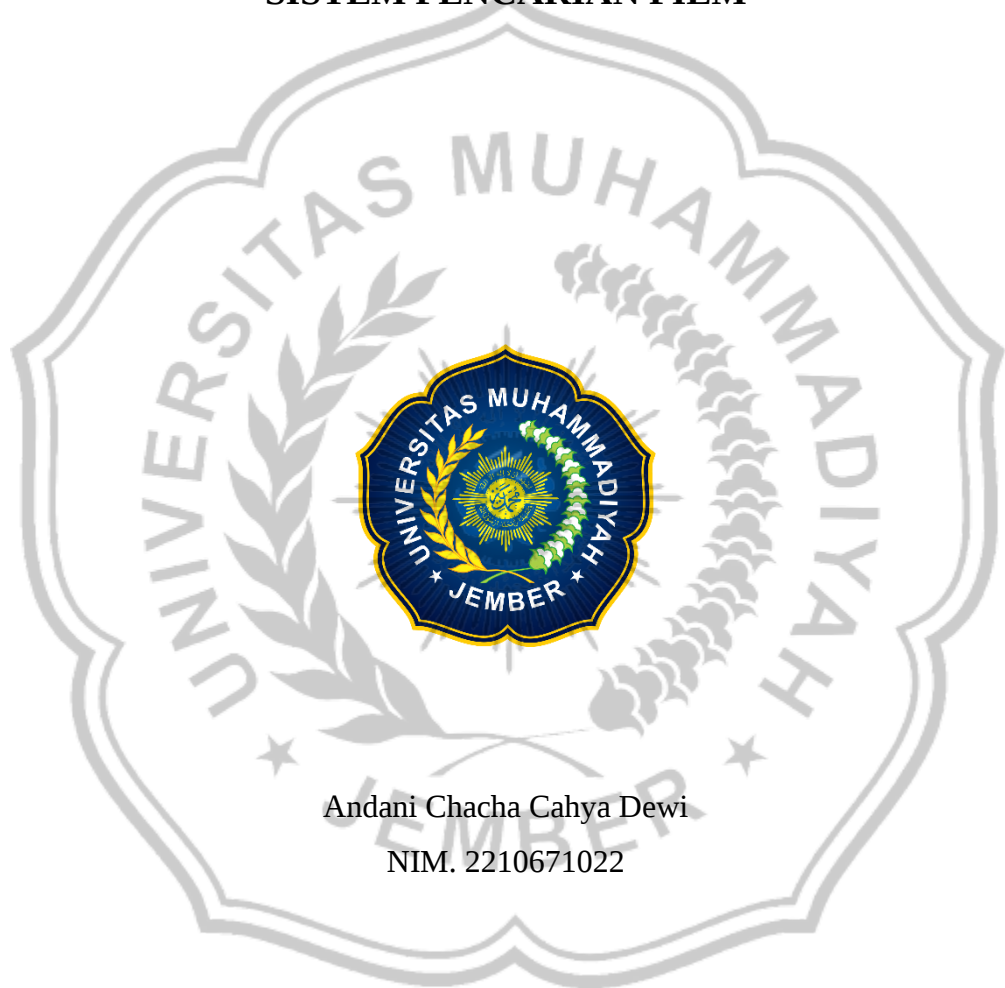


LAPORAN TUGAS AKHIR

**IMPLEMENTASI *CONTENT-BASED FILTERING*
MENGUNAKAN *WORD EMBEDDINGS* PADA
SISTEM PENCARIAN FILM**



Andani Chacha Cahya Dewi
NIM. 2210671022

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER
2026**

LAPORAN TUGAS AKHIR

IMPLEMENTASI *CONTENT-BASED FILTERING* MENGUNAKAN *WORD EMBEDDINGS* PADA SISTEM PENCARIAN FILM

Disusun sebagai salah satu syarat untuk kelulusan
Strata Satu (S-1) Program Studi Sistem Informasi Fakultas Teknik
Universitas Muhammadiyah Jember



Oleh:
Andani Chacha Cahya Dewi
NIM. 2210671022

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER
2026**

LEMBAR PERSETUJUAN SIDANG TUGAS AKHIR

Yang bertanda tangan dibawah ini:

Nama Dosen Pembimbing I : Deni Arifianto., M.Kom.
NPK/NIDN : 0718068103
Nama Dosen Pembimbing II : Nanda Kurnia Wardati., S.Kom., M.Kom
NPK/NIDN : 0715109401

Sebagai Dosen Pembimbing Tugas Akhir (TA), pada Mahasiswa:

Nama : Andani Chacha Cahya Dewi
NIM : 2210671022
Program Studi : Sistem Informasi

Bersama ini menyatakan:

Menyetujui mahasiswa tersebut diatas untuk maju dalam Sidang Tugas Akhir dengan judul:

IMPLEMENTASI *CONTENT-BASED FILTERING* MENGGUNAKAN WORD EMBEDDING PADA SISTEM PENCARIAN FILM

Jember, 31 Juli 2026

Pembimbing I


Deni Arifianto., M.Kom.
NIDN. 0718068103

Pembimbing II


Nanda Kurnia Wardati., S.Kom., M.Kom.
NIDN. 0715109401

Mengetahui
Ketua Program Studi Sistem Informasi


Deni Arifianto., M.Kom.
NIDN. 0718068103

LEMBAR PENGESAHAN

IMPLEMENTASI CONTENT-BASED FILTERING MENGGUNAKAN WORD EMBEDDINGS PADA SISTEM Pencarian FILM

Oleh:

Andani Chacha Cahya Dewi

2210671022

Telah mempertanggung jawabkan Laporan Tugas Akhirnya pada sidang Tugas Akhir tanggal 31 Juli 2026 sebagai salah satu syarat kelulusan dan mendapat gelar

Sarjana Komputer (S.Kom)

di

Universitas Muhammadiyah Jember

Disetujui Oleh:

Penguji I

Ulya Anisatur Rosyidah., M.Kom.

NIDN. 0710037903

Penguji II

Wiwik Suharso., M.Kom

NIDN. 0006097601

Mengesahkan,
Dekan Fakultas Teknik,

Prof. Dr. Ir. Muhtar, S.T., M.T., IPM

NIDN. 0010067301

Pembimbing I

Deni Arifianto., M.Kom

NIDN. 0718068103

Pembimbing II

Nanda Kurnia Wardati., S.Kom., M.Kom

NIDN. 0715109401

Mengetahui,
Ketua Program Studi Sistem Informasi

Deni Arifianto., M.Kom

NIDN. 0718068103

PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama : Andani Chacha Cahya Dewi
NIM : 2210671022
Program Studi : Sistem Informasi

Menyatakan bahwa Tugas Akhir yang berjudul **“IMPLEMENTASI *CONTENT-BASED FILTERING* MENGGUNAKAN WORD EMBEDDING PADA SISTEM Pencarian Film”** adalah benar benar hasil karya saya sendiri (kecuali kutipan yang telah saya sebutkan sebelumnya) dan belum pernah diajukan pada institusi manapun.

Demikian pernyataan ini saya buat dengan sebenarnya tanpa ada paksaan dan tekanan dari pihak manapun. Saya siap bertanggung jawab dan bersedia menerima sanksi apabila dikemudian hari pernyataan ini tidak benar.

Jember, 31 Juli 2026

Yang membuat pernyataan



Andani Chacha Cahya Dewi
NIM. 2210671022

PRAKATA

Puji syukur kehadiran Allah SWT atas segala rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul “Implementasi *Content-Based Filtering* Menggunakan *Word Embeddings* Pada Sistem Pencarian Film” sebagai salah satu syarat untuk memperoleh gelar Sarjana Sistem Informasi di Universitas Muhammadiyah Jember. Dalam proses penyusunan Tugas Akhir ini, penulis menyadari bahwa perjalanan yang dilalui tidak selalu mudah. Banyak hal yang harus dipelajari, diperbaiki, dilewati, dan diselesaikan satu per satu. Namun, berkat doa, dukungan, bantuan, bimbingan, serta kehadiran berbagai pihak, akhirnya Tugas Akhir ini dapat diselesaikan dengan baik. Oleh karena itu, dengan segala kerendahan hati, penulis menyampaikan terima kasih kepada:

1. Allah SWT yang telah memberikan kesehatan, kekuatan, kesabaran, kemudahan, serta kelancaran kepada penulis dalam melewati setiap proses penyusunan Tugas Akhir ini.
2. Kepada orang tua saya tercinta, Mama Ida Nurani., S.H. Terima kasih yang tak terhingga atas setiap doa yang tidak pernah putus, kasih sayang yang tidak pernah berkurang, serta pengorbanan yang mungkin tidak akan pernah mampu penulis balas dengan apa pun. Terima kasih telah menjadi tempat pulang terbaik, sumber kekuatan terbesar, dan alasan penulis untuk terus bertahan dan menyelesaikan apa yang telah dimulai. Di balik setiap pencapaian penulis, selalu ada doa, perjuangan, kerja keras, dan keikhlasan Mama. Semoga Allah SWT senantiasa memberikan kesehatan, kebahagiaan, umur yang panjang, serta keberkahan dalam setiap langkah Mama.
3. Bapak Prof. Dr. Ir. Muhtar, S.T., M.T., IPM., selaku Dekan Fakultas Teknik Universitas Muhammadiyah Jember.
4. Bapak Deni Arifianto, M.Kom., selaku Ketua Program Studi Sistem Informasi Fakultas Teknik Universitas Muhammadiyah Jember.
5. Bapak Deni Arifianto, M.Kom., selaku Dosen Pembimbing I dan Ibu Nanda Kurnia Wardati, S.Kom., M.Kom, selaku Dosen Pembimbing II yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan,

arahan, motivasi, serta masukan yang sangat berarti selama proses penyusunan Tugas Akhir ini hingga dapat terselesaikan dengan baik. Penulis sangat berterima kasih sebesar-besarnya kepada beliau, karena telah membimbing, megajar, memberikan pengetahuan serta ilmu yang bermanfaat, dan tentunya membuat penulis yakin dan optimis terhadap apa yang penulis kerjakan.

6. Ibu Ulya Anisatur Rosyidah, M.Kom., selaku Dosen Penguji I dan Bapak Wiwik Suharso., M.Kom selaku Dosen Penguji II dalam penyusunan skripsi penulis sendiri. Tentunya, penulis sangat berterima kasih sebesar-besarnya kepada beliau, karena telah membimbing, megajar, memberikan pengetahuan serta ilmu yang bermanfaat, dan tentunya membuat penulis yakin dan optimis terhadap apa yang penulis kerjakan.
7. Ibu Dr. Eka Nova Ali Vardani., M.Pd dan Bapak Agus Milu Susetyo., M.Pd selaku Validator Pakar pada penelitian ini. Penulis juga mengucapkan terima kasih yang sebesar-besarnya, karena kesediaan waktu, tenaga, serta ketelitiannya dalam memvalidasi data pengujian *ground truth*, sehingga evaluasi sistem pencarian film ini dapat diselesaikan dengan baik dan objektif.
8. Seluruh Bapak dan Ibu Dosen Program Studi Sistem Informasi Fakultas Teknik Universitas Muhammadiyah Jember yang telah memberikan ilmu, pengalaman, serta bimbingan kepada penulis selama menempuh pendidikan di bangku perkuliahan.
9. Kepada seseorang dengan NIM 2210671032, terima kasih karena telah menjadi salah satu bagian baik dalam perjalanan penulis hingga sampai pada titik ini. Terima kasih atas waktu, perhatian, kesabaran, dukungan, serta hal-hal sederhana yang mungkin terlihat kecil, tetapi memiliki arti bagi penulis. Terima kasih telah hadir dan menemani penulis di tengah proses penyelesaian Tugas Akhir ini, mendengarkan berbagai keluhan, memberikan semangat ketika penulis mulai lelah, serta menjadi salah satu tempat berbagi cerita di tengah hari-hari yang tidak selalu mudah. Mungkin waktu yang dilalui bersama belum terlalu panjang, tetapi kehadiranmu telah memberikan warna tersendiri dalam perjalanan ini. Semoga segala hal baik yang sedang kita

jalani dapat terus tumbuh dan membawa kita pada hal-hal baik lainnya di kemudian hari.

10. Kepada Irawati Kusumadewi dan Retno Anjani, terima kasih telah menjadi bagian dari perjalanan penulis sejak awal perkuliahan hingga sampai pada titik ini. Terima kasih atas kebersamaan, doa, dukungan, perhatian, serta segala cerita yang telah kita lalui bersama sejak masa-masa awal menjadi mahasiswa hingga akhirnya sampai di penghujung perjalanan perkuliahan. Terima kasih telah hadir di berbagai keadaan, baik ketika hari-hari terasa menyenangkan maupun ketika semuanya terasa melelahkan. Terima kasih untuk setiap tawa, keluh kesah, cerita, bantuan, dan semangat yang telah dibagikan selama ini. Banyak hal yang mungkin terlihat sederhana, tetapi justru menjadi bagian dari kenangan yang akan selalu penulis ingat. Semoga setelah perjalanan ini, kita tetap memiliki jalan masing-masing yang penuh dengan hal-hal baik, dan semoga persahabatan serta cerita yang telah terjalin sejak awal perkuliahan tetap menjadi bagian indah yang tidak terlupakan.
11. Kepada Fidyah Masyitara dan Cantika Putri Anastasya, terima kasih telah menjadi salah satu bagian baik yang penulis temui dalam perjalanan perkuliahan ini. Meskipun kebersamaan kita berawal dari kegiatan KKN, penulis bersyukur dapat mengenal dua pribadi yang begitu baik, tulus, dan menyenangkan. Terima kasih atas segala perhatian, bantuan, kepedulian, dukungan, serta kebaikan yang telah diberikan selama kita bersama. Terima kasih untuk setiap cerita, tawa, dan momen sederhana yang membuat masa KKN menjadi salah satu bagian yang begitu berkesan bagi penulis. Bahkan setelah KKN selesai, kebaikan dan perhatian yang kalian berikan tetap menjadi hal yang penulis syukuri hingga saat ini. Semoga segala kebaikan yang telah kalian berikan kembali kepada kalian dalam bentuk kebahagiaan, keberkahan, dan hal-hal baik yang jauh lebih besar. Terima kasih sudah menjadi bagian dari cerita yang mungkin singkat, tetapi meninggalkan kesan yang begitu berarti.
12. Penulis juga ingin menyampaikan rasa terima kasih kepada teman-teman sekelas dan juga angkatan Sistem Informasi 22 yang tidak bisa saya sebutkan satu satu namanya, terimakasih banyak yang telah menjadi keluarga kedua

selama masa perkuliahan. Terima kasih atas segala tawa, diskusi panjang, kekompakan, serta keluh kesah yang kita lalui bersama dari semester pertama hingga masa pengerjaan skripsi ini. Pengalaman berjuang bersama kalian akan selalu menjadi kenangan terbaik.

13. *Last but not least*, yang paling utama, peluk hangat dan apresiasi tertinggi saya persembahkan untuk diri saya sendiri, Andani Chacha Cahya Dewi. Terima kasih sudah memilih untuk bertahan dan tidak menyerah. Mengingat kembali masa lalu, tidak pernah terbayangkan sedikit pun akan berhasil menyanggah gelar dari program studi Sistem Informasi ini. Berangkat dari seorang anak yang sama sekali buta akan bidang teknologi, yang harus menelan kecewa berkali-kali karena penolakan PTN, dan yang sempat patah hati karena gagal menggapai mimpi di jurusan Farmasi maupun Kedokteran. Namun, jalan Tuhan ternyata jauh lebih indah. Dari rentetan ketidaktahuan dan rasa salah arah itu, kamu berjuang keras hingga akhirnya berhasil membuktikan kepada diri sendiri dan dunia bahwa kamu bisa melampaui batasmu. Terima kasih sudah menangis tapi tetap mengetik baris demi baris skripsi ini. Selamat merayakan keberhasilanmu, dan bersiaplah menghadapi kejutan di babak kehidupan selanjutnya. Kamu hebat, dan kamu berhasil membuktikannya.

Jember, 31 Juli 2026



Andani Chacha Cahya Dewi
2210671022

MOTTO

" Allah tidak mengatakan hidup ini mudah. Tetapi Allah berjanji,bahwa
sesungguhnya bersama kesulitan ada kemudahan "

(Q.S Al-insyirah: 5-6)

“Orang lain ga akan paham struggle dan masa sulit nya kita yang mereka ingin
tabu hanya bagian success stories. Berjuanglah untuk diri sendiri walaupun tidak
ada tepuk tangan. Kelak diri kita dimasa depan akan sangat banga dengan apa
yang kita perjuangkan hari ini”

*"Life can be heavy,especially if you try to carry it all at once. Part of growing up
and moving into new chapters of your life is about catch and release"*

(Taylor Swift)

“Semua jatuh bangun mu hal yang biasa,angan dan pertanyaan waktu yang
menjawab nya,berikan tenggat waktu bersedihlah secukupnya,rayakan perasaan
mu sebagai manusia”

(Baskara Putra-Hindia)

" Apapun yang sudah terjadi dalam hidupmu, jangan katakan "seandai nya" tapi
katakan " Qadarullah " karna semua yang terjadi adalah takdir dan takdir Allah itu
selalu baik,karna allah itu maha baik"

(ustadz Hannan Attaki)

“Terkadang, kegagalan di satu pintu adalah cara Tuhan untuk mengarahkan kita
ke pintu lain tempat kita benar-benar bisa bersinar. Mekarlah di tempat yang tidak
pernah kamu rencanakan.”

(Achaaa)

DAFTAR ISI

LEMBAR PERSETUJUAN SIDANG TUGAS AKHIR	ii
LEMBAR PERNYATAAN	iii
PERNYATAAN.....	iv
PRAKATA.....	v
MOTTO	ix
ABSTRAK	x
DAFTAR ISI.....	xi
DAFTAR GAMBAR	xiv
DAFTAR TABEL.....	xv
DAFTAR LAMPIRAN.....	xvi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah.....	6
BAB 2 TINJAUAN PUSTAKA	8
2.1 Sistem Temu Kembali	8
2.2 <i>Content-Based Filtering</i>	9
2.3 Teknik <i>Web Scraping</i>	10
2.4 <i>Text Mining</i> dan <i>Natural Language Processing</i> (NLP)	11
2.5 <i>Word Embeddings</i> (Word2Vec).....	12
2.6 <i>Cosine Similarity</i>	17
2.7 <i>Inter-Annotator Agreement</i> (IAA).....	19
2.8 Metode Validasi Holdout (<i>Holdout Method</i>)	20
2.9 Evaluasi Kinerja Dan Relevansi Sistem.....	21
2.9.1 <i>Precision@K</i>	23
2.9.2 <i>Recall@K</i>	24
2.9.3 <i>Mean Reciprocal Rank</i> (MRR)	25
2.10 Penelitian Terdahulu.....	26
BAB 3 METODOLOGI PENELITIAN.....	30
3.1 Tahapan Penelitian	30
3.2 Identifikasi Masalah dan Kebutuhan Sistem	31

3.2.1	Identifikasi Masalah	31
3.2.2	Identifikasi Kebutuhan Sistem	31
3.3	Pengumpulan Data (<i>Web Scraping</i>)	33
3.4	Pembersihan Data Awal (<i>Data Cleaning</i>)	36
3.5	Pra-Pemrosesan Data (<i>Text Preprocessing</i>)	37
3.6	Pembagian Data (<i>Holdout</i>)	39
3.7	Pemodelan Representasi Teks (<i>Word Embeddings Word2Vec</i>)	40
3.8	Penerjemahan Kueri Pengguna (<i>Deep-Translator</i>)	43
3.9	Pengujian Kemiripan Menggunakan <i>Cosine Similarity</i>	45
3.10	Visualisasi Antarmuka Pengguna	45
3.11	Evaluasi Kinerja Algoritma	47
3.11.1	Validasi <i>Ground Truth</i> Pakar (<i>Inter-Annotator Agreement</i>)	48
3.11.2	Metrik Evaluasi Kinerja (<i>Precision@K</i> , <i>Recall@K</i> , dan <i>MRR</i>)	49
BAB 4 HASIL DAN PEMBAHASAN		52
4.1	Pengumpulan Data	52
4.2	Pembersihan Data Awal (<i>Data Cleaning</i>)	53
4.3	Hasil Pra-Pemrosesan Data (<i>Text-Preprocessing</i>)	56
4.3.1	Hasil Penormalan Huruf Kapital (<i>Case Folding</i>)	56
4.3.2	Hasil Pembersihan Teks (<i>Clean Text</i>)	57
4.3.3	Hasil Pemotongan Kata (<i>Tokenizing</i>)	59
4.3.4	Hasil Pembuangan Kata Umum (<i>Stopword Removal</i>)	60
4.3.5	Hasil Lematisasi (<i>Lemmatization</i>)	62
4.3.6	Hasil Pembentukan Teks Terproses (<i>Processed Text</i>)	63
4.4	Hasil Pemodelan Representasi Teks (<i>Word Embeddings Word2Vec</i>) ...	64
4.4.1	Hasil Pembentukan Vektor Kata (<i>Word Embeddings</i>)	65
4.4.2	Hasil Pembentukan Vektor Dokumen (<i>Movie Vectors</i>)	66
4.5	Hasil Perhitungan Kemiripan Menggunakan <i>Cosine Similarity</i>	68
4.6	Implementasi Antarmuka Pengguna	71
4.6.1	Antarmuka Halaman Utama Pencarian (<i>Search Page</i>)	71
4.6.2	Antarmuka Halaman Hasil Pencarian (<i>Search Results Page</i>)	72
4.7	Hasil Evaluasi Kinerja Algoritma	73
4.7.1	Hasil Validasi <i>Ground Truth</i> Pakar (<i>Inter-Annotator Agreement</i>) .	74
4.7.2	Hasil Metrik Evaluasi Kinerja (<i>Precision@K</i> , <i>Recall@K</i> , dan <i>MRR</i>) 76	
4.7.3	Hasil Scenario Pengujian Berdasarkan Variasi Panjang Kata Kunci (<i>Query Length Analysis</i>)	79

BAB 5 PENUTUP	81
5.1 Kesimpulan.....	81
5.2 Saran	81
DAFTAR PUSTAKA	83
LAMPIRAN.....	88
BIODATA PENULIS	94



DAFTAR GAMBAR

Gambar 4.1 Distribusi 15 Genre Terpopuler pada Pangkalan Data.....	54
Gambar 4.2 Antarmuka Halaman Utama Pencarian (Search Page).....	72
Gambar 4.3 Antarmuka Halaman Hasil Pencarian (Search Results Page)	73
Gambar 4.4 Visualisasi Fluktuasi Kinerja Sistem terhadap 25 Kueri Pengujian..	79



DAFTAR TABEL

Tabel 2.1 Matriks Dasar Klasifikasi (Confusion Matrix)	22
Tabel 2.2 Penelitian Terdahulu	26
Tabel 3.1 Spesifikasi Perangkat Keras dan Perangkat Lunak.....	32
Tabel 3.2 Struktur Atribut Dataset Film	34
Tabel 3.3 Contoh Dataset Film IMDb.....	35
Tabel 3.4 Ilustrasi Hasil Pra-pemrosesan Teks Sinopsis	39
Tabel 3.5 Rincian Pembagian Dataset (Holdout Split)	40
Tabel 3.6 Konfigurasi Parameter Pelatihan Word2Vec.....	41
Tabel 3.7 Ilustrasi Representasi Vektor Kata (Word Embeddings).....	42
Tabel 3.8 Contoh Sajian Hasil Pencarian Film pada Antarmuka Pengguna.....	47
Tabel 3.9 Rincian Mekanisme Pengambilan Keputusan Kesepakatan (IAA)	48
Tabel 3.10 Tingkat Kesepakatan Cohen's Kappa.....	49
Tabel 4.1 Contoh Data Mentah Hasil Ekstraksi IMDb	52
Tabel 4.2 Cuplikan Baris Awal Pangkalan Data Bersih (Clean Dataset)	54
Tabel 4.3 Cuplikan Perbandingan Atribut Sinopsis Awal dan Luaran Case Folding	56
Tabel 4.4 Cuplikan Perbandingan Atribut Case Folding dan Luaran Clean Text	58
Tabel 4.5 Cuplikan Perbandingan Atribut Clean Text dan Luaran Tokenizing ...	59
Tabel 4.6 Cuplikan Perbandingan Atribut Tokenizing dan Luaran Stopword Removal	61
Tabel 4.7 Cuplikan Perbandingan Atribut Stopword Removal dan Luaran Lemmatization.....	62
Tabel 4.8 Cuplikan Perbandingan Atribut Lemmatized	64
Tabel 4.9 Cuplikan Matriks Vektor Kata (Word Embeddings)	66
Tabel 4.10 Cuplikan Matriks Vektor Dokumen (Movie Vectors)	68
Tabel 4.11 Hasil Perhitungan Cosine Similarity untuk.....	70
Tabel 4.12 Sampel Penentuan Ground Truth Berdasarkan Konsensus Pakar	76
Tabel 4.13 Simulasi Perhitungan Metrik Evaluasi per Kueri	77
Tabel 4.14 Rekapitulasi Akhir Capaian Metrik Evaluasi Kinerja Sistem.....	77
Tabel 4.15 Hasil Pencarian Top-3 Berdasarkan Variasi Panjang Kata Kunci.....	79

DAFTAR LAMPIRAN

Lampiran 1 Source Code Web Scraping IMDb	88
Lampiran 2 Source Code Data Cleaning.....	88
Lampiran 3 Source Code Text Preprocessing.....	88
Lampiran 4 Source Code Text Preprocessing.....	88
Lampiran 5 Source Code Movie Embeddings	88
Lampiran 6 Source Code Sistem Pencarian Film dengan Cosine Similarity.....	88
Lampiran 7 Source Code Inter-Annotator Agreement (IAA).....	88
Lampiran 8 Source Code dan Hasil Evaluasi Sistem (Precision@K, Recall@K, MRR)	88
Lampiran 9 Hasil Pengujian Sistem Pencarian.....	89
Lampiran 10 Surat Permohonan Izin Penelitian	90
Lampiran 11 Lembar Pernyataan Ahli Pakar (Agus Milu Susetyo., M.Pd)	91
Lampiran 12 Lembar Pernyataan Ahli Pakar (Dr. Eka Nova Ali Vardani., M.Pd)	92
Lampiran 13 Dokumentasi.....	93