%
...	...	...	...	...
3021	SMA Negeri 114 Jakarta Utara	EM5 - Analisis Data	AR-Arsitektur dan Organisasi	38%
3022	SMA Negeri 114 Jakarta Utara	EM6 - Algoritma dan Pemrograman	AR-Arsitektur dan Organisasi	31%
3023	SMA Negeri 114 Jakarta Utara	EM7 - Dampak Sosial Informatika	AR-Arsitektur dan Organisasi	26%
3024	SMA Negeri 114 Jakarta Utara	EM8 - Praktik Lintas Bidang	AR-Arsitektur dan Organisasi	27%

Tabel 4.3 menjelaskan bahwa hasil persentase pemenuhan *Knowledge Area* sebanyak 3024 dokumen yang diujikan, maka dihasilkan persentase pemenuhan yang bervariasi dengan hasil yang tidak berbeda terlalu jauh dengan ujicoba pertama dan kedua. Perbedaan signifikan skor pemenuhan pada perbandingan dokumen elemen materi modul ajar informatika dengan konten *Knowledge Area* mendapatkan hasil pemenuhan memiliki tren *similarity* yang rendah dengan nilai dibawah 0.5. Terdapat satu dokumen pada elemen materi Algoritma dan Pemrograman memiliki pemenuhan *Knowledge Area* sebesar 50%. Visualisasi dari hasil pemenuhan *Knowledge Area* dapat dilihat pada Gambar 4.8.



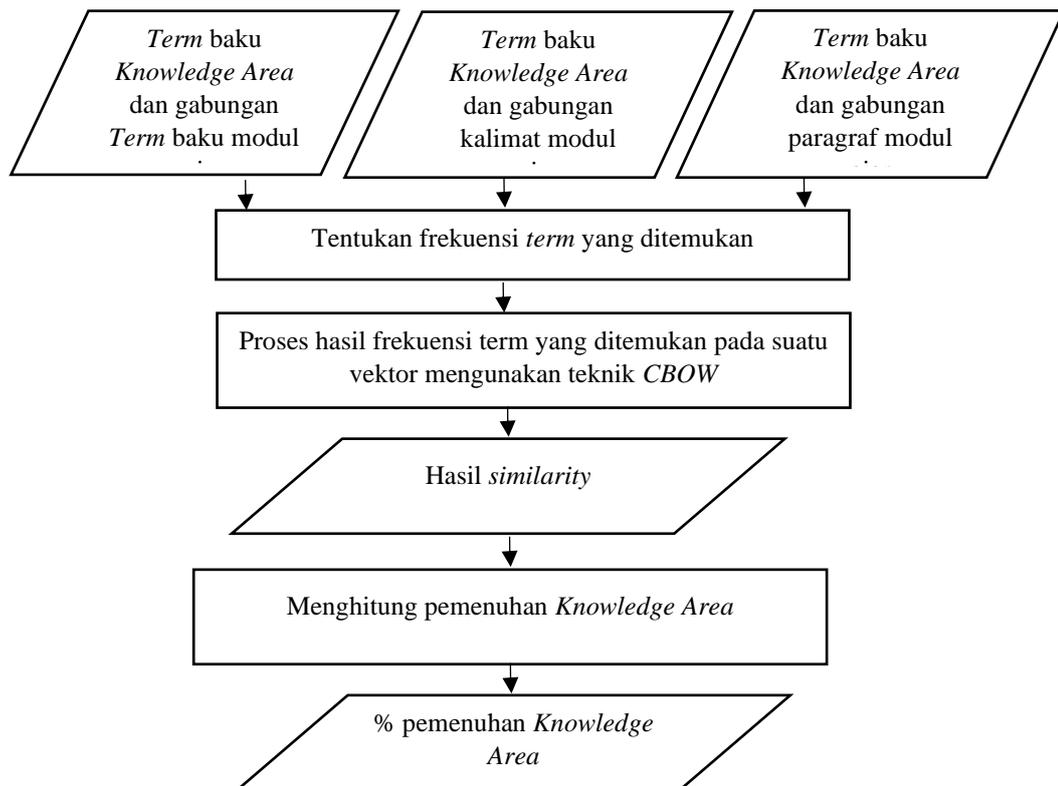
Gambar 4.8. Visualisasi pemenuhan *Knowledge Area* ujicoba ketiga

Gambar 4.8 menjelaskan bahwa dari hasil ujicoba kedua pada pemenuhan *Knowledge Area* dengan skor diatas 50% terdapat pada *Knowledge Area* PD-Komputasi Paralel dan Terdistribusi, SF-Dasar-Dasar Sistem, AL-Algorithm dan Kompleksitas, Dasar- SDF-Dasar Pengembangan Perangkat Lunak, CN-Ilmu Komputasi, dan OS-Sistem Operasi. Secara umum hasil pemenuhan *Knowledge Area* dengan membandingkan elemen materi modul ajar informatika sebanyak delapan dokumen didapatkan skor pemenuhan rata-rata dibawah 50%. Hal ini dapat di analisa berdasarkan konten isi materi pada masing-masing elemen materi yang diukur *similarity*-nya dan skor pemenuhannya masih belum memenuhi standar kurikulum yang berlaku, sehingga dapat dianalisis bahwa pada ujicoba ketiga tren skor pemenuhan *Knowledge Area* yang sudah diujikan juga tergolong masih rendah dengan skor pemenuhan rata-rata dibawah 50%.

## BAB V METODE WORD2VEC

### 5.1. Desain Metode

Pada tahap implementasi menggunakan metode *Word2Vec* dapat dilihat pada Gambar 5.1.



Gambar 5.1. Desain metode Word2Vec

Gambar 5.1 menjelaskan tahapan keseluruhan metodologi yang dilakukan dengan menggunakan tiga skenario input dokumen, antara lain input *term* baku *Knowledge Area* dengan gabungan *term* baku modul ajar, input *term* baku *Knowledge Area* dengan gabungan kalimat modul ajar dan input *term* baku *Knowledge Area* dengan gabungan paragraf modul ajar. Langkah selanjutnya adalah dengan menghitung perkalian skalar antara dokumen 1 (*Knowledge Area*) dengan dokumen 2 (modul ajar). Langkah selanjutnya adalah menghitung panjang vektor setiap dokumen. Setelah didapatkan nilai panjang vektor, maka langkah

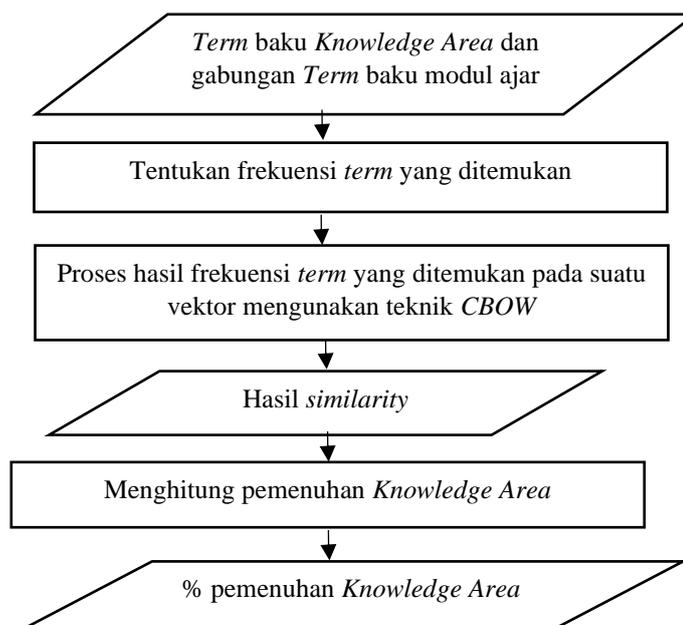
selanjutnya adalah menghitung *similarity* dokumen dengan menggunakan rumus *cosine similarity*. Hasil dari skor *similarity* diproses untuk menghitung pemenuhan *Knowledge Area* dengan menggunakan rumus yang ada pada bab 3 sub bab 3.2.2.5.2.

## 5.2. Ujicoba

Pada tahap ujicoba metode Word2Vec dilakukan sebanyak tiga skenario. Skenario pertama adalah dengan membandingkan *similarity term* dasar *Knowledge Area* dengan gabungan *term* modul ajar. Sedangkan pada skenario kedua adalah dengan membandingkan *similarity term* dasar *Knowledge Area* dengan kumpulan kalimat modul ajar. Pada skenario ketiga yang dilakukan adalah dengan membandingkan *similarity term* dasar *Knowledge Area* dengan kumpulan paragraf pada modul ajar. Ketiga skenario tersebut dapat dilihat pada bab 5.2.1, 5.2.2 dan 5.2.3.

### 5.2.1. Skenario Ujicoba Pertama

Pada tahapan skenario ujicoba pertama, langkah yang dilakukan adalah sebagai berikut.



Gambar 5.2. Tahapan skenario ujicoba pertama dengan metode Word2Vec

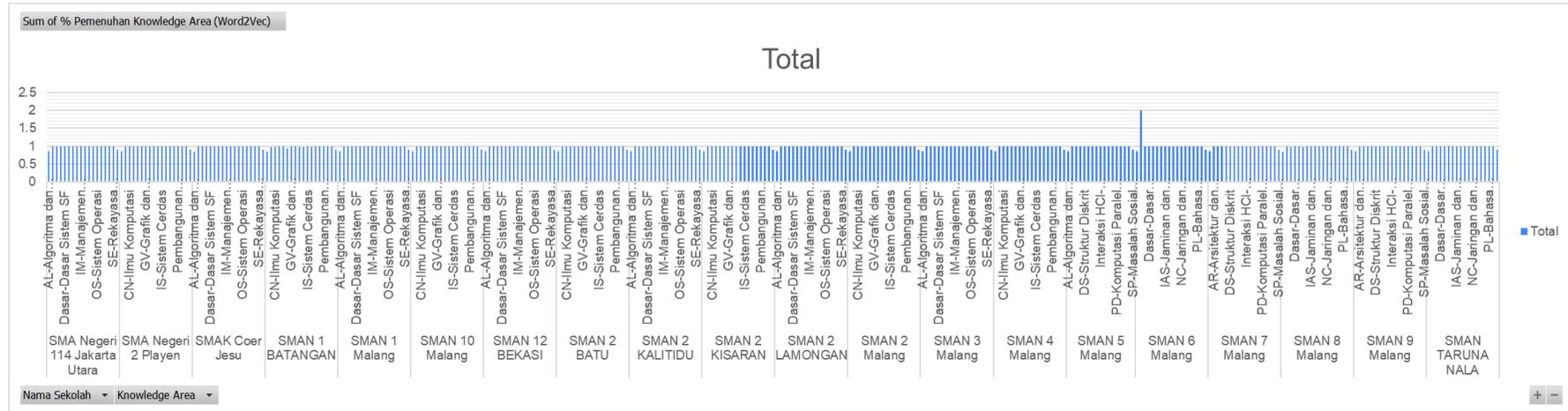
Gambar 5.2 menjelaskan skenario ujicoba pertama, input yang diberikan adalah *term* dasar dokumen *Knowledge Area* dengan gabungan *term* dokumen modul ajar. Langkah selanjutnya adalah dengan menghitung perkalian skalar antara dokumen 1 (*Knowledge Area*) dengan dokumen 2 (modul ajar). Kemudian menghitung panjang vektor setiap dokumen. Setelah didapatkan nilai panjang vektor, maka langkah selanjutnya adalah menghitung *similarity* dokumen dengan menggunakan rumus cosine-similarity. Hasil skenario ujicoba pertama dapat dilihat pada Tabel 5.1.

Tabel 5.1. Hasil persentase pemenuhan *Knowledge Area* ujicoba pertama dengan metode Word2Vec

No.	Nama Sekolah	Elemen Materi	Nama Dokumen	% pemenuhan <i>Knowledge Area</i> (Word2Vec)
1	SMAN 1 Malang	EM1 - Berpikir Komputasional	AL-Algoritma dan Kompleksitas	85.70604%
2	SMAN 1 Malang	EM2 - Teknologi Infirmasi dan Komunikasi	AL-Algoritma dan Kompleksitas	85.92604%
3	SMAN 1 Malang	EM3 - Sistem Komputer	AL-Algoritma dan Kompleksitas	85.92604%
4	SMAN 1 Malang	EM4 - Jaringan Komputer dan Internet	AL-Algoritma dan Kompleksitas	85.80604%
5	SMAN 1 Malang	EM5 - Analisis Data	AL-Algoritma dan Kompleksitas	85.72604%
6	SMAN 1 Malang	EM6 - Algoritma dan Pemrograman	AL-Algoritma dan Kompleksitas	86.01604%
7	SMAN 1 Malang	EM7 - Dampak Sosial Informatika	AL-Algoritma dan Kompleksitas	85.74604%
8	SMAN 1 Malang	EM8 - Praktik Lintas Bidang	AL-Algoritma dan Kompleksitas	85.84604%
9	SMAN 1 Malang	EM1 - Berpikir Komputasional	SP-Masalah Sosial dan Praktek Profesi	89.98950%
10	SMAN 1 Malang	EM2 - Teknologi Infirmasi dan Komunikasi	SP-Masalah Sosial dan Praktek Profesi	89.79950%
11	SMAN 1 Malang	EM3 - Sistem Komputer	SP-Masalah Sosial dan Praktek Profesi	89.77950%
12	SMAN 1 Malang	EM4 - Jaringan Komputer dan Internet	SP-Masalah Sosial dan Praktek Profesi	89.67950%
13	SMAN 1 Malang	EM5 - Analisis Data	SP-Masalah Sosial dan Praktek Profesi	89.87950%
14	SMAN 1 Malang	EM6 - Algoritma dan Pemrograman	SP-Masalah Sosial dan Praktek Profesi	89.92950%
15	SMAN 1 Malang	EM7 - Dampak Sosial Informatika	SP-Masalah Sosial dan Praktek Profesi	90.02950%
16	SMAN 1 Malang	EM8 - Praktik Lintas Bidang	SP-Masalah Sosial dan Praktek Profesi	90.13950%
18	SMAN 1 Malang	EM1 - Berpikir Komputasional	Dasar-Dasar Sistem SF	99.99374%
19	SMAN 1 Malang	EM2 - Teknologi Infirmasi dan Komunikasi	Dasar-Dasar Sistem SF	99.88374%
20	SMAN 1 Malang	EM3 - Sistem Komputer	Dasar-Dasar Sistem SF	99.91774%
21	SMAN 1 Malang	EM4 - Jaringan Komputer dan Internet	Dasar-Dasar Sistem SF	99.59774%
22	SMAN 1 Malang	EM5 - Analisis Data	Dasar-Dasar Sistem SF	99.66774%

No.	Nama Sekolah	Elemen Materi	Nama Dokumen	% pemenuhan <i>Knowledge Area</i> ( <i>Word2Vec</i> )
23	SMAN 1 Malang	EM6 - Algoritma dan Pemrograman	Dasar-Dasar Sistem SF	99.66774%
24	SMAN 1 Malang	EM7 - Dampak Sosial Informatika	Dasar-Dasar Sistem SF	99.61774%
...	...	...	...	...
3021	SMA Negeri 114 Jakarta Utara	EM1 - Berpikir Komputasional	AR-Arsitektur dan Organisasi	99.96512%
3022	SMA Negeri 114 Jakarta Utara	EM2 - Teknologi Infirmasi dan Komunikasi	AR-Arsitektur dan Organisasi	99.97111%
3023	SMA Negeri 114 Jakarta Utara	EM3 - Sistem Komputer	AR-Arsitektur dan Organisasi	99.91111%
3024	SMA Negeri 114 Jakarta Utara	EM4 - Jaringan Komputer dan Internet	AR-Arsitektur dan Organisasi	99.88111%

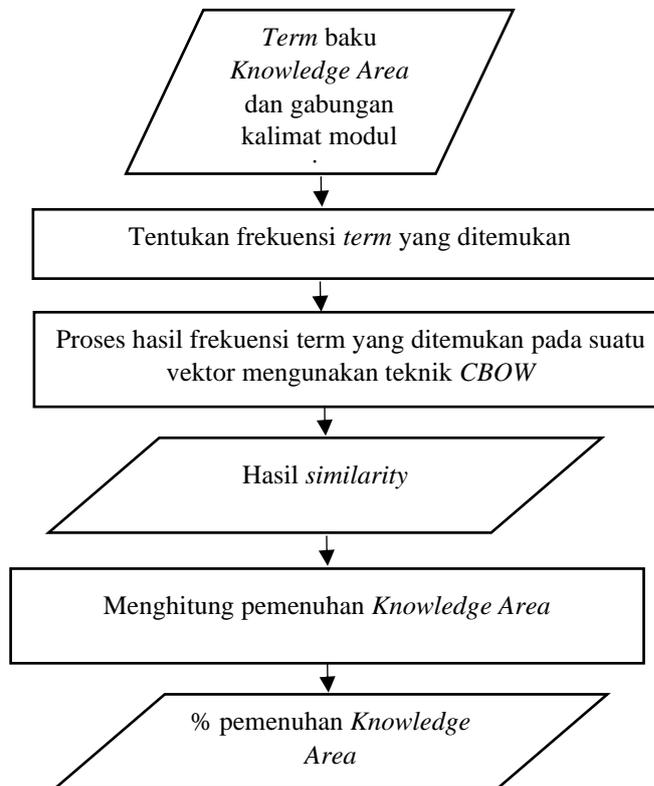
Tabel 5.1 menjelaskan bahwa hasil persentase pemenuhan *Knowledge Area* sebanyak 3024 dokumen yang diujikan, maka dihasilkan pemenuhan yang bervariasi dari setiap dokumen yang diujikan dengan target pemenuhan 18 *Knowledge Area* pada kurikulum *Computer Science Curricula 2013* dengan perbedaan signifikan antar perbandingan dokumen elemen materi modul ajar informatika dengan konten *Knowledge Area*. Pada hasil pemenuhan memiliki tren *similarity* yang baik dengan nilai diatas 50% mendekati 100% pada masing-masing elemen materi yang dibandingkan. Visualiasasi dari hasil pemenuhan *Knowledge Area* dapat dilihat pada Gambar 5.3. Gambar 5.3 menjelaskan pada pemenuhan *Knowledge Area* rata-rata mendapatkan skor diatas 50%. Secara umum hasil pemenuhan *Knowledge Area* dengan membandingkan elemen materi modul ajar informatika sebanyak delapan dokumen didapatkan skor pemenuhan rata-rata diatas 50% ke 100%. Hal ini dapat di analisa berdasarkan konten isi materi pada masing-masing elemen materi yang diukur *similarity*-nya dan skor pemenuhannya memenuhi standar kurikulum yang berlaku, sehingga dapat dianalisis bahwa pada ujicoba pertama tren skor pemenuhan *Knowledge Area* yang sudah diujikan tinggi.



Gambar 5.3. Visualisasi pemenuhan Knowledge Area ujicoba pertama

### 5.2.2. Skenario Ujicoba Kedua

Pada tahapan skenario ujicoba kedua, langkah yang dilakukan adalah sebagai berikut.



Gambar 5.4. Tahapan skenario ujicoba kedua dengan metode Word2Vec

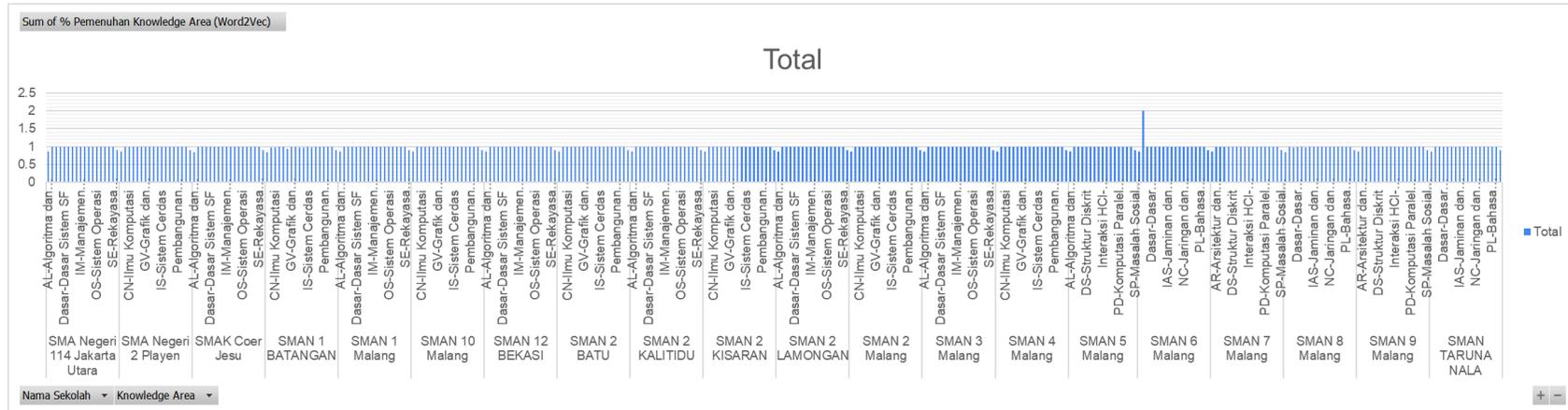
Gambar 5.4 menjelaskan skenario ujicoba kedua, input yang diberikan adalah *term* dasar dokumen *Knowledge Area* dengan gabungan kalimat dokumen modul ajar. Langkah selanjutnya adalah dengan menghitung perkalian skalar antara dokumen 1 (*Knowledge Area*) dengan dokumen 2 (modul ajar). Kemudian menghitung panjang vektor setiap dokumen. Setelah didapatkan nilai panjang vektor, maka langkah selanjutnya adalah menghitung similarity dokumen dengan menggunakan rumus cosine-similarity. Hasil skenario ujicoba kedua dapat dilihat pada Tabel 5.2.

Tabel 5.2. Hasil persentase pemenuhan *Knowledge Area* ujicoba kedua dengan metode Word2Vec

No.	Nama Sekolah	Elemen Materi	Nama Dokumen	% pemenuhan <i>Knowledge Area</i> (Word2Vec)
1	SMAN 1 Malang	EM1 - Berpikir Komputasional	AL-Algoritma dan Kompleksitas	84.70604%
2	SMAN 1 Malang	EM2 - Teknologi Infirmasi dan Komunikasi	AL-Algoritma dan Kompleksitas	85.94504%
3	SMAN 1 Malang	EM3 - Sistem Komputer	AL-Algoritma dan Kompleksitas	84.96054%
4	SMAN 1 Malang	EM4 - Jaringan Komputer dan Internet	AL-Algoritma dan Kompleksitas	85.80604%
5	SMAN 1 Malang	EM5 - Analisis Data	AL-Algoritma dan Kompleksitas	85.72604%
6	SMAN 1 Malang	EM6 - Algoritma dan Pemrograman	AL-Algoritma dan Kompleksitas	86.10804%
7	SMAN 1 Malang	EM7 - Dampak Sosial Informatika	AL-Algoritma dan Kompleksitas	85.88004%
8	SMAN 1 Malang	EM8 - Praktik Lintas Bidang	AL-Algoritma dan Kompleksitas	85.90604%
9	SMAN 1 Malang	EM1 - Berpikir Komputasional	SP-Masalah Sosial dan Praktek Profesi	89.95950%
10	SMAN 1 Malang	EM2 - Teknologi Infirmasi dan Komunikasi	SP-Masalah Sosial dan Praktek Profesi	89.77950%
11	SMAN 1 Malang	EM3 - Sistem Komputer	SP-Masalah Sosial dan Praktek Profesi	89.80950%
12	SMAN 1 Malang	EM4 - Jaringan Komputer dan Internet	SP-Masalah Sosial dan Praktek Profesi	89.76850%
13	SMAN 1 Malang	EM5 - Analisis Data	SP-Masalah Sosial dan Praktek Profesi	89.87950%
14	SMAN 1 Malang	EM6 - Algoritma dan Pemrograman	SP-Masalah Sosial dan Praktek Profesi	89.93950%
15	SMAN 1 Malang	EM7 - Dampak Sosial Informatika	SP-Masalah Sosial dan Praktek Profesi	90.03850%
16	SMAN 1 Malang	EM8 - Praktik Lintas Bidang	SP-Masalah Sosial dan Praktek Profesi	90.12850%
18	SMAN 1 Malang	EM1 - Berpikir Komputasional	Dasar-Dasar Sistem SF	99.98374%
19	SMAN 1 Malang	EM2 - Teknologi Infirmasi dan Komunikasi	Dasar-Dasar Sistem SF	99.89374%
20	SMAN 1 Malang	EM3 - Sistem Komputer	Dasar-Dasar Sistem SF	99.80874%
21	SMAN 1 Malang	EM4 - Jaringan Komputer dan Internet	Dasar-Dasar Sistem SF	99.59774%
22	SMAN 1 Malang	EM5 - Analisis Data	Dasar-Dasar Sistem SF	99.77874%
23	SMAN 1 Malang	EM6 - Algoritma dan Pemrograman	Dasar-Dasar Sistem SF	99.60988%
24	SMAN 1 Malang	EM7 - Dampak Sosial Informatika	Dasar-Dasar Sistem SF	99.63765%
...	...	...	...	...
3021	SMA Negeri 114 Jakarta Utara	EM1 - Berpikir Komputasional	AR-Arsitektur dan Organisasi	99.97812%
3022	SMA Negeri 114 Jakarta Utara	EM2 - Teknologi Infirmasi dan Komunikasi	AR-Arsitektur dan Organisasi	99.9689%
3023	SMA Negeri 114 Jakarta Utara	EM3 - Sistem Komputer	AR-Arsitektur dan Organisasi	99.92211%
3024	SMA Negeri 114 Jakarta Utara	EM4 - Jaringan Komputer dan Internet	AR-Arsitektur dan Organisasi	99.89098%

Tabel 5.2 menjelaskan bahwa hasil persentase pemenuhan dari dokumen sebanyak 3024 dokumen yang diujikan dengan target pemenuhan *Knowledge Area* sebanyak 18 KA, maka dihasilkan pemenuhan yang sangat bervariasi namun

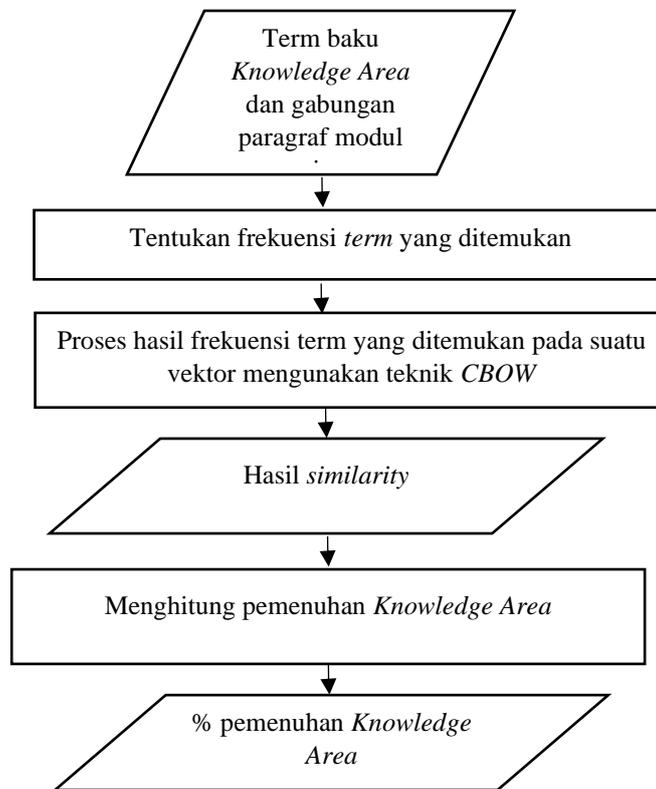
cenderung signifikan. Pada kurikulum *Computer Science Curricula 2013* dengan perbedaan signifikan antar perbandingan dokumen elemen materi modul ajar informatika dengan konten *Knowledge Area*. Pada hasil pemenuhan memiliki tren *similarity* yang baik dengan nilai diatas 50% mendekati 100% pada masing-masing elemen materi yang dibandingkan. Visualiasasi dari hasil pemenuhan *Knowledge Area* dapat dilihat pada Gambar 5.5. Gambar 5.5 menjelaskan pada pemenuhan *Knowledge Area* rata-rata mendapatkan skor diatas 50%. Secara umum hasil pemenuhan *Knowledge Area* dengan membandingkan elemen materi modul ajar informatika sebanyak delapan dokumen didapatkan skor pemenuhan rata-rata diatas 50% ke 100%. Dapat di analisa berdasarkan konten isi materi pada masing-masing elemen materi yang diukur *similarity*-nya dan skor pemenuhannya sudah memenuhi standar kurikulum yang berlaku, sehingga dapat dianalisis bahwa pada ujicoba pertama tren skor pemenuhan *Knowledge Area* yang sudah diujikan tinggi.



Gambar 5.5. Visualisasi pemenuhan Knowledge Area ujicoba kedua

### 5.2.3. Skenario Ujicoba Ketiga

Pada tahapan skenario ujicoba ketiga, langkah yang dilakukan adalah sebagai berikut.



Gambar 5.6. Tahapan skenario ujicoba ketiga dengan metode Word2Vec

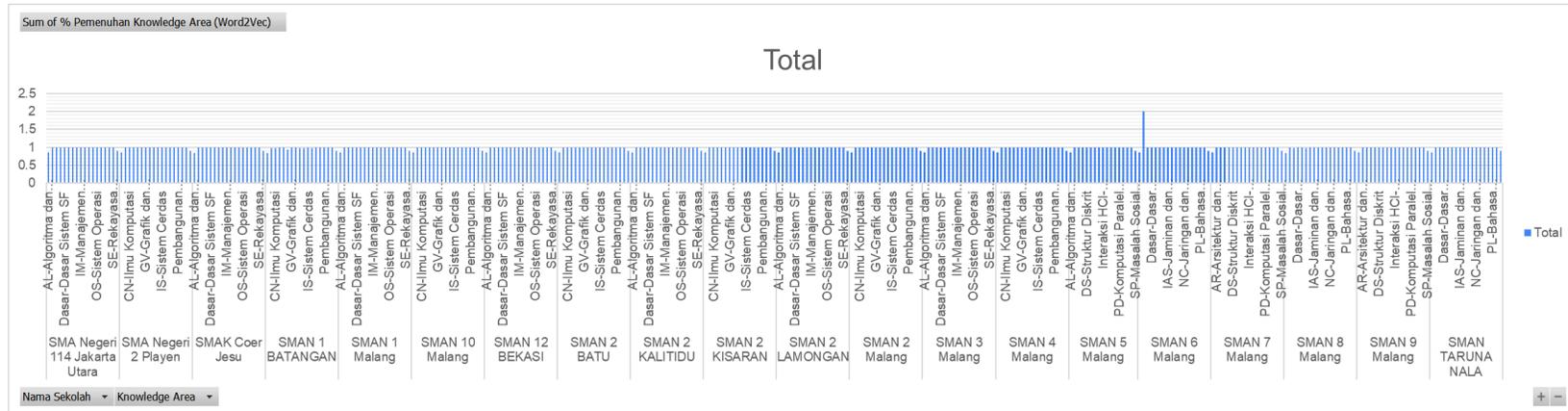
Gambar 5.6 menjelaskan skenario ujicoba ketiga, input yang diberikan adalah *term* dasar dokumen *Knowledge Area* dengan gabungan paragraf dokumen modul ajar. Langkah selanjutnya adalah dengan menghitung perkalian skalar antara dokumen 1 (*Knowledge Area*) dengan dokumen 2 (modul ajar). Kemudian menghitung panjang vektor setiap dokumen. Setelah didapatkan nilai panjang vektor, maka langkah selanjutnya adalah menghitung similarity dokumen dengan menggunakan rumus cosine similarity. Hasil skenario ujicoba ketiga dapat dilihat pada Tabel 5.3.

Tabel 5.3. Hasil persentase pemenuhan *Knowledge Area* ujicoba ketiga dengan metode Word2Vec

No.	Nama Sekolah	Elemen Materi	Nama Dokumen	% pemenuhan <i>Knowledge Area</i> (Word2Vec)
1	SMAN 1 Malang	EM1 - Berpikir Komputasional	AL-Algoritma dan Kompleksitas	85.70604%
2	SMAN 1 Malang	EM2 - Teknologi Infirmasi dan Komunikasi	SP-Masalah Sosial dan Praktek Profesi	89.98950%
3	SMAN 1 Malang	EM3 - Sistem Komputer	Dasar-Dasar Sistem SF	99.99374%
4	SMAN 1 Malang	EM4 - Jaringan Komputer dan Internet	SE-Rekayasa Perangkat Lunak	99.98646%
5	SMAN 1 Malang	EM5 - Analisis Data	Dasar-Dasar Pengembangan Perangkat Lunak SDF	99.99258%
6	SMAN 1 Malang	EM6 - Algoritma dan Pemrograman	PL-Bahasa Pemrograman	99.99512%
7	SMAN 1 Malang	EM7 - Dampak Sosial Informatika	PD-Komputasi Paralel dan Terdistribusi	99.99446%
8	SMAN 1 Malang	EM8 - Praktik Lintas Bidang	Pembangunan Berbasis Platform PBD	99.99407%
9	SMAN 1 Malang	EM1 - Berpikir Komputasional	OS-Sistem Operasi	99.98774%
10	SMAN 1 Malang	EM2 - Teknologi Infirmasi dan Komunikasi	NC-Jaringan dan Komunikasi	99.73085%
11	SMAN 1 Malang	EM3 - Sistem Komputer	IS-Sistem Cerdas	99.98825%
12	SMAN 1 Malang	EM4 - Jaringan Komputer dan Internet	IM-Manajemen Informasi	99.99216%
13	SMAN 1 Malang	EM5 - Analisis Data	IAS-Jaminan dan Keamanan Informasi	99.94721%
14	SMAN 1 Malang	EM6 - Algoritma dan Pemrograman	Interaksi HCI-Manusia-Komputer	99.98795%
15	SMAN 1 Malang	EM7 - Dampak Sosial Informatika	GV-Grafik dan Visualisasi	99.99071%
16	SMAN 1 Malang	EM8 - Praktik Lintas Bidang	DS-Struktur Diskrit	99.99014%
17	SMAN 1 Malang	EM1 - Berpikir Komputasional	CN-Ilmu Komputasi	99.99536%
18	SMAN 1 Malang	EM2 - Teknologi Infirmasi dan Komunikasi	AR-Arsitektur dan Organisasi	99.95188%
19	SMAN 2 Malang	EM3 - Sistem Komputer	AL-Algoritma dan Kompleksitas	85.70604%
20	SMAN 2 Malang	EM4 - Jaringan Komputer dan Internet	SP-Masalah Sosial dan Praktek Profesi	89.98950%
21	SMAN 2 Malang	EM5 - Analisis Data	Dasar-Dasar Sistem SF	99.99374%
22	SMAN 2 Malang	EM6 - Algoritma dan Pemrograman	SE-Rekayasa Perangkat Lunak	99.98646%
23	SMAN 2 Malang	EM7 - Dampak Sosial Informatika	Dasar-Dasar Pengembangan Perangkat Lunak SDF	99.99258%
24	SMAN 2 Malang	EM1 - Berpikir Komputasional	PL-Bahasa Pemrograman	99.99512%
...	...	...	...	...
3021	SMA Negeri 114 Jakarta Utara	EM1 - Berpikir Komputasional	Interaksi HCI-Manusia-Komputer	99.98339%
3022	SMA Negeri 114 Jakarta Utara	EM2 - Teknologi Infirmasi dan Komunikasi	GV-Grafik dan Visualisasi	99.98547%
3023	SMA Negeri 114 Jakarta Utara	EM3 - Sistem Komputer	DS-Struktur Diskrit	99.97486%

No.	Nama Sekolah	Elemen Materi	Nama Dokumen	% pemenuhan <i>Knowledge Area</i> ( <i>Word2Vec</i> )
3024	SMA Negeri 114 Jakarta Utara	EM4 - Jaringan Komputer dan Internet	CN-Ilmu Komputasi	99.98277%

Tabel 5.3 menjelaskan bahwa hasil persentase pemenuhan dari dokumen sebanyak 3024 dokumen yang diujikan dengan target pemenuhan *Knowledge Area* sebanyak 18 KA, maka dihasilkan pemenuhan yang sangat bervariasi namun cenderung signifikan dengan uji coba pertama dan kedua. Kurikulum *Computer Science Curricula 2013* dengan perbedaan hasil yang signifikan antara perbandingan dokumen elemen materi modul ajar informatika dengan konten *Knowledge Area*. Pada hasil pemenuhan memiliki tren *similarity* yang baik dengan nilai di atas 50% mendekati 100% pada masing-masing elemen materi yang dibandingkan. Visualisasi dari hasil pemenuhan *Knowledge Area* dapat dilihat pada Gambar 5.7. Gambar 5.7 menjelaskan pada pemenuhan *Knowledge Area* pada uji coba ketiga ini rata-rata mendapatkan skor di atas 50%. Secara umum hasil pemenuhan *Knowledge Area* dengan membandingkan elemen materi modul ajar informatika sebanyak delapan dokumen didapatkan skor pemenuhan rata-rata di atas 50% ke 100%. Hal ini dapat di analisa berdasarkan konten isi materi pada masing-masing elemen materi yang diukur *similarity*-nya dan skor pemenuhannya memenuhi standar kurikulum yang berlaku, sehingga dapat dianalisis bahwa pada uji coba pertama tren skor pemenuhan *Knowledge Area* yang sudah diujikan tinggi.



Gambar 5.7. Visualisasi pemenuhan Knowledge Area uicoba ketiga

## BAB VI PEMBAHASAN

### 6.1. Perbandingan Hasil Pemenuhan *Knowledge Area*

Pada pembahasan ini akan menyampaikan perbandingan hasil pemenuhan *Knowledge Area* antara kedua metode. Dimana hasil dari uji coba kedua metode akan memiliki karakteristik masing-masing yaitu persentase pemenuhan *Knowledge Area* yang bervariasi. Hasil uji coba dapat dilihat pada Tabel 6.1.

Tabel 6.1. Hasil persentase pemenuhan *Knowledge Area* menggunakan metode *Cosine Similarity* dan *Word2Vec* uji coba pertama

No.	Nama Sekolah	Elemen Materi	Nama Dokumen	% pemenuhan <i>Knowledge Area (Cosine-sim)</i>	% pemenuhan <i>Knowledge Area (Word2Vec)</i>
1	SMAN 1 Malang	EM1 - Berpikir Komputasional	AL-Algoritma dan Kompleksitas	19%	85.70604%
2	SMAN 1 Malang	EM2 - Teknologi Informasi dan Komunikasi	AL-Algoritma dan Kompleksitas	41%	85.92604%
3	SMAN 1 Malang	EM3 - Sistem Komputer	AL-Algoritma dan Kompleksitas	41%	85.92604%
4	SMAN 1 Malang	EM4 - Jaringan Komputer dan Internet	AL-Algoritma dan Kompleksitas	29%	85.80604%
5	SMAN 1 Malang	EM5 - Analisis Data	AL-Algoritma dan Kompleksitas	21%	85.72604%
6	SMAN 1 Malang	EM6 - Algoritma dan Pemrograman	AL-Algoritma dan Kompleksitas	50%	86.01604%
7	SMAN 1 Malang	EM7 - Dampak Sosial Informatika	AL-Algoritma dan Kompleksitas	23%	85.74604%
8	SMAN 1 Malang	EM8 - Praktik Lintas Bidang	AL-Algoritma dan Kompleksitas	33%	85.84604%
9	SMAN 1 Malang	EM1 - Berpikir Komputasional	SP-Masalah Sosial dan Praktek Profesi	41%	89.98950%
10	SMAN 1 Malang	EM2 - Teknologi Informasi dan Komunikasi	SP-Masalah Sosial dan Praktek Profesi	22%	89.79950%
11	SMAN 1 Malang	EM3 - Sistem Komputer	SP-Masalah Sosial dan Praktek Profesi	20%	89.77950%
...	...	...	...	...	...
3021	SMA Negeri 114 Jakarta Utara	EM5 - Analisis Data	AR-Arsitektur dan Organisasi	35%	99.96512%
3022	SMA Negeri 114 Jakarta Utara	EM6 - Algoritma dan Pemrograman	AR-Arsitektur dan Organisasi	30%	99.9711%
3023	SMA Negeri 114 Jakarta Utara	EM7 - Dampak Sosial Informatika	AR-Arsitektur dan Organisasi	28%	99.9111%
3024	SMA Negeri 114 Jakarta Utara	EM8 - Praktik Lintas Bidang	AR-Arsitektur dan Organisasi	24%	99.8811%

Tabel 6.1 menjelaskan hasil pemenuhan *Knowledge Area* dari kedua metode pada uji coba pertama yaitu membandingkan *term* dasar dengan gabungan *term*.

Perbedaan hasil pemenuhan dengan rentang yang jauh dapat dianalisis bahwasannya dalam konteks konten dokumen, faktor term dasar yang dihasilkan dari proses *preprocessing* dokumen teks berbahasa Indonesia memiliki kemunculan kata yang beragam sehingga istilah yang mungkin sama namun diindikasikan berbeda pada saat di proses *similarity* menggunakan metode. Faktor kedua masih terdapat istilah yang belum familiar pada teks Bahasa Indonesia karena dokumen pembandingnya adalah dokumen berbahasa Inggris yang di terjemahkan menjadi teks berbahasa Indonesia. Untuk uji coba kedua pada proses pemenuhan *Knowledge Area* dengan menggunakan kedua metode tersebut dapat dilihat pada Tabel 6.2.

Tabel 6.2. Hasil persentase pemenuhan *Knowledge Area* menggunakan metode *Cosine Similarity* dan *Word2Vec* uji coba kedua

No.	Nama Sekolah	Elemen Materi	Nama Dokumen	% pemenuhan <i>Knowledge Area (Cosine-sim)</i>	% pemenuhan <i>Knowledge Area (Word2Vec)</i>
1	SMAN 1 Malang	EM1 - Berpikir Komputasional	AL-Algorithm dan Kompleksitas	20%	84.70604%
2	SMAN 1 Malang	EM2 - Teknologi Informasi dan Komunikasi	AL-Algorithm dan Kompleksitas	42%	85.94504%
3	SMAN 1 Malang	EM3 - Sistem Komputer	AL-Algorithm dan Kompleksitas	40%	84.96054%
4	SMAN 1 Malang	EM4 - Jaringan Komputer dan Internet	AL-Algorithm dan Kompleksitas	31%	85.80604%
5	SMAN 1 Malang	EM5 - Analisis Data	AL-Algorithm dan Kompleksitas	26%	85.72604%
6	SMAN 1 Malang	EM6 - Algoritma dan Pemrograman	AL-Algorithm dan Kompleksitas	51%	86.10804%
7	SMAN 1 Malang	EM7 - Dampak Sosial Informatika	AL-Algorithm dan Kompleksitas	25%	85.88004%
8	SMAN 1 Malang	EM8 - Praktik Lintas Bidang	AL-Algorithm dan Kompleksitas	36%	85.90604%
9	SMAN 1 Malang	EM1 - Berpikir Komputasional	SP-Masalah Sosial dan Praktek Profesi	47%	89.95950%
10	SMAN 1 Malang	EM2 - Teknologi Informasi dan Komunikasi	SP-Masalah Sosial dan Praktek Profesi	22%	89.77950%
11	SMAN 1 Malang	EM3 - Sistem Komputer	SP-Masalah Sosial dan Praktek Profesi	25%	89.80950%
...	...	...	...	...	...
3021	SMA Negeri 114 Jakarta Utara	EM5 - Analisis Data	AR-Arsitektur dan Organisasi	35%	99.97812%
3022	SMA Negeri 114 Jakarta Utara	EM6 - Algoritma dan Pemrograman	AR-Arsitektur dan Organisasi	33%	99.9689%
3023	SMA Negeri 114 Jakarta Utara	EM7 - Dampak Sosial Informatika	AR-Arsitektur dan Organisasi	28%	99.92211%
3024	SMA Negeri 114 Jakarta Utara	EM8 - Praktik Lintas Bidang	AR-Arsitektur dan Organisasi	25%	99.89098%

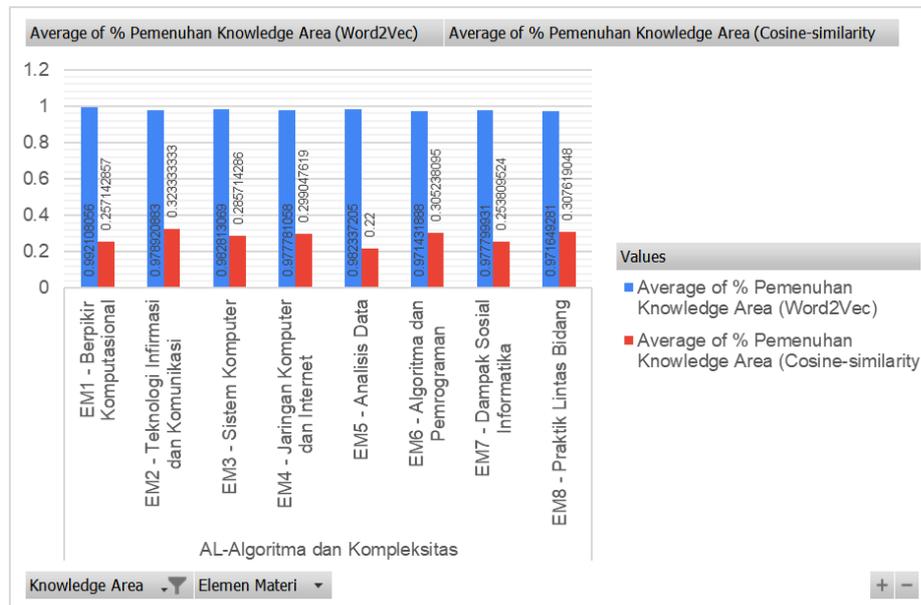
Tabel 6.2 menjelaskan hasil pemenuhan *Knowledge Area* dari kedua metode pada ujicoba kedua yaitu membandingkan *term* dasar dengan gabungan kalimat. Perbedaan hasil pemenuhan dengan rentang hasil delta yang berbeda jauh (cosine-sim dibawah 50% dan Word2Vec hamper mendekati 100%) dapat dianalisis bahwasannya dalam kontek konten dokumen, faktor gabungan kalimat yang diproses *similarity* pada dokumen teks berbahasa Indonesia memiliki kemunculan kata yang beragam sehingga istilah yang mungkin sama namun diindikasi berbeda pada saat di proses *similarity* menggunakan kedua metode. Pada ujicoba kedua ini faktor lain yang mempengaruhi hasil pemenuhannya adalah masih terdapat istilah yang belum familiar pada teks bahasa indonesia karena dokumen pembandingnya adalah dokumen berbahasa inggris yang di terjemahkan menjadi teks berbahasa Indonesia serta konten elemen materi yang dikembangkan masih sederhana. Pada ujicoba ketiga pada proses pemenuhan *Knowledge Area* dengan menggunakan kedua metode tersebut dapat dilihat pada Tabel 6.3.

Tabel 6.3. Hasil persentase pemenuhan *Knowledge Area* menggunakan metode *Cosine Similarity* dan Word2Vec ujicoba ketiga

No.	Nama Sekolah	Elemen Materi	Nama Dokumen	% pemenuhan <i>Knowledge Area</i> ( <i>Cosine-sim</i> )	% pemenuhan <i>Knowledge Area</i> ( <i>Word2Vec</i> )
1	SMAN 1 Malang	EM1 - Berpikir Komputasional	AL-Algorithm dan Kompleksitas	18%	85.70604%
2	SMAN 1 Malang	EM2 - Teknologi Infiriasi dan Komunikasi	AL-Algorithm dan Kompleksitas	40%	89.98950%
3	SMAN 1 Malang	EM3 - Sistem Komputer	AL-Algorithm dan Kompleksitas	42%	99.99374%
4	SMAN 1 Malang	EM4 - Jaringan Komputer dan Internet	AL-Algorithm dan Kompleksitas	30%	99.98646%
5	SMAN 1 Malang	EM5 - Analisis Data	AL-Algorithm dan Kompleksitas	25%	99.99258%
6	SMAN 1 Malang	EM6 - Algoritma dan Pemrograman	AL-Algorithm dan Kompleksitas	52%	99.99512%
7	SMAN 1 Malang	EM7 - Dampak Sosial Informatika	AL-Algorithm dan Kompleksitas	26%	99.99446%
8	SMAN 1 Malang	EM8 - Praktik Lintas Bidang	AL-Algorithm dan Kompleksitas	35%	99.99407%
9	SMAN 1 Malang	EM1 - Berpikir Komputasional	SP-Masalah Sosial dan Praktek Profesi	45%	99.98774%
10	SMAN 1 Malang	EM2 - Teknologi Infiriasi dan Komunikasi	SP-Masalah Sosial dan Praktek Profesi	25%	99.73085%
11	SMAN 1 Malang	EM3 - Sistem Komputer	SP-Masalah Sosial dan Praktek Profesi	22%	99.98825%
...	...	...	...	...	...

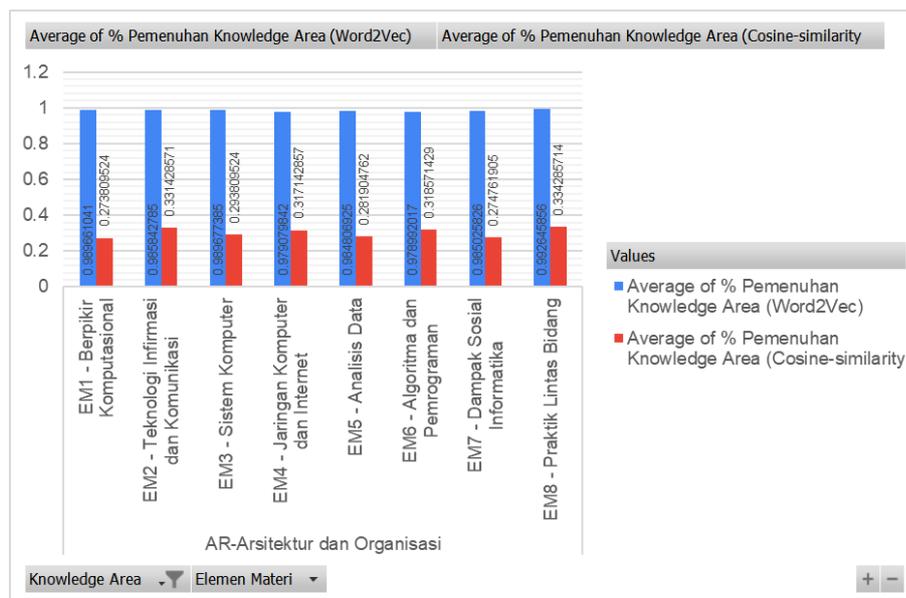
No.	Nama Sekolah	Elemen Materi	Nama Dokumen	% pemenuhan <i>Knowledge Area</i> (Cosine- <i>sim</i> )	% pemenuhan <i>Knowledge Area</i> ( <i>Word2Vec</i> )
3021	SMA Negeri 114 Jakarta Utara	EM5 - Analisis Data	AR-Arsitektur dan Organisasi	38%	99.98339%
3022	SMA Negeri 114 Jakarta Utara	EM6 - Algoritma dan Pemrograman	AR-Arsitektur dan Organisasi	31%	99.98547%
3023	SMA Negeri 114 Jakarta Utara	EM7 - Dampak Sosial Informatika	AR-Arsitektur dan Organisasi	26%	99.97486%
3024	SMA Negeri 114 Jakarta Utara	EM8 - Praktik Lintas Bidang	AR-Arsitektur dan Organisasi	27%	99.98277%

Tabel 6.3 menjelaskan hasil pemenuhan *Knowledge Area* dari kedua metode pada ujicoba ketiga yaitu membandingkan *term* dasar dengan gabungan paragraf. Perbedaan hasil pemenuhan dengan rentang hasil delta yang berbeda jauh (cosine-sim dibawah 50% dan Word2Vec hamper mendekati 100%) dapat dianalisis bahwasannya dalam konteks konten dokumen, faktor gabungan paragraf yang diproses *similarity* pada dokumen teks berbahasa Indonesia memiliki kemunculan kata yang beragam sehingga istilah yang mungkin sama namun diindikasikan berbeda pada saat di proses *similarity* menggunakan kedua metode. Pada ujicoba ketiga ini faktor lain yang mempengaruhi hasil pemenuhannya adalah masih terdapat istilah yang belum familiar pada teks bahasa Indonesia karena dokumen pembandingnya adalah dokumen berbahasa Inggris yang di terjemahkan menjadi teks berbahasa Indonesia serta konten elemen materi yang dikembangkan masih sederhana dan menggunakan ejaan bahasa pada elemen materi modul ajar yang dikembangkan belum lengkap. Untuk menjelaskan hasil ujicoba pertama untuk dokumen sebanyak 3024 dokumen, maka akan dibahas hasil per *Knowledge Area*. Pada pembahasan pertama yaitu pemenuhan *Knowledge Area* AL-Algoritma dan Kompleksitas yang dapat dilihat pada Gambar 6.1.



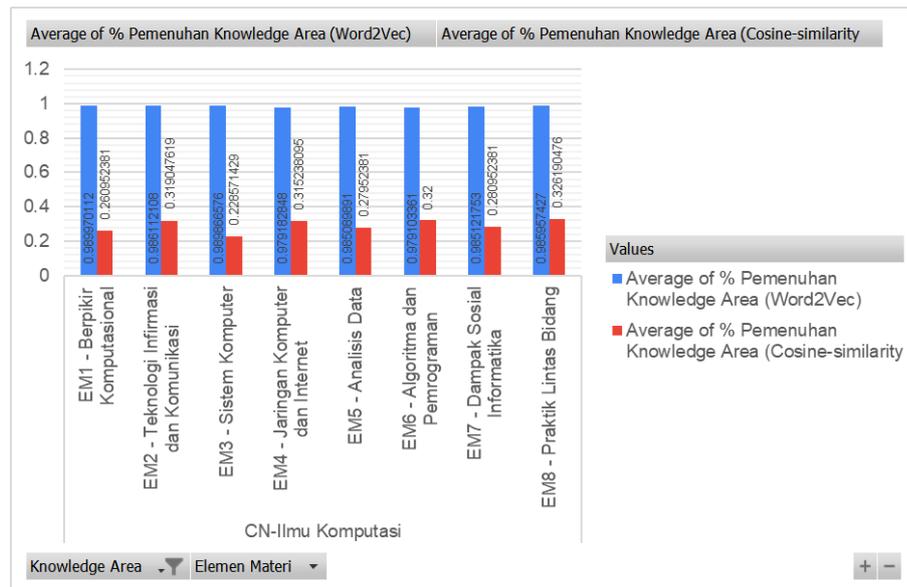
Gambar 6.1. Hasil rata-rata pemenuhan *Knowledge Area* AI-Algorithm dan Kompleksitas dari perbandingan elemen materi menggunakan metode *Cosine Similarity* dan *Word2Vec*

Gambar 6.1 menjelaskan bahwa perbedaan hasil pemenuhan KA antara kedua metode dengan rata-rata pemenuhan KA terbaik untuk KA. AI-Algorithm dan Kompleksitas dengan perbandingan dokumen EM1 - Berpikir Komputasional sebesar 99.2108% dengan pemenuhan tertinggi menggunakan *Word2Vec*.



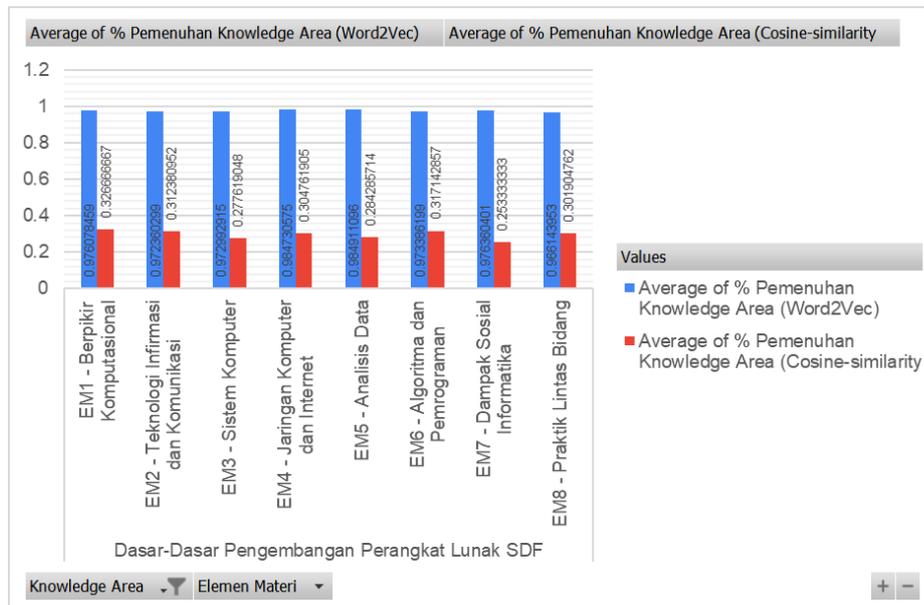
Gambar 6.2. Hasil rata-rata pemenuhan *Knowledge Area* AR-Arsitektur dan Organisasi dari perbandingan elemen materi menggunakan metode *Cosine Similarity* dan *Word2Vec*

Gambar 6.2 menjelaskan bahwa perbedaan hasil pemenuhan KA antara kedua metode dengan rata-rata pemenuhan KA terbaik untuk KA. AR-Arsitektur dan Organisasi dengan pembanding dokumen EM8 – Praktik Lintas Bidang sebesar 99.2645% dengan pemenuhan tertinggi menggunakan Word2Vec.



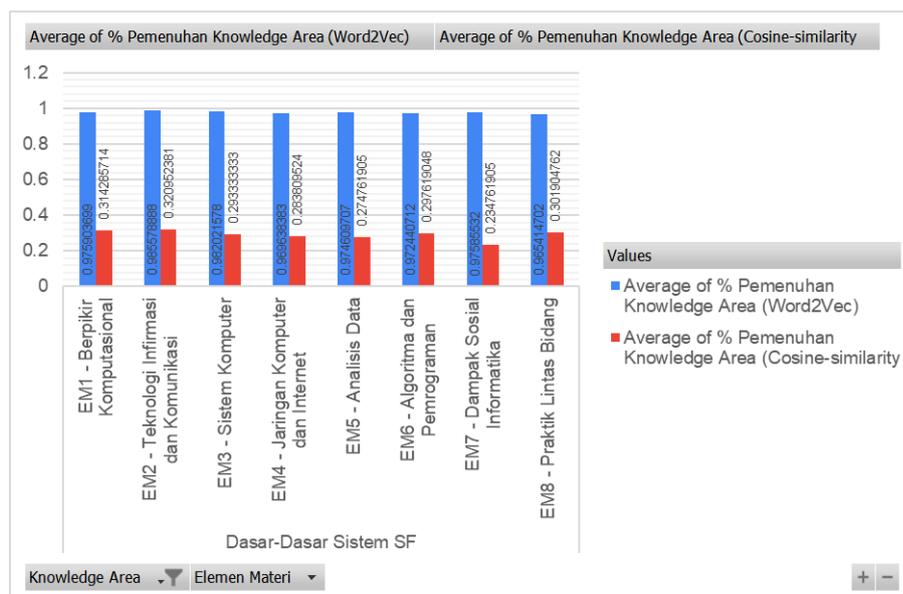
Gambar 6.3. Hasil rata-rata pemenuhan *Knowledge Area* CN-Ilmu Komputasi Organisasi dari pembandingan elemen materi menggunakan metode *Cosine Similarity* dan Word2Vec

Gambar 6.3 menjelaskan bahwa perbedaan hasil pemenuhan KA antara kedua metode dengan rata-rata pemenuhan KA terbaik untuk CN-Ilmu Komputasi dengan pembanding dokumen EM1 - Berpikir Komputasional sebesar 99.897% dengan pemenuhan tertinggi menggunakan Word2Vec.



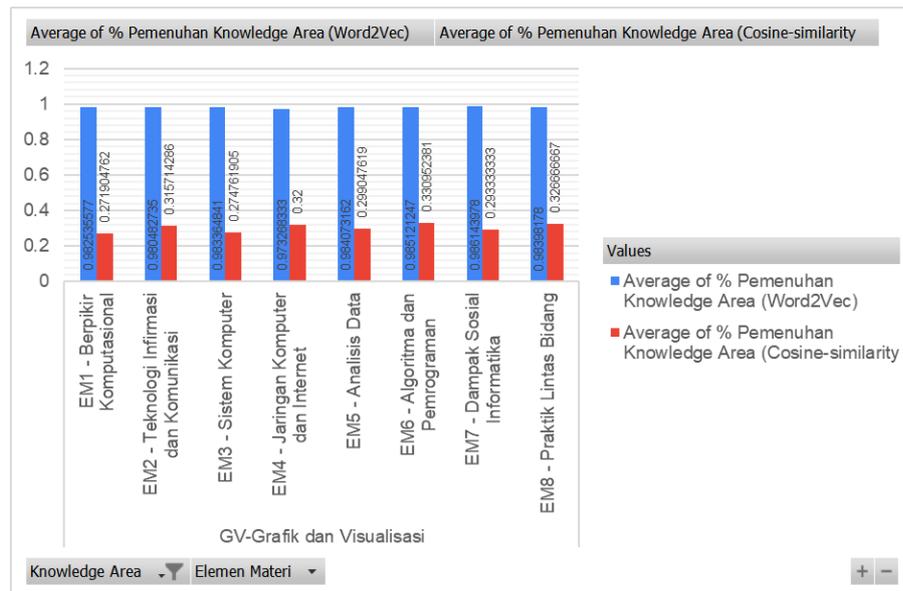
Gambar 6.4. Hasil pemenuhan *Knowledge Area* SDF- Dasar-Dasar Pengembangan Perangkat Lunak Organisasi dari perbandingan elemen materi menggunakan metode *Cosine Similarity* dan *Word2Vec*

Gambar 6.4 menjelaskan bahwa perbedaan hasil pemenuhan KA antara kedua metode dengan rata-rata pemenuhan KA terbaik untuk SDF- Dasar-Dasar Pengembangan Perangkat Lunak dengan perbandingan dokumen EM8 – Analisis Data sebesar 98.4911% dengan pemenuhan tertinggi menggunakan *Word2Vec*.



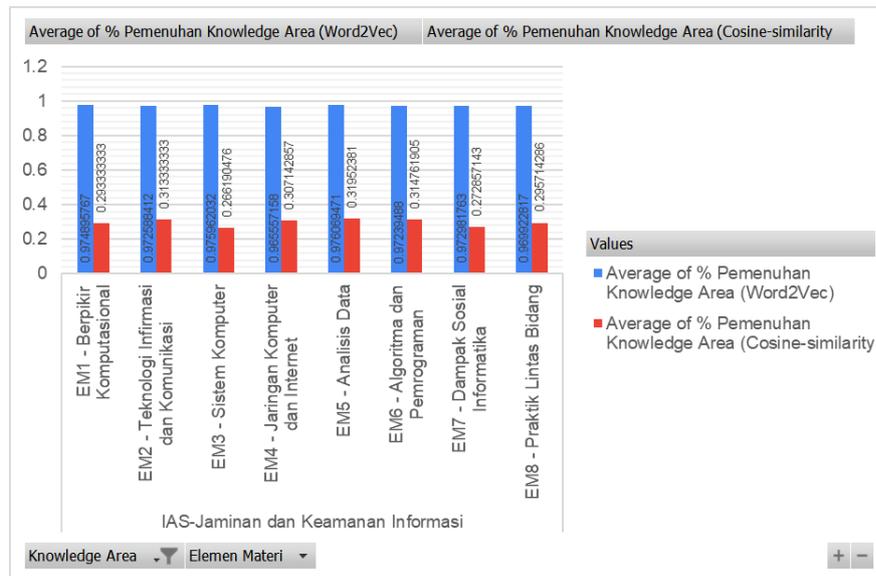
Gambar 6.5. Hasil pemenuhan *Knowledge Area* Dasar-Dasar Sistem-SF Organisasi dari perbandingan elemen materi menggunakan metode *Cosine Similarity* dan *Word2Vec*

Gambar 6.5 menjelaskan bahwa perbedaan hasil pemenuhan KA antara kedua metode dengan rata-rata pemenuhan KA terbaik untuk Dasar-Dasar Sistem-SF dengan pembandingan dokumen EM2 – Teknologi Informasi dan komunikasi sebesar 98.557% dengan pemenuhan tertinggi menggunakan Word2Vec.



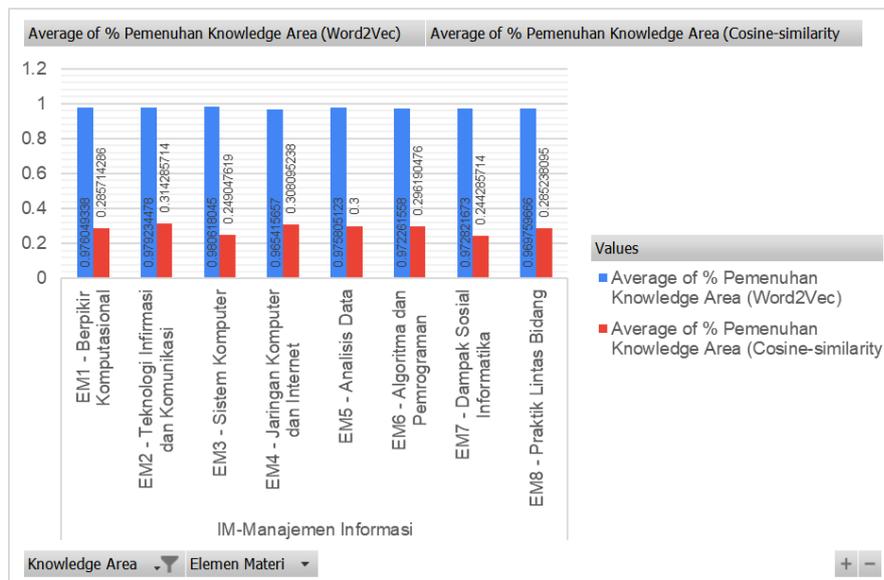
Gambar 6.6. Hasil pemenuhan *Knowledge Area* GV-Grafik dan Visualisasi Organisasi dari pembandingan elemen materi menggunakan metode *Cosine Similarity* dan *Word2Vec*

Gambar 6.6 menjelaskan bahwa perbedaan hasil pemenuhan KA antara kedua metode dengan rata-rata pemenuhan KA terbaik untuk GV-Grafik dan Visualisasi dengan pembandingan dokumen EM7 – Dampak Sosial Informatika sebesar 98.614% dengan pemenuhan tertinggi menggunakan Word2Vec.



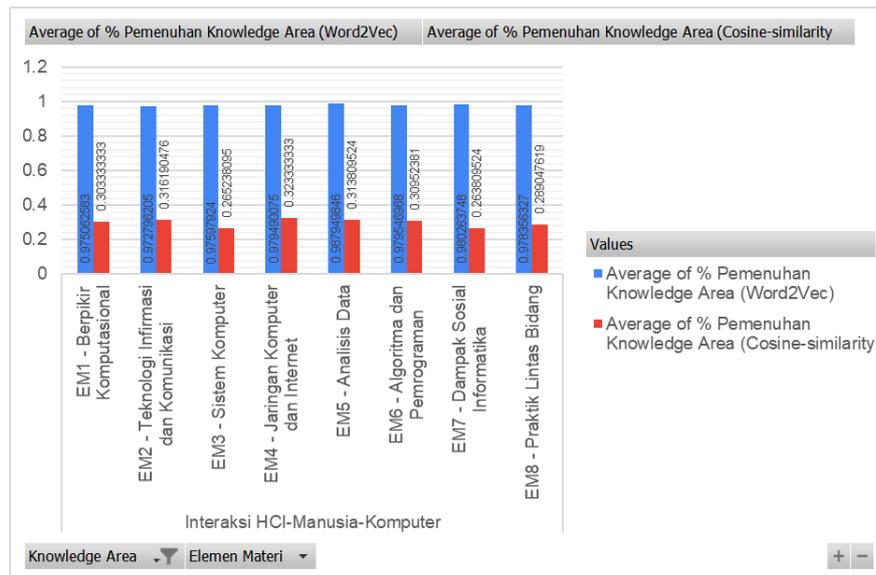
Gambar 6.7. Hasil pemenuhan *Knowledge Area* IAS-Jaringan dan Keamanan Informasi Organisasi dari perbandingan elemen materi menggunakan metode *Cosine Similarity* dan *Word2Vec*

Gambar 6.7 menjelaskan bahwa perbedaan hasil pemenuhan KA antara kedua metode dengan rata-rata pemenuhan KA terbaik untuk IAS-Jaringan dan Keamanan Informasi dengan pembandingan dokumen EM5 – Analisis Data sebesar 97.609% dengan pemenuhan terendah menggunakan *Word2Vec*.



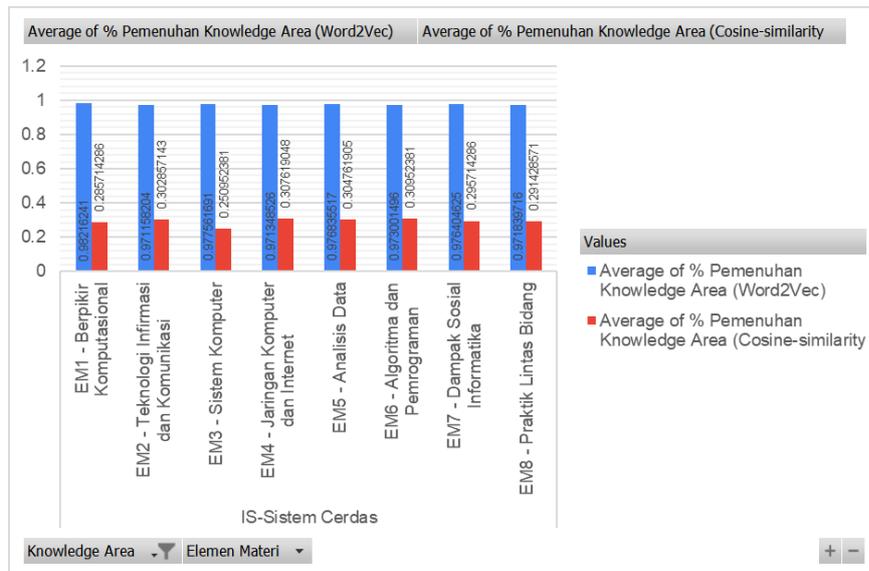
Gambar 6.8. Hasil pemenuhan *Knowledge Area* IM-Manajemen Informasi Organisasi dari perbandingan elemen materi menggunakan metode *Cosine Similarity* dan *Word2Vec*

Gambar 6.8 menjelaskan bahwa perbedaan hasil pemenuhan KA antara kedua metode dengan rata-rata pemenuhan KA terbaik untuk IM-Manajemen Informasi dengan pembandingan dokumen EM3 – Sistem Komputer sebesar 98.0618% dengan pemenuhan tertinggi menggunakan Word2Vec.



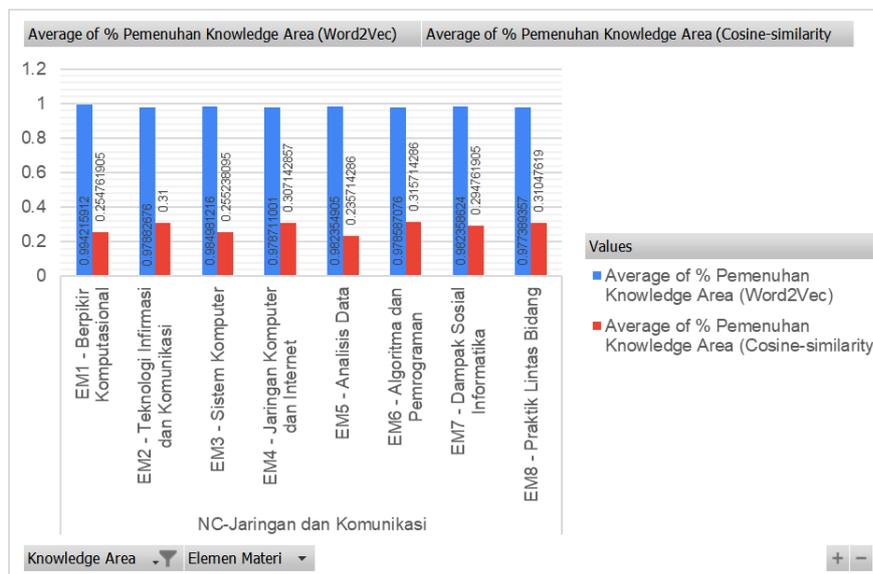
Gambar 6.9. Hasil pemenuhan *Knowledge Area* HCI-Interaksi Manusia dan Komputer Organisasi dari pembandingan elemen materi menggunakan metode *Cosine Similarity* dan *Word2Vec*

Gambar 6.9 menjelaskan bahwa perbedaan hasil pemenuhan KA antara kedua metode dengan rata-rata pemenuhan KA terbaik untuk HCI-Interaksi Manusia dan Komputer dengan pembandingan dokumen EM5 – Analisis Data sebesar 98.794% dengan pemenuhan tertinggi menggunakan Word2Vec.



Gambar 6.10. Hasil pemenuhan *Knowledge Area* IS-Sistem Cerdas Organisasi dari perbandingan elemen materi menggunakan metode *Cosine Similarity* dan *Word2Vec*

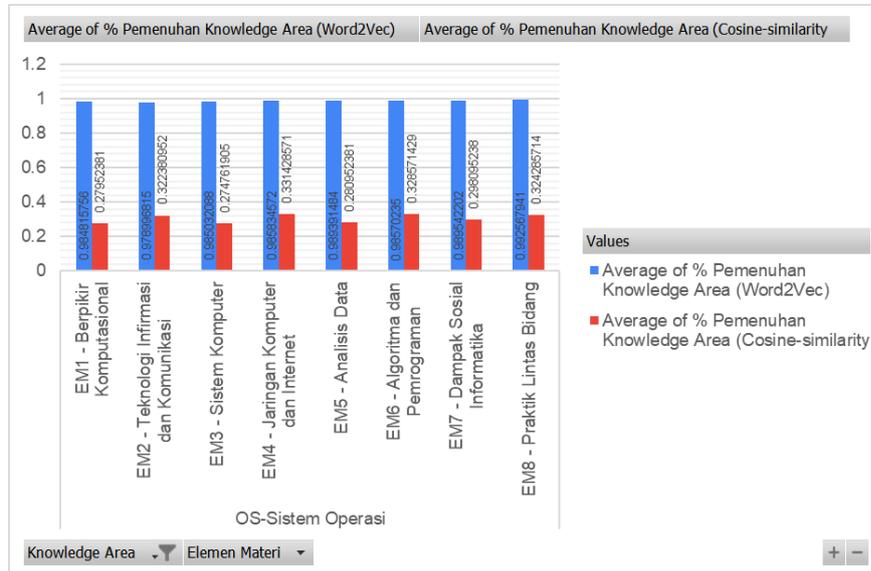
Gambar 6.10 menjelaskan bahwa perbedaan hasil pemenuhan KA antara kedua metode dengan rata-rata pemenuhan KA terbaik untuk IS-Sistem Cerdas dengan pembandingan dokumen EM1 - Berpikir Komputasional sebesar 98.216% dengan pemenuhan terendah menggunakan *Word2Vec*.



Gambar 6.11. Hasil pemenuhan *Knowledge Area* NC-Jaringan dan Komunikasi Organisasi dari perbandingan elemen materi menggunakan metode *Cosine Similarity* dan *Word2Vec*

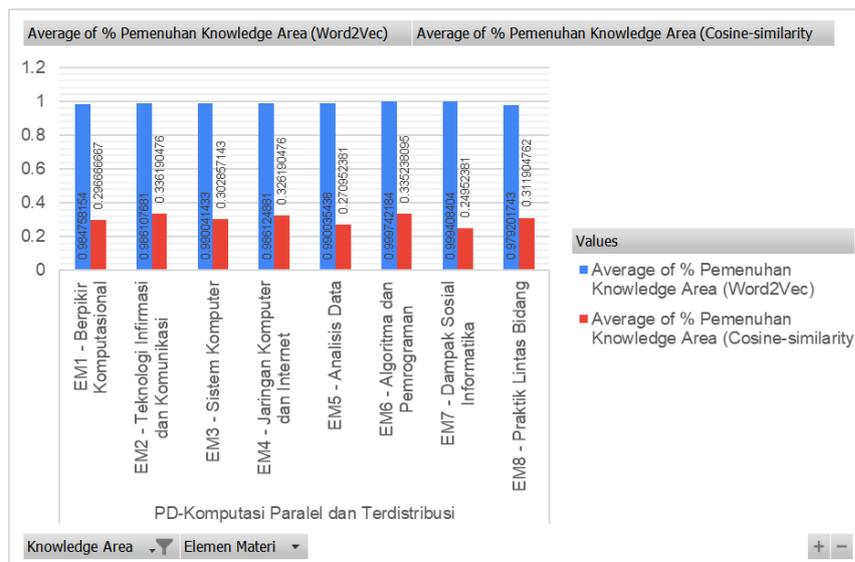
Gambar 6.11 menjelaskan bahwa perbedaan hasil pemenuhan KA antara kedua metode dengan rata-rata pemenuhan KA terbaik untuk NC-Jaringan dan

Komunikasi dengan pembandingan dokumen EM1 - Berpikir Komputasional sebesar 99.421% dengan pemenuhan tertinggi menggunakan Word2Vec.



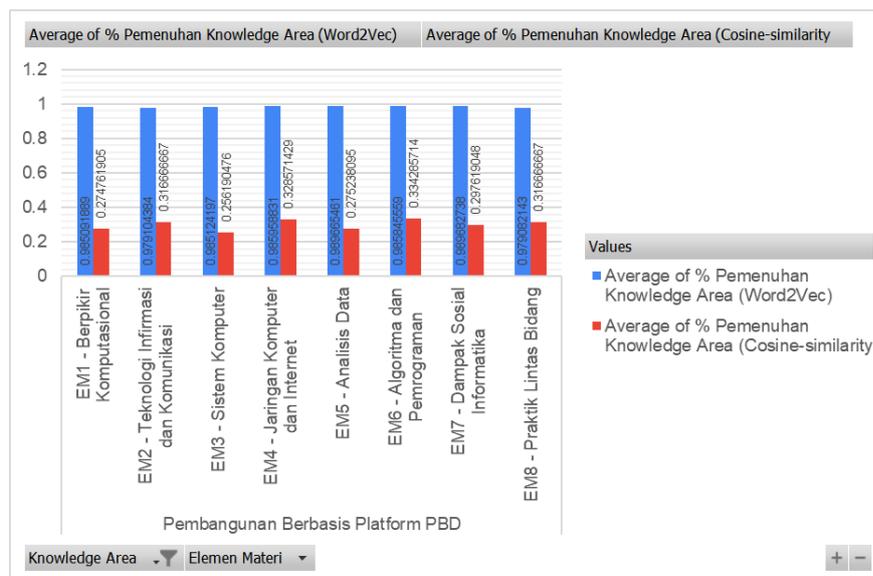
Gambar 6.12. Hasil pemenuhan *Knowledge Area* OS-Sistem Operasi Organisasi dari pembandingan elemen materi menggunakan metode *Cosine Similarity* dan *Word2Vec*

Gambar 6.12 menjelaskan bahwa perbedaan hasil pemenuhan KA antara kedua metode dengan rata-rata pemenuhan KA terbaik untuk OS-Sistem Operasi dengan pembandingan dokumen EM3 – Sistem Komputer sebesar 98.5032% dengan pemenuhan tertinggi menggunakan *Word2Vec*.



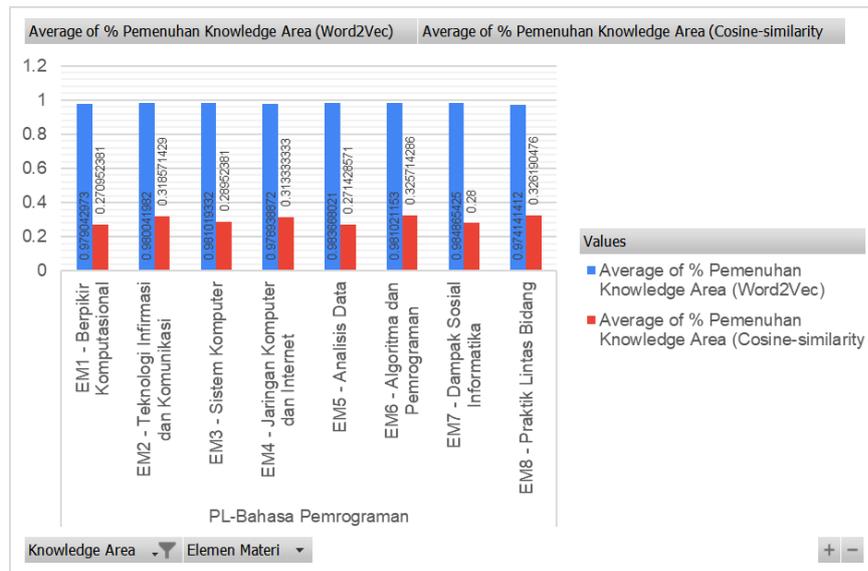
Gambar 6.13. Hasil pemenuhan *Knowledge Area* PD-Komputasi Paralel dan Terdistribusi Organisasi dari pembandingan elemen materi menggunakan metode *Cosine Similarity* dan *Word2Vec*

Gambar 6.13 menjelaskan bahwa perbedaan hasil pemenuhan KA antara kedua metode dengan rata-rata pemenuhan KA terbaik untuk PD-Komputasi Paralel dan Terdistribusi dengan pembanding dokumen EM6 – Algoritma dan pemrograman sebesar 99.974% dengan pemenuhan tertinggi menggunakan Word2Vec.



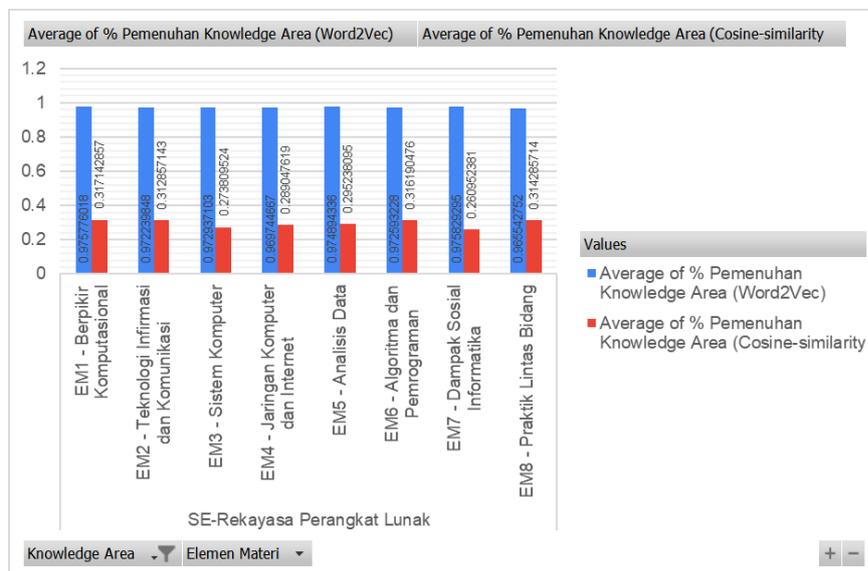
Gambar 6.14. Hasil pemenuhan Knowledge Area PBD-Pembangunan Berbasis Platform Organisasi dari pembandingan elemen materi menggunakan metode Cosine Similarity dan Word2Vec

Gambar 6.14 menjelaskan bahwa perbedaan hasil pemenuhan KA antara kedua metode dengan rata-rata pemenuhan KA terbaik untuk PBD-Pembangunan Berbasis Platform dengan pembanding dokumen EM7 – Dampak Sosial Informatika sebesar 98.968% dengan pemenuhan tertinggi menggunakan Word2Vec.



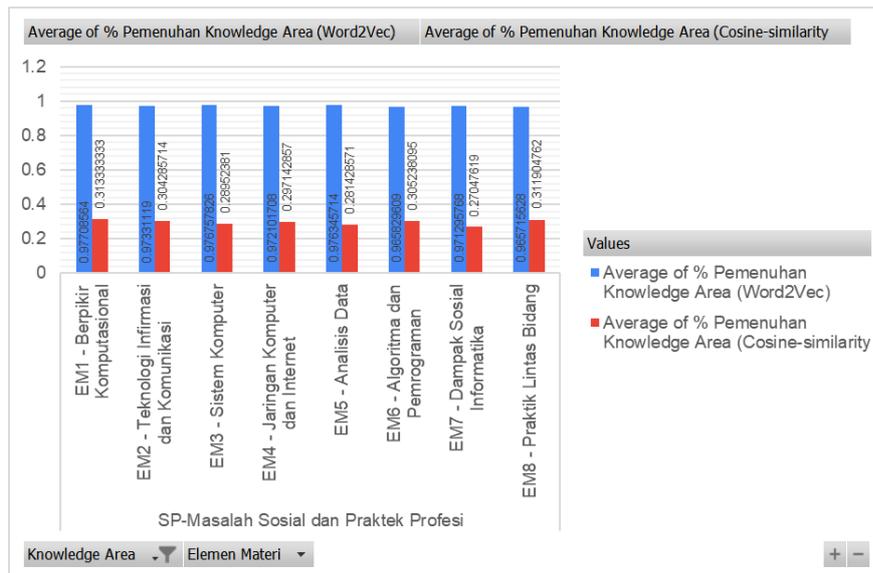
Gambar 6.15. Hasil pemenuhan *Knowledge Area* PL-Bahasa Pemrograman Organisasi dari perbandingan elemen materi menggunakan metode *Cosine Similarity* dan *Word2Vec*

Gambar 6.15 menjelaskan bahwa perbedaan hasil pemenuhan KA antara kedua metode dengan rata-rata pemenuhan KA terbaik untuk PL-Bahasa Pemrograman dengan pembandingan dokumen EM6 – Algoritma dan Pemrograman sebesar 99.102% dengan pemenuhan tertinggi menggunakan *Word2Vec*.



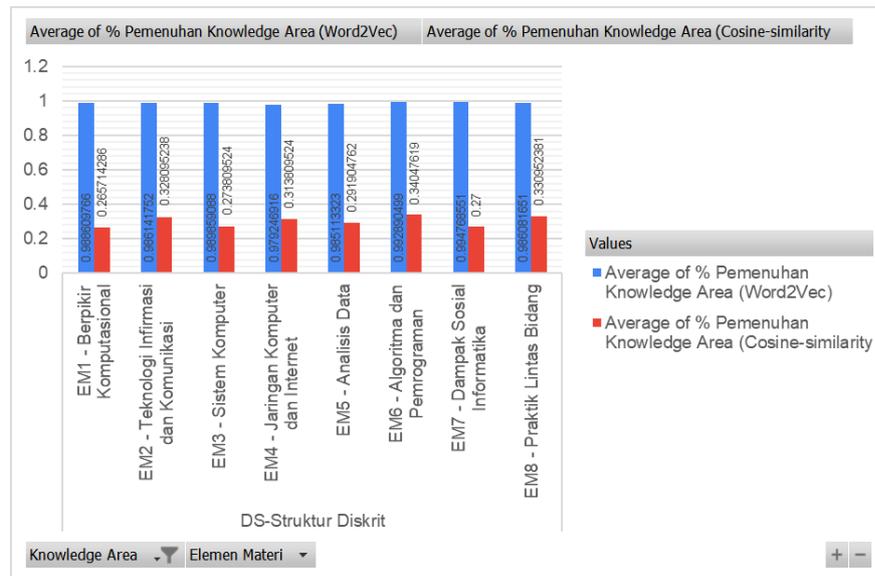
Gambar 6.16. Hasil pemenuhan *Knowledge Area* SE-Rekayasa Perangkat Lunak Organisasi dari perbandingan elemen materi menggunakan metode *Cosine Similarity* dan *Word2Vec*

Gambar 6.16 menjelaskan bahwa perbedaan hasil pemenuhan KA antara kedua metode dengan rata-rata pemenuhan KA terbaik untuk SE-Rekayasa Perangkat Lunak dengan pembanding dokumen EM8 – Praktik Lintas Bidang sebesar 96.655% dengan pemenuhan tertinggi menggunakan Word2Vec.



Gambar 6.17. Hasil pemenuhan *Knowledge Area* SP-Masalah Sosial dan Praktik Profesi Organisasi dari pembandingan elemen materi menggunakan metode *Cosine Similarity* dan *Word2Vec*

Gambar 6.17 menjelaskan bahwa perbedaan hasil pemenuhan KA antara kedua metode dengan rata-rata pemenuhan KA terbaik untuk SP-Masalah Sosial dan Praktik Profesi dengan pembanding dokumen EM7 – Dampak Sosial Informatika sebesar 97.129% dengan pemenuhan tertinggi menggunakan *Word2Vec*.



Gambar 6.18. Hasil pemenuhan *Knowledge Area* DS-Struktur Diskrit Organisasi dari perbandingan elemen materi menggunakan metode *Cosine Similarity* dan *Word2Vec*

Gambar 6.18 menjelaskan bahwa perbedaan hasil pemenuhan KA antara kedua metode dengan rata-rata pemenuhan KA terbaik untuk DS-Struktur Diskrit dengan pembandingan dokumen EM7 – Dampak Sosial Informatika sebesar 99.476% dengan pemenuhan tertinggi menggunakan *Word2Vec*.

## 6.2. Perbandingan Pengukuran Akurasi Pemenuhan *Knowledge Area*

Pada pembahasan pengukuran akurasi dalam pemenuhan *Knowledge Area* pada kedua metode menggunakan data demografi responden dan *Ground Truth* yang diperoleh dari narasumber guru informatika SMA yang dapat dilihat pada Tabel 6.4 dan Tabel 6.5.

Tabel 6.4. Sampel Demografi Responden

No.	Jenis Kelamin	Umur	Tingkat Pendidikan	Keahlian
1	Laki-laki	40	S1	Sistem Informasi
2	Laki-laki	39	S2	Sistem Cerdas
3	Laki-laki	46	S1	Software Engineering
4	Laki-laki	43	S1	Web dan Mobile
5	Perempuan	42	S1	Software Engineering
6	Perempuan	39	S2	Sistem Cerdas
7	Perempuan	35	S1	Sistem Informasi
8	Laki-laki	32	S1	Multimedia dan Teknologi Informasi
9	Laki-laki	36	S1	Sistem dan Jaringan
10	Laki-laki	38	S1	Software Engineering

Tabel 6.4 menjelaskan demografi latar belakang responden yang berasal dari profesi Guru Informatika dan Dosen Informatika.

Tabel 6.5. Sampel Ground Thruth

No.	Nama Sekolah	Elemen Materi	GT
1	SMAN 1 Malang	EM1 – Berpikir Komputasional	PL-Bahasa Pemrograman
2	SMAN 1 Malang	EM2 – Teknologi Informasi dan Komunikasi	PL-Bahasa Pemrograman
3	SMAN 1 Malang	EM3 – Sistem Komputer	OS-Sistem Operasi
4	SMAN 1 Malang	EM4 – Jaringan Komputer dan Internet	NC-Jaringan dan Komunikasi
5	SMAN 1 Malang	EM5 – Analisis Data	PL-Bahasa Pemrograman
6	SMAN 1 Malang	EM6 – Algoritma dan Pemrograman	PL-Bahasa Pemrograman
7	SMAN 1 Malang	EM7 – Dampak Sosial Informatika	PL-Bahasa Pemrograman
8	SMAN 1 Malang	EM8 – Praktik Lintas Bidang	PL-Bahasa Pemrograman
...	...	...	...
158	SMA Negeri 114 Jakarta Utara	EM6 – Algoritma dan Pemrograman	AL-Algoritma dan Kompleksitas
159	SMA Negeri 114 Jakarta Utara	EM7 – Dampak Sosial Informatika	SP-Masalah Sosial dan Praktek Profesi
160	SMA Negeri 114 Jakarta Utara	EM8 – Praktik Lintas Bidang	PBD-Pembangunan Berbasis Platform

Tabel 6.5 menjelaskan data Ground Truth dari narasumber sebanyak 160 baris item data dengan kelas poembanding sebanyak 13 kelas antara lain AL-Algoritma dan Kompleksitas, GV-Grafik dan Visualisasi, IAS-Jaminan dan Keamanan Informasi, IM-Manajemen Informasi, Interaksi HCI-Manusia-Komputer, IS-Sistem Cerdas, NC-Jaringan dan Komunikasi, OS-Sistem Operasi, PBD-Pembangunan Berbasis Platform, PD-Komputasi Paralel dan Terdistribusi, SDF-Dasar-Dasar Pengembangan Perangkat Lunak, SE-Rekayasa Perangkat Lunak, SF-Dasar-Dasar Sistem.

Tabel 6.6. Sampel Hasil Uji Validasi Ground Truth pada kelas KA. AL-Algorithm dan Kompleksitas

No.	Nama Sekolah	Nama Dokumen	Ground Thruth	Prediksi Cosine-similarity	Prediksi Word2Vec	TP	TN	FP	FN	TP1	TN1	FP1	FN1
1	SMAN 1 Malang	EM1 - BERPIKIR KOMPUTASIONAL	AL-Algorithm dan Kompleksitas	PL-Bahasa Pemrograman	SE-Rekayasa Perangkat Lunak	0	1	0	0	0	0	0	1
2	SMAN 1 Malang	EM2 - TEKNOLOGI INFORMASI DAN KOMUNIKASI	IM-Manajemen Informasi	PL-Bahasa Pemrograman	SDF - Dasar-Dasar Pengembangan Perangkat Lunak	0	1	0	0	0	1	0	0
3	SMAN 1 Malang	EM3 - SISTEM KOMPUTER	OS-Sistem Operasi	OS-Sistem Operasi	OS-Sistem Operasi	1	0	0	0	1	0	0	0
4	SMAN 1 Malang	EM4 - JARINGAN KOMPUTER DAN INTERNET	NC-Jaringan dan Komunikasi	NC-Jaringan dan Komunikasi	NC-Jaringan dan Komunikasi	1	0	0	0	1	0	1	0
5	SMAN 1 Malang	EM5 - ANALISIS DATA	IAS-Jaminan dan Keamanan Informasi	PL-Bahasa Pemrograman	OS-Sistem Operasi	0	1	0	0	0	0	1	0
6	SMAN 1 Malang	EM6 - ALGORITMA DAN PEMROGRAMAN	AL-Algorithm dan Kompleksitas	PL-Bahasa Pemrograman	GV-Grafik dan Visualisasi	0	1	0	0	0	1	0	0
7	SMAN 1 Malang	EM7 - DAMPAK SOSIAL INFORMATIKA	SP-Masalah Sosial dan Praktek Profesi	PL-Bahasa Pemrograman	PD-Komputasi Paralel dan Terdistribusi	0	1	0	0	0	1	0	0
8	SMAN 1 Malang	EM8 - PRAKTIK LINTAS BIDANG	SE-Rekayasa Perangkat Lunak	PL-Bahasa Pemrograman	PD-Komputasi Paralel dan Terdistribusi	0	1	0	0	0	1	0	0
9	SMAN 2 Malang	EM1 - BERPIKIR KOMPUTASIONAL	AL-Algorithm dan Kompleksitas	PL-Bahasa Pemrograman	SE-Rekayasa Perangkat Lunak	0	1	0	0	0	1	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
159	SMA Negeri 114 Jakarta Utara	EM7 - DAMPAK SOSIAL INFORMATIKA	SP-Masalah Sosial dan Praktek Profesi	PL-Bahasa Pemrograman	PD-Komputasi Paralel dan Terdistribusi	0	1	0	0	0	1	0	0
160	SMA Negeri 114 Jakarta Utara	EM8 - PRAKTIK LINTAS BIDANG	PBD-Pembangunan Berbasis Platform	PL-Bahasa Pemrograman	PD-Komputasi Paralel dan Terdistribusi	0	1	0	0	0	1	0	0
<b>Total</b>						<b>28</b>	<b>110</b>	<b>0</b>	<b>23</b>	<b>28</b>	<b>102</b>	<b>28</b>	<b>4</b>

Untuk menghitung tingkat akurasi metode Cosine-similarity dan Word2Vec terhadap *Knowledge Area*, maka hasil perhitungan akurasi sebagai berikut:

$$Akurasi = \frac{Jumlah (TP+TN)}{Jumlah (TP+FN)+Jumlah (FP+TN)} \times 100\% \quad (6.1)$$

Hasil dari menghitung keseluruhan akurasi pada *Knowledge Area* dapat dilihat pada Tabel 6.7.

Tabel 6.7. Hasil perhitungan akurasi metode Cosine-similarity dan Word2Vec

No.	Kelas <i>Knowledge Area</i>	%Pemenuhan <i>Knowledge Area</i>		Akurasi	
		Cosine-sim	Word2Vec	Cosine-sim	Word2Vec
1	AL-Algoritma dan Kompleksitas	28.1488	97.9355	0.857	0.802
2	GV-Grafik dan Visualisasi	29.6667	97.2305	0.913	0.811
3	IAS-Jaminan dan Keamanan Informasi	29.0179	97.5183	0.181	0.181
4	IM-Manajemen Informasi	29.7440	97.2445	0.175	0.292
5	Interaksi HCI-Manusia-Komputer	29.7262	97.5870	1.000	0.801
6	IS-Sistem Cerdas	29.9464	98.0342	0.919	0.806
7	NC-Jaringan dan Komunikasi	30.3690	98.9427	0.181	0.175
8	OS-Sistem Operasi	30.0000	98.4944	0.269	0.356
9	PBD-Pembangunan Berbasis Platform	30.5000	98.6485	0.831	0.767
10	PD-Komputasi Paralel dan Terdistribusi	28.5476	98.2178	0.175	0.425
11	SDF-Dasar-Dasar Pengembangan Perangkat Lunak	29.3571	97.5039	1.000	0.750

No.	Kelas <i>Knowledge Area</i>	%Pemenuhan <i>Knowledge Area</i>		Akurasi	
		Cosine-sim	Word2Vec	Cosine-sim	Word2Vec
12	SE-Rekayasa Perangkat Lunak	28.5357	97.3996	0.988	0.753
13	SF-Dasar-Dasar Sistem	29.7857	97.2547	0.994	0.744
<b>Rata-rata</b>		<b>29.6217</b>	<b>97.9988</b>	<b>0.6525</b>	<b>0.5895</b>

Dari hasil perhitungan akurasi kedua metode mendapatkan rata-rata akurasi pada metode Cosine-similarity sebesar 0.6525 sedangkan metode Word2Vec sebesar 0.5895.

### 6.3. Analisis Similarity pada Dokumen dalam pandangan Islam

Pada analisis similarity yang dilakukan pada dokumen kurikulum informatika pada jenjang SMA menurut pandangan islam memiliki dua prinsip dalam yang dapat dijadikan pedoman antara lain kejujuran yang tertuang dalam ayat Al-Qur'an surat Al-Baqarah ayat (2:42) (Bahrun Abubakar, 2008).

وَلَا تَلْبِسُوا الْحَقَّ بِالْبَاطِلِ وَتَكْتُمُوا الْحَقَّ وَأَنْتُمْ تَعْلَمُونَ

“Janganlah kamu campuradukkan kebenaran dengan kebatilan dan (jangan pula) kamu sembunyikan kebenaran, sedangkan kamu mengetahui(-nya).” (QS. Al-Baqarah 2:42)

وَلَا تَلْبِسُوا (Dan janganlah kalian campur aduk) – الْحَقَّ (barang yang hak) yang telah Kuturunkan kepada kalian – بِالْبَاطِلِ (dengan yang batil) yang kamu adakan – وَ (dan) jangan pula - تَكْتُمُوا الْحَقَّ (kalian sembunyikan yang hak itu) berupamsifat dan ciri-ciri Muhammad - وَأَنْتُمْ تَعْلَمُونَ (sedangkan kalian mengetahui) bahwa ia hak adanya.

Dalam pandangan islam, ayat tersebut memiliki makna bahwa ketika mengungkap suatu kebenaran dalam penelitian dibutuhkan kejujuran dan keadilan sehingga dihasilkan informasi yang akurat dan adil dalam pengungkapan kemiripan dokumen. Prinsip yang kedua adalah penghormatan atas ilmu dan kebenaran yang tertuang dalam QS. Al-Mujadila (58:11) (Bahrun Abubakar, 2008).

يَا أَيُّهَا الَّذِينَ آمَنُوا إِذَا قِيلَ لَكُمْ تَفَسَّحُوا فِي الْمَجَالِسِ فَافْسَحُوا يَفْسَحَ اللَّهُ لَكُمْ وَإِذَا قِيلَ فَانْشُرُوا فَانْشُرُوا  
رَفَعَ اللَّهُ الَّذِينَ آمَنُوا مِنْكُمْ وَالَّذِينَ أُوتُوا الْعِلْمَ دَرَجَاتٍ وَاللَّهُ بِمَا تَعْمَلُونَ خَبِيرٌ

*“Wahai orang-orang yang beriman! Apabila dikatakan kepadamu, “Berilah kelapangan di dalam majelis-majelis,” maka lapangkanlah, niscaya Allah akan memberi kelapangan untukmu. Dan apabila dikatakan, “Berdirilah kamu,” maka berdirilah, niscaya Allah akan mengangkat (derajat) orang-orang yang beriman di antarmu dan orang-orang yang diberi ilmu beberapa derajat. Dan Allah Mahateliti apa yang kamu kerjakan. QS. Al-Mujadila (58:11)*

Dalam pandangan islam menurut tafsir Aljalalain *يَا أَيُّهَا الَّذِينَ آمَنُوا إِذَا قِيلَ لَكُمْ تَفَسَّحُوا* (Hai orang-orang yang beriman, apabila dikatakan kepada kalian, "Berlapang-lapanglah) berluas-luaslah *فِي الْمَجَالِسِ* (dalam majelis") yaitu majelis tempat Nabi SAW. berada, dan majelis zikir sehingga orang-orang yang datang kepada kalian dapat tempat duduk. Menurut suatu qiraat lafal al-majalis dibaca al-majlis dalam bentuk mufrad *لَيْسَ فَافْسَحُوا يَفْسَحَ اللَّهُ لَكُمْ* (maka lapangkanlah, niscaya Allah akan memberi kelapangan untuk kalian) di surga nanti. *وَإِذَا قِيلَ فَانْشُرُوا* (Dan apabila dikatakan, "Berdirilah kalian") untuk melakukan salat dan hal-hal lainnya yang termasuk amal-amal kebaikan *فَانْشُرُوا* (maka berdirilah) menurut qiraat lainnya kedua-duanya dibaca fansyuzuu dengan memakai harakat damah pada huruf Syinnya *رَفَعَ اللَّهُ الَّذِينَ آمَنُوا مِنْكُمْ* (niscaya Allah akan meninggikan orang-orang yang beriman di antara kalian) karena ketaatannya dalam hal tersebut - *وَ* (dan) Dia meninggikan pula *الَّذِينَ أُوتُوا الْعِلْمَ دَرَجَاتٍ* (orang-orang yang diberi ilmu pengetahuan beberapa derajat) di surga nanti. *وَاللَّهُ بِمَا تَعْمَلُونَ خَبِيرٌ* (Dan Allah Maha Mengetahui apa yang kalian kerjakan).

Dalam ayat ini memiliki makna bahwa dalam melakukan penelitian, maka lakukan penghormatan atas kebenaran keilmuan yang sudah ada dan yang dikembangkan untuk memperoleh pengungkapan kebenaran keilmuan yang nyata. Konteks menganalisis similarity dokumen mempertimbangkan kedua prinsip, dimana proses dalam menganalisis similarity dokumen dibutuhkan keilmuan yang pas dan benar serta berpedoman kepada penelitian yang sudah pernah dilakukan sebelumnya untuk memperoleh kebenaran yang nyata atas penelitian yang dilakukan.

## BAB VII

### KESIMPULAN

#### 7.1. Kesimpulan

Hasil pengujian dan pembahasan analisis *similarity* menggunakan metode Cosine-similarity dan Word2Vec dapat disimpulkan sebagai berikut:

- a. Input data dokumen yang digunakan pada penelitian ini adalah dokumen kurikulum informatika pada jenjang SMA yang bersumber dari komunitas guru mata pelajaran informatika di Kota Malang, Platform Merdeka Mengajar (PMM) dan komunitas guru mata Pelajaran informatika di provinsi Jawa Timur. Dan input dokumen pembandingnya adalah dokumen *Computer Science Curricula 2013* yang bersumber dari <https://csed.acm.org/wp-content/uploads/2023/03/Version-Beta-v2.pdf>.
- b. Pada tahapan proses setelah input dokumen dilakukan adalah dengan melakukan *preprocessing* teks yaitu dengan melakukan *case folding*, *filtering*, *stemming* dan *tokenizing*. Dan hasil pada tahap ini adalah kumpulan term dasar pada masing-masing dokumen yang diproses.
- c. Tahap perhitungan *similarity* pada kedua metode dihasilkan skor *similarity* yang bervariasi. Dengan menggunakan metode Cosine-similarity dihasilkan kecenderungan skor dibawah 0.5 dengan rata-rata *similarity* sebesar 0.29606 sedangkan dengan menggunakan metode Word2Vec mendapatkan hasil skor *similarity* diatas 0.5 dengan rata-rata *similarity* sebesar 0.98449.
- d. Pada tahap pemenuhan *Knowledge Area* dari skor *similarity* yang didapatkan dengan metode Cosine-similarity pada seluruh *Knowledge Area* yang diujikan mendapatkan rata-rata pemenuhan sebesar 29.6%, sedangkan dengan metode Word2Vec pada seluruh *Knowledge Area* yang diujikan mendapatkan rata-rata pemenuhan sebesar 98.44%.
- e. Tahap pengukuran akurasi kedua metode, dengan metode Cosine-similarity mendapatkan rata-rata akurasi pada seluruh *Knowledge Area* yang diujikan sebesar 0.6525 Dan dengan metode Word2Vec

mendapatkan rata-rata akurasi pada seluruh *Knowledge Area* yang diujikan sebesar 0.5895.

- f. Faktor yang mempengaruhi hasil pemenuhan *Knowledge Area* dari dokumen yang diukur *similarity*-nya adalah konten elemen materi modul ajar informatika yang dikembangkan belum memenuhi secara utuh ruang lingkup materi pada *Knowledge Area*, struktur dan gaya bahasa pada konten dokumen masing-masing, dan kemunculan kata yang memiliki makna sama namun berbeda sukukata sehingga skor pemenuhan yang dihasilkan lebih kecil.
- g. Hasil dari analisis *similarity* ini dapat digunakan sebagai evaluasi pengukuran dalam pengembangan perangkat kurikulum informatika baik dilingkungan sekolah, komunitas guru informatika, Dinas Pendidikan dan Kementerian Pendidikan dan Kebudayaan Republik Indonesia.

## 7.2. Saran

Pada penelitian yang sudah dilakukan dapat dijadikan pertimbangan dalam melakukan evaluasi terukur dalam pengembangan perangkat kurikulum pada jenjang SMA. Dan oleh karena itu terdapat beberapa saran yang bisa dikembangkan pada penelitian ini antara lain:

- a. Penelitian ini dapat dilakukan dengan data input dokumen teks yang lebih banyak pada sekolah di seluruh wilayah Indonesia baik kurikulum informatika pada jenjang dasar fase C, D, dan jenjang menengah pada fase F. Sehingga akan didapatkan evaluasi terukur atas pengembangan capaian pembelajaran dan menjadi standar dalam mengambil kebijakan baik satuan Pendidikan, Dinas Pendidikan maupun Kementerian Pendidikan dan Kebudayaan dalam kualitas dalam pengembangan perangkat kurikulum yang berlaku.
- b. Pada analisis kemiripan dokumen kurikulum informatika dapat dilakukan dengan menambah tahapan dengan menggunakan metode Word2Vec yaitu arsitektur skip-gram sehingga pembandingan untuk mengoptimalkan dalam menganalisis *similarity* akan lebih baik.

## DAFTAR PUSTAKA

- Cios, K.J., Swiniarski, R.W., Pedrycz, W., Kurgan, L.A. (2007). Text Mining. In: Data Mining. Springer, Boston, MA. [https://doi.org/10.1007/978-0-387-36795-8\\_14](https://doi.org/10.1007/978-0-387-36795-8_14)
- Bahrin Abubakar, Al-Mahalli, Jalaluddin dan Jalaluddin As-Suyuti, Tafsir Al-Jalalain, diterjemahkan Bahrin Abubakar, Terjemahan tafsir Jalalain Berikut Asbabun Nuzul, Jilid 1. Bandung : Penerbit Sinar Baru Algensindo, 2008. Penerbit Sinar Baru Algensindo, 2008.
- Bahrin Abubakar, Al-Mahalli, Jalaluddin dan Jalaluddin As-Suyuti, Tafsir Al-Jalalain, diterjemahkan Bahrin Abubakar, Terjemahan tafsir Jalalain Berikut Asbabun Nuzul, Jilid 2. Bandung : Penerbit Sinar Baru Algensindo, 2008. Penerbit Sinar Baru Algensindo, 2008.
- Tardan, Pandu Prakoso, Alva Erwin, Kho I Eng, and Wahyu Muliady. “Automatic Text Summarization Based on Semantic Analysis Approach for Documents in Indonesian Language.” In *2013 International Conference on Information Technology and Electrical Engineering (ICITEE)*, 47–52. Yogyakarta, Indonesia: IEEE, 2013. <https://doi.org/10.1109/ICITEED.2013.6676209>.
- Fuddoly, Aini, Jafreezal Jaafar, and Norshuhani Zamin. “Keywords Similarity Based Topic Identification for Indonesian News Documents.” In *2013 European Modelling Symposium*, 14–20. Manchester, United Kingdom: IEEE, 2013. <https://doi.org/10.1109/EMS.2013.3>.
- Lahitani, Alfirna Rizqi, Adhistya Erna Permanasari, and Noor Akhmad Setiawan. “Cosine Similarity to Determine Similarity Measure: Study Case in Online Essay Assessment.” In *2016 4th International Conference on Cyber and IT Service Management*, 1–6. Bandung, Indonesia: IEEE, 2016. <https://doi.org/10.1109/CITSM.2016.7577578>.
- Gokul, P.P, B K Akhil, and Kumar K. M Shiva. “Sentence Similarity Detection in Malayalam Language Using Cosine Similarity.” In *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, 221–25. Bangalore: IEEE, 2017. <https://doi.org/10.1109/RTEICT.2017.8256590>.
- Sohangir, Sahar, and Dingding Wang. “Improved Sqrt-Cosine Similarity Measurement.” *Journal of Big Data* 4, no. 1 (December 2017): 25. <https://doi.org/10.1186/s40537-017-0083-6>.
- Suyanto, 1974- (pengarang). Data mining : untuk klasifikasi dan klusterisasi data / Suyanto. Bandung :: Penerbit Informatika,, 2017.

- Astawa, Ida Bagus Made dan Adnyana. 2018. Belajar dan Pembelajaran. Depok: Rajawali Grafindo Persada.
- Slamet, C, A R Atmadja, D S Maylawati, R S Lestari, W Darmalaksana, and M A Ramdhani. "Automated Text Summarization for Indonesian Article Using Vector Space Model." *IOP Conference Series: Materials Science and Engineering* 288 (January 2018): 012037. <https://doi.org/10.1088/1757-899X/288/1/012037>.
- Soyusiawaty, Dewi, Anna Hendri Soleliza Jones, and Pangah Widiandana. "Similarity Detection of Student Assignments Using Rocchio Method." In *2018 12th International Conference on Telecommunication Systems, Services, and Applications (TSSA)*, 1–4. Yogyakarta, Indonesia: IEEE, 2018. <https://doi.org/10.1109/TSSA.2018.8708827>.
- Amalia, A, D Gunawan, Y Fithri, and I Aulia. "Automated Bahasa Indonesia Essay Evaluation with Latent Semantic Analysis." *Journal of Physics: Conference Series* 1235, no. 1 (June 1, 2019): 012100. <https://doi.org/10.1088/1742-6596/1235/1/012100>.
- Ramadhanti, Nahda Rosa, and Siti Mariyah. "Document Similarity Detection Using Indonesian Language Word2vec Model." In *2019 3rd International Conference on Informatics and Computational Sciences (ICICoS)*, 1–6. Semarang, Indonesia: IEEE, 2019. <https://doi.org/10.1109/ICICoS48119.2019.8982432>.
- Rosid, Mochamad Alfian, Arif Senja Fitriani, Ika Ratna Indra Astutik, Nasrudin Iqrok Mulloh, and Haris Ahmad Gozali. "Improving Text Preprocessing For Student Complaint Document Classification Using Sastrawi." *IOP Conference Series: Materials Science and Engineering* 874, no. 1 (June 1, 2020): 012017. <https://doi.org/10.1088/1757-899X/874/1/012017>.
- Mubai, Akrimullah & Jalinus, Nizwardi & Ambiyar, Ambiyar & Wakhinuddin, Wakhinuddin & Abdullah, Rijal & Rizal, Fahmi & Waskito, Waskito. (2021). Implementasi Model Cipp Dalam Evaluasi Kurikulum Pendidikan Teknik Informatika. *EDUKATIF : JURNAL ILMU PENDIDIKAN*. 3. 1383-1394. 10.31004/edukatif.v3i4.549.
- Madyatmadja, Evaristus Didik, Bernardo Nugroho Yahya, and Cristofer Wijaya. "Contextual Text Analytics Framework for Citizen Report Classification: A Case Study Using the Indonesian Language." *IEEE Access* 10 (2022): 31432–44. <https://doi.org/10.1109/ACCESS.2022.3158940>.
- D. A. Diartono, I. Nugroho, and J. A. Razaq, "PENINGKATAN HASIL SISTEM TEMU KEMBALI INFORMASI BERBASIS PADA KATA MAJEMUK MENGGUNAKAN JACCARD SIMILARITY," *JDI*, vol. 14, no. 1, pp. 1–10, Mar. 2022, doi: 10.35315/informatika.v14i1.9160.

- P. N. Raharjo, A. Handoyo, and H. Juwiantho, "Sistem Rekomendasi Content Based Filtering Pekerjaan dan Tenaga Kerja Potensial menggunakan Cosine Similarity".
- D. S. E. I. Faisal Rahutomo and Obby Auliyaur Rohman, "EVALUASI FITUR WORD2VEC PADA SISTEM UJIAN ESAI ONLINE," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. Volume 04, Nomor 01, pp. 36–45, Jun. 2019, doi: <https://doi.org/10.29100/jipi.v4i1.1098>.
- S. M. D. Abdul Kadir and S. Vahlepi, "Mendalami Informasi dengan Bertabayyun Menurut Al-Qur'an di Tinjau Dari Tafsir Klasik dan Kontemporer," *JIUBJ*, vol. 21, no. 2, p. 825, Jul. 2021, doi: 10.33087/jiubj.v21i2.1570.
- Han, J, Kamber, M, & Pei, J. 2012. *Data Mining: Concept and Techniques*, Third Edition. Waltham: Morgan Kaufmann Publishers
- C. Crysdiyan, "The evaluation of higher education policy to drive university entrepreneurial activities in information technology learning," *Cogent Education*, vol. 9, no. 1, p. 2104012, Dec. 2022, <https://doi:10.1080/2331186X.2022.2104012>.
- Faisal, Muhammad & Malang, Ibrahim & Sulthan, Maulana & Malang, Ibrahim & Amini, Fauziyah & Amin, M & Sedayu, Agung & Nugroho, Fresy. (2020). Plagiarism Detection Using Manber and Winnowing Algorithm. *International Journal of Advanced Science and Technology*. 29. 2130-2136.

# **LAMPIRAN**

### **A. Studi Literatur**

Studi literatur yang dilakukan selama penelitian dapat diakses pada link [https://sman4mlg-my.sharepoint.com/:f:/g/personal/dhani\\_sman4malang\\_sch\\_id/EjvhgU3MNJpBuDgHkW5Y6igB-Ru3GVyjfrzVGZD-pAFwww?e=MuY3Pi](https://sman4mlg-my.sharepoint.com/:f:/g/personal/dhani_sman4malang_sch_id/EjvhgU3MNJpBuDgHkW5Y6igB-Ru3GVyjfrzVGZD-pAFwww?e=MuY3Pi)

### **B. Sumber Dokumen**

1. Dokumen *Computer Science Curricula 2013* diambil dari link [https://www.acm.org/binaries/content/assets/education/cs2013\\_web\\_final.pdf](https://www.acm.org/binaries/content/assets/education/cs2013_web_final.pdf)
2. Dokumen Modul Ajar Informatika dapat diakses pada link [https://sman4mlg-my.sharepoint.com/:f:/g/personal/dhani\\_sman4malang\\_sch\\_id/EhTHg6401nVII5k3WLjwb5oBPx7brUvI75tJBULgZ1nv\\_w?e=IRh4Zf](https://sman4mlg-my.sharepoint.com/:f:/g/personal/dhani_sman4malang_sch_id/EhTHg6401nVII5k3WLjwb5oBPx7brUvI75tJBULgZ1nv_w?e=IRh4Zf)

### **C. Implementasi Sistem**

1. Kode program implementasi sistem keseluruhan dapat diakses pada link [https://sman4mlg-my.sharepoint.com/:f:/g/personal/dhani\\_sman4malang\\_sch\\_id/EoKBWi5KzrBAkpvDaADqKS8BXuLQPWjs41qHGblgJAV8Hw?e=XpKeyO](https://sman4mlg-my.sharepoint.com/:f:/g/personal/dhani_sman4malang_sch_id/EoKBWi5KzrBAkpvDaADqKS8BXuLQPWjs41qHGblgJAV8Hw?e=XpKeyO)
2. Tampilan User Interface dapat diakses pada link [https://sman4mlg-my.sharepoint.com/:f:/g/personal/dhani\\_sman4malang\\_sch\\_id/EiJXIOC4q7dInG1TvrJBzZEBxHzp7EP7h9MNWEAswPghmw?e=U1IPV1](https://sman4mlg-my.sharepoint.com/:f:/g/personal/dhani_sman4malang_sch_id/EiJXIOC4q7dInG1TvrJBzZEBxHzp7EP7h9MNWEAswPghmw?e=U1IPV1)

### **D. Hasil Ujicoba Metode**

Hasil Ujicoba keseluruhan dapat diakses pada link [https://sman4mlg-my.sharepoint.com/:f:/g/personal/dhani\\_sman4malang\\_sch\\_id/EphjTNKkSmVFkpf6KEgduskBMwlt-qrxIQo5sOGvueXruA?e=CeYYoi](https://sman4mlg-my.sharepoint.com/:f:/g/personal/dhani_sman4malang_sch_id/EphjTNKkSmVFkpf6KEgduskBMwlt-qrxIQo5sOGvueXruA?e=CeYYoi)

### **E. Hasil Perhitungan Akurasi Metode**

Hasil perhitungan akurasi dapat diakses pada link [https://sman4mlg-my.sharepoint.com/:f:/g/personal/dhani\\_sman4malang\\_sch\\_id/ErcWITl-uw1Bn24acXMI1z4BGgn9rIoK4fMDIbOx5WeYJQ?e=RnJ3AX](https://sman4mlg-my.sharepoint.com/:f:/g/personal/dhani_sman4malang_sch_id/ErcWITl-uw1Bn24acXMI1z4BGgn9rIoK4fMDIbOx5WeYJQ?e=RnJ3AX)