

**PENGECEKAN KESAMAAN MAKNA
KALIMAT DARI REFERENSI JURNAL
DENGAN METODE *COSINE SIMILARITY***

SKRIPSI



**Disusun Oleh :
HAFIDZ HAIKAL
(1955201000977)**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS MERDEKA PASURUAN**

2023

**PENGECEKAN KESAMAAN MAKNA KALIMAT DARI REFERENSI
JURNAL DENGAN METODE *COSINE SIMILARITY***

SKRIPSI



OLEH :
HAFIDZ HAIKAL
NPM : 1955201000977

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS MERDEKA PASURUAN**

2023

**PENGECEKAN KESAMAAN MAKNA KALIMAT DARI REFERENSI
JURNAL DENGAN METODE *COSINE SIMILARITY***

SKRIPSI

Diajukan sebagai salah satu persyaratan
Dalam menyelesaikan program Sarjana Komputer (S.Kom)



Oleh :

Hafidz Haikal

NPM 1955201000977

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS MERDEKA PASURUAN**

2023

PERSETUJUAN PEMBIMBING SKRIPSI

**PENGECEKAN KESAMAAN MAKNA KALIMAT DARI REFERENSI
JURNAL DENGAN METODE *COSINE SIMILARITY***

Nama : Hafidz Haikal
NPM : 1955201000977
Program Studi : Informatika

Pasuruan, 04 Juli 2023

Telah diperiksa dan disetujui Oleh :

Pembimbing I



Moh. Zoqi Sarwani, S.Pd., M.Kom.
NIDN. 0716049101

Pembimbing II



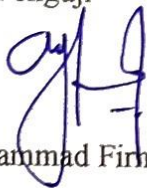
Dian Ahkam Sani, S.Kom., M.Kom.
NIDN. 0724079202

PERSETUJUAN DAN PENGESAHAN SKRIPSI

Skripsi oleh Hafidz Haikal ini
Telah dipertahankan di depan dewan penguji
Pada tanggal 04 juli 2023

Dewan Penguji

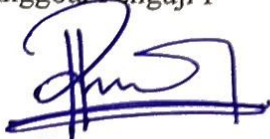
Ketua Penguji



(Mochammad Firman Arif, S.Kom., M.Kom.)

NIDN. 0704029201

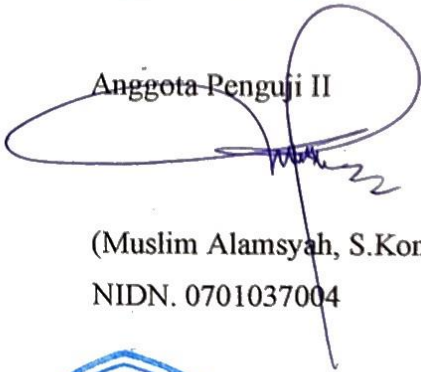
Anggota Penguji I



(Rudi Hariyanto, S.Kom., M.T.)

NIDN. 0718018401

Anggota Penguji II



(Muslim Alamsyah, S.Kom., M.Kom.)

NIDN. 0701037004

Mengesahkan,
Dekan Fakultas Teknologi Informasi



(Muhammad Misdrum, S.Kom., M.Kom.)

NIDN. 0717046704

Mengetahui,

Ketua Program Studi Informatika



(Dian Ahkam Sani, S.Kom., M.Kom.)

NIDN. 0724079202

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Hafidz Haikal
NPM : 1955201000977
Jurusan/Program Studi : Informatika
Fakultas/Program : Teknologi Informasi

menyatakan dengan sesungguhnya bahwa **skripsi** yang saya tulis ini benar-benar tulisan saya, dan bukan merupakan plagiasi baik sebagian atau seluruhnya.

Apabila di kemudian hari terbukti atau dapat dibuktikan **bahwa skripsi** ini hasil plagiasi, baik sebagian atau seluruhnya, maka saya bersedia menerima sanksi atas perbuatan tersebut sesuai dengan ketentuan yang berlaku.

Pasuruan, 04 Juli 2023

Yang membuat pernyataan



HAFIDZ HAIKAL

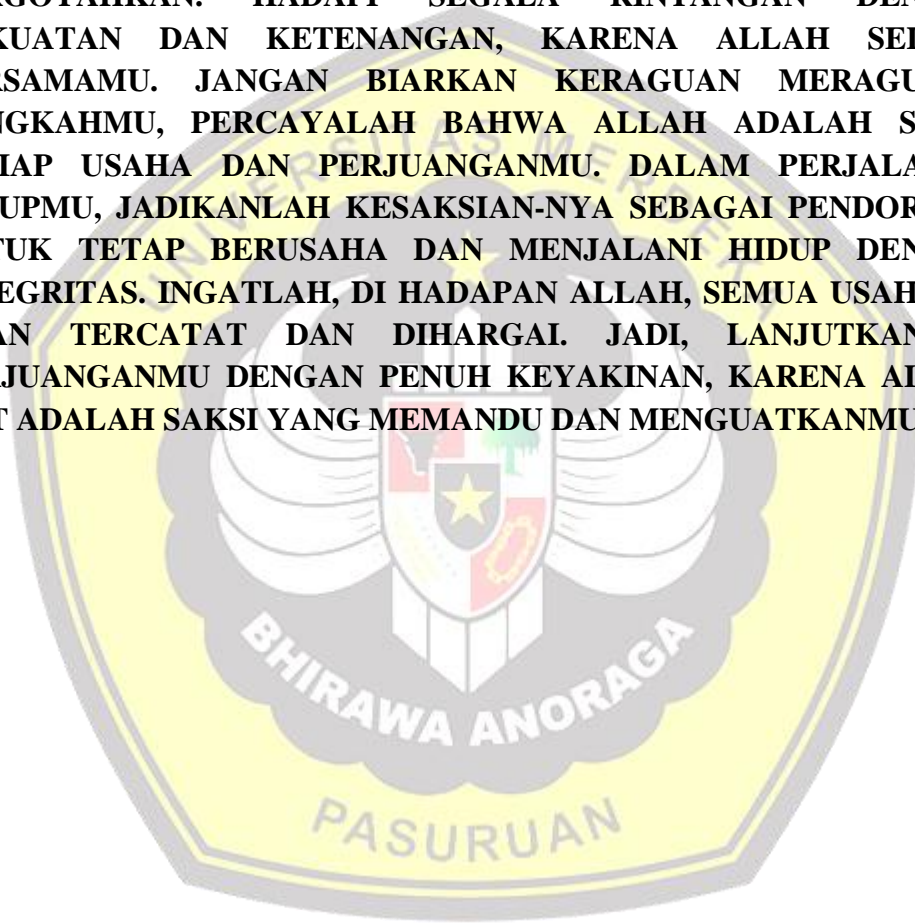
1955201000977

MOTTO

“WAKAFA BILLAHI SYAHIDA”

“DAN CUKUPLAH ALLAH SWT SEBAGAI SAKSI”

ALLAH SWT SEBAGAI SAKSI SUDAH CUKUP, TUMBUHLAH DENGAN KEYAKINAN DAN SEMANGAT YANG TAK TERGOYAHKAN. HADAPI SEGALA RINTANGAN DENGAN KEKUATAN DAN KETENANGAN, KARENA ALLAH SELALU BERSAMAMU. JANGAN BIARKAN KERAGUAN MERAGUKAN LANGKAHMU, PERCAYALAH BAHWA ALLAH ADALAH SAKSI SETIAP USAHA DAN PERJUANGANMU. DALAM PERJALANAN HIDUPMU, JADIKANLAH KESAKSIAN-NYA SEBAGAI PENDORONG UNTUK TETAP BERUSAHA DAN MENJALANI HIDUP DENGAN INTEGRITAS. INGATLAH, DI HADAPAN ALLAH, SEMUA USAHAMU AKAN TERCATAT DAN DIHARGAI. JADI, LANJUTKANLAH PERJUANGANMU DENGAN PENUH KEYAKINAN, KARENA ALLAH SWT ADALAH SAKSI YANG MEMANDU DAN MENGUATKANMU.



PERSEMBAHAN

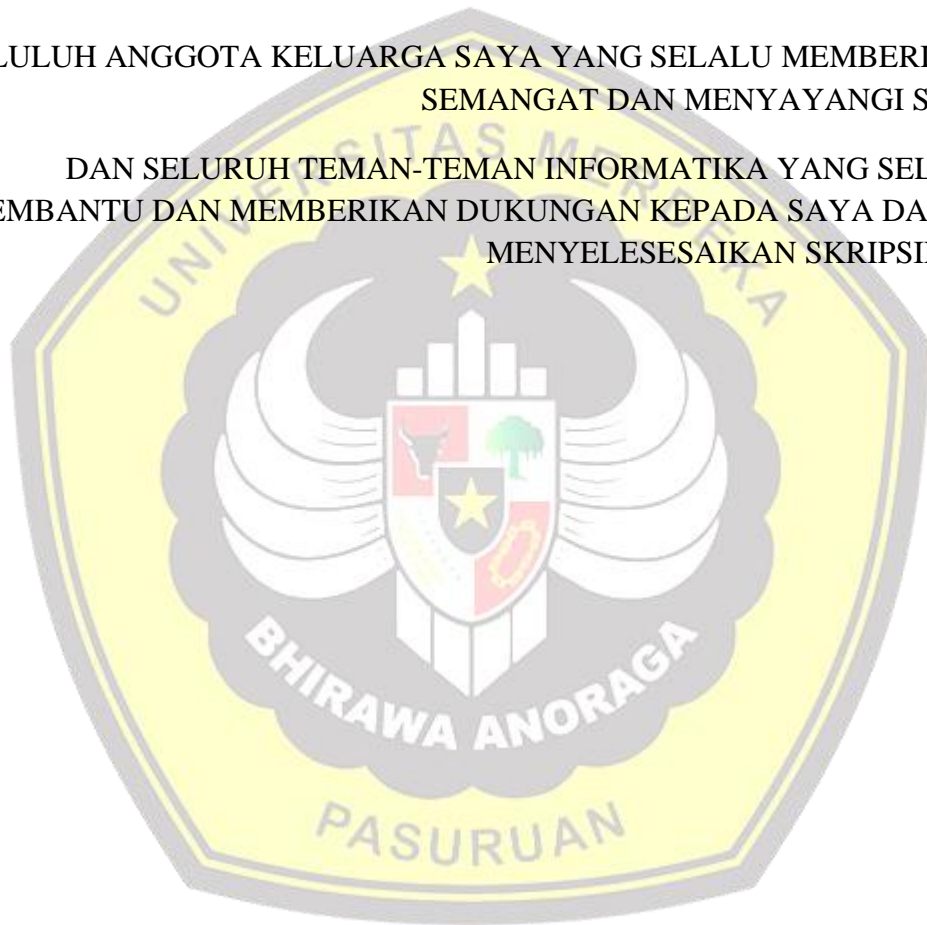
SKRIPSI INI SAYA PERSEMBAHKAN KEPADA :

ALLAH SWT YANG SELALU SAYA HARAPKAN RIDHO DAN
HIDAYAHNYA

ORANG TUA SAYA YANG SELALU MENDUKUNG DAN MENDOAKAN
SAYA DALAM SEGALA HAL

SELULUH ANGGOTA KELUARGA SAYA YANG SELALU MEMBERIKAN
SEMANGAT DAN MENYAYANGI SAYA

DAN SELURUH TEMAN-TEMAN INFORMATIKA YANG SELALU
MEMBANTU DAN MEMBERIKAN DUKUNGAN KEPADA SAYA DALAM
MENYELESESAIKAN SKRIPSIN INI



ABSTRAK

Hafidz Haikal. 2023. *Pengecekan Kesamaan Makna Kalimat Dari Referensi Jurnal Dengan Metode Cosine Similarity*, Mahasiswa. Program Studi Informatika, Fakultas Teknologi Informasi Universitas Merdeka Pasuruan.

Pembimbing : (I) Mohammad Zoqi Sarwani, S.Pd., M.Kom., (II) Dian Ahkam Sani, S.Kom., M.Kom.

Kata Kunci : Kalimat Referensi, Kesamaan Makna, *FastText*, *Cosine Similarity*

Penelitian ini bertujuan melakukan pengecekan kalimat referensi kesamaan makna, sistem yang mampu melakukan pengecekan kesamaan makna antara kalimat menjadi sangat penting untuk meningkatkan kualitas dan efektivitas. Untuk mencapai tujuan penelitian, peneliti menggunakan algoritma *FastText* yang dikembangkan oleh Facebook untuk melakukan pemodelan kalimat. Algoritma ini mampu memperhitungkan konteks kata dalam suatu kalimat dan menghasilkan representasi vektor untuk setiap kata dalam kalimat. Kemudian, representasi vektor tersebut digunakan untuk menghitung representasi vektor untuk keseluruhan kalimat. Setelah mendapatkan representasi vektor untuk kalimat referensi dan kalimat uji, peneliti menggunakan metode *cosine similarity* untuk menghitung kedekatan antara kedua vektor tersebut. Metode *cosine similarity* digunakan karena kemampuannya mengenali makna yang sama meskipun kalimat memiliki perbedaan sintaksis atau penggunaan kata yang sedikit berbeda. Peneliti menguji sistem yang dibuat pada sebuah dataset yang terdiri dari 273 data kalimat referensi dan kalimat uji. Akurasi yang diperoleh dengan rata-rata 71,64%, yang menunjukkan bahwa sistem yang dikerjakan mampu menemukan kesamaan makna antara kalimat dengan cukup baik. Hasil penelitian ini dapat menjadi dasar untuk pengembangan sistem pengecekan kesamaan yang lebih maju dan efisien di masa mendatang.

ABSTRACT

Hafidz Haikal, 2023. *Checking Sentence Similarity of Journal References Using Cosine Similarity Method*, Thesis. Informatics Study Program, Faculty of Information Technology, Merdeka University of Pasuruan

Adviser : (I) Mohammad Zoqi Sarwani, S.Pd., M.Kom., (II) Dian Ahkam Sani, S.Kom., M.Kom.

Keyword : *Reference Sentence, Semantic Similarity, FastText, Cosine Similarity*

This research aims to perform a check on the similarity of sentence meanings. A system capable of checking the similarity of meanings between sentences is crucial for improving quality and effectiveness. To achieve the research objective, the researcher used the FastText algorithm developed by Facebook to model sentences. This algorithm can consider the context of words in a sentence and generate vector representations for each word in the sentence. Then, these vector representations are used to calculate the vector representation for the entire sentence. After obtaining the vector representations for the reference and test sentences, the researcher employed the cosine similarity method to measure the proximity between the two vectors. The cosine similarity method is used because of its ability to recognize the same meaning even when sentences have slight syntactic differences or variations in word usage. The researcher tested the developed system on a dataset consisting of 273 pairs of reference and test sentences. The accuracy obtained was an average of 71.64%, indicating that the implemented system can effectively identify the similarity of sentence meanings. The results of this research can serve as a foundation for the development of more advanced and efficient similarity checking systems in the future.

KATA PENGANTAR

Puji syukur atas kehadiran Allah SWT, yang telah memberikan rahmat dan hidaya-Nya kepada kita sekalian, khususnya kepada penulis, sehingga skripsi dengan judul " *Pengecekan Kesamaan Makna Kalimat Dari Referensi Jurnal Dengan Metode Cosine Similarity* " dapat terselesaikan dengan baik.

Di dalam penyelesaiannya penulis banyak sekali dibantu oleh beberapa pihak, oleh karenanya pada kesempatan ini, penulis mengucapkan rasa terimakasih kepada :

1. Allah SWT yang selalu memberikan kekuatan kepada penulis, khususnya pada saat penyelesaian skripsi ini yang walaupun diri ini banyak sekali berbuat salah dan dosa kepada-MU.
2. Ayah dan Ibu tercinta yang selalu ikhlas dan penuh dengan kesabaran membesarkan dan mendidik penulis selama ini. Semoga rahmat Allah SWT selalu menyertaimu.
3. Bapak Muhammad Misdram, S.Kom., M.Kom., selaku Dekan Fakultas Teknologi Informasi Universitas Merdeka Pasuruan.
4. Bapak Dian Ahkam Sani, S.Kom., M.Kom., selaku Ketua Program Studi Informatika Universitas Merdeka Pasuruan sekaligus selaku Dosen Pembimbing II yang telah meluangkan waktu untuk membimbing penulis selama pembuatan skripsi ini.
5. Bapak M. Zoqi Sarwani, S.Pd., M.Kom., selaku Dosen Pembimbing I yang telah meluangkan waktu untuk membimbing penulis selama pembuatan skripsi ini.
6. Teman - Teman Informatika angkatan tahun 2019 yang selalu memberikan semangat dan bantuan selama menyelesaikan skripsi ini.
7. Serta semua pihak yang secara langsung maupun tidak langsung telah membantu penulis dalam menyelesaikan skripsi ini.

Penulis dalam menulis skripsi ini sudah berusaha semaksimal mungkin, namun penulis juga menyadari juga bahwa skripsi ini masih banyak kekurangan dan jauh dari kata sempurna. Oleh sebab itu penulis sangat mengharapkan kritik dan saran dari pembaca yang dapat bermanfaat bagi perbaikan pada masa mendatang.

Penulis berharap semoga skripsi ini dapat bermanfaat bagi pihak-pihak yang berkepentingan untuk menambah wawasan dan pengetahuan dan juga bagi pembaca.

Pasuruan, 04 Juli 2023

Penulis

Hafidz Haikal



DAFTAR ISI

	Halaman
HALAMAN COVER	i
HALAMAN PERSETUJUAN DOSEN PENDAMPING.....	ii
HALAMAN PENGESAHAN 1	iii
HALAMAN PENGESAHAN 2	iv
HALAMAN PERTANYAAN KEASLIAN SKRIPSI	v
HALAMAN MOTTO	vi
HALAMAN PERSEMBAHAN	vii
ABSTRAK	viii
<i>ABSTRACT</i>	ix
KATA PENGANTAR	x
DAFTAR ISI	xii
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	
2.1 Penelitian Terkait	7
2.2 Landasan Teori	10
2.2.1 Kalimat	10
2.2.2 <i>Text Mining</i>	12
2.2.3 <i>Natural Language Processing</i>	13
2.2.4 <i>Word Embedding</i>	13
2.2.5 Sistem Temu Kembali (<i>Information Retrival</i>)	14
2.2.6 <i>Preprocessing Text</i>	15
2.2.7 Tahap <i>Preprocessing Text</i>	15
2.2.8 <i>Library Scikit - Learn (Sklearn)</i>	16
2.2.9 <i>Library Pandas</i>	17

2.2.10 <i>FastText</i>	18
2.2.11 <i>Vector Size</i>	19
2.2.12 <i>Cosine Similarity</i>	19
BAB III METODOLOGI PENELITIAN	
3.1 Alur Penelitian	22
3.2 Perancangan Sistem	25
3.3 <i>Dataset</i>	26
3.4 <i>Text Preprocessing</i>	27
3.4.1 <i>Case Folding</i>	28
3.4.2 <i>Tokenization</i>	29
3.5 <i>FastText Modeling</i>	29
3.5.1 Pembuatan Modeling	30
3.5.2 Struktur Ulasan	31
3.6 <i>Cosine Similariy</i>	32
3.7 Kebutuhan Sistem	32
3.7.1 Perangkat Keras	32
3.7.2 Perangkat Lunak	33
BAB VI HASIL DAN PEMBAHASAN	
4.1 Implementasi Sistem	34
4.2 <i>Download</i> Wikipedia Bahasa Indonesia	35
4.3 <i>Library</i>	35
4.4 Pemodelan <i>FastText</i>	36
4.5 <i>Input Data</i>	37
4.6 <i>Input Kalimat</i>	38
4.7 <i>Preprocessing Text</i>	39
4.8 <i>Vektor</i>	40
4.9 <i>Dataset</i>	40
4.10 <i>Cosine Similariy</i>	41
BAB V PENUTUP	
5.1 Kesimpulan	42
5.2 Saran	42
DAFTAR PUSTAKA	44
LAMPIRAN	47

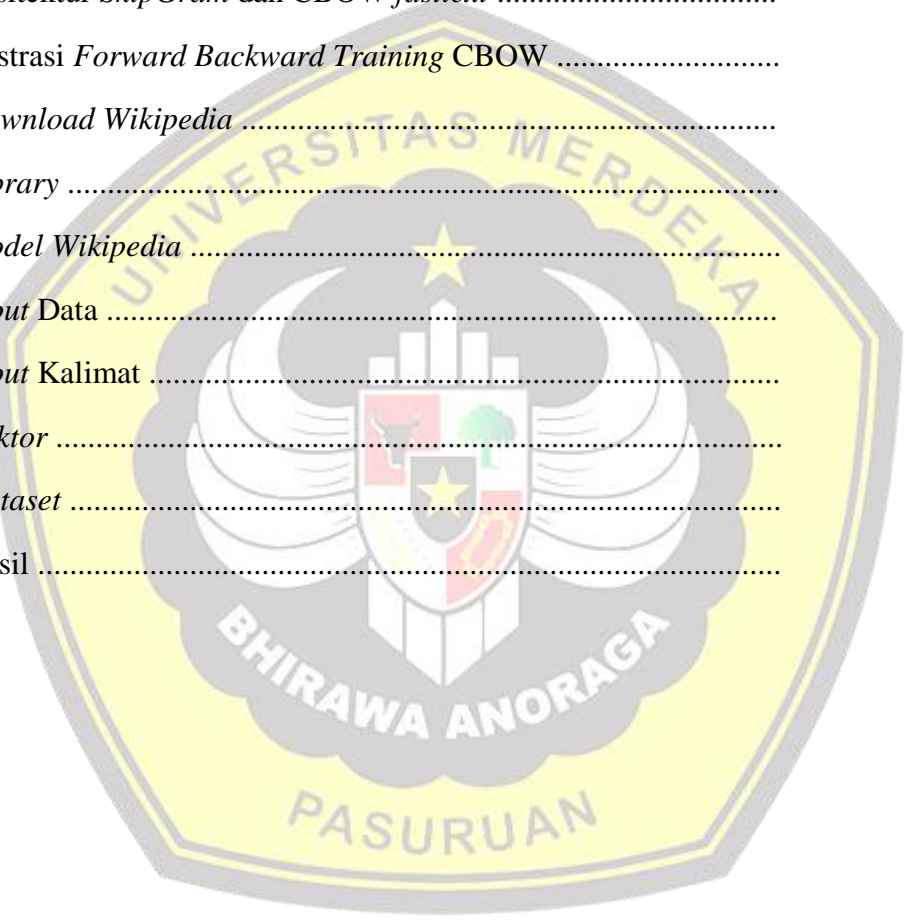
DAFTAR TABEL

Tabel	Halaman
3.1 Hasil Proses <i>Case Folding</i>	28
3.2 Hasil Proses <i>Tokenization</i>	29
4.1 Hasil <i>Case Folding</i>	39
4.2 Hasil <i>Tokenization</i>	39



DAFTAR GAMBAR

	Halaman
3.1 Alur Penelitian	22
3.2 Perancangan Sistem	25
3.3 <i>Text Preprocessing</i>	28
3.4 Arsitektur <i>SkipGram</i> dan <i>CBOW fasttext</i>	31
3.5 Ilustrasi <i>Forward Backward Training CBOW</i>	31
4.1 <i>Download Wikipedia</i>	35
4.2 <i>Library</i>	35
4.3 <i>Model Wikipedia</i>	36
4.4 <i>Input Data</i>	37
4.5 <i>Input Kalimat</i>	38
4.6 <i>Vektor</i>	40
4.7 <i>Dataset</i>	40
4.8 Hasil	41



DAFTAR LAMPIRAN

	Halaman
1. Lampiran 1 <i>Source Code Program</i>	48
2. Lampiran 2 Link Program dan Link Dataset	53
3. Lampiran 3 Daftar Riwayat Hidup	54
4. Lampiran 3 Surat Keterangan Cek Plagiasi	55

