

**KLASIFIKASI BERBASIS ONTOLOGI UNTUK DOKUMEN
TUGAS AKHIR DENGAN METODE
PEMBOBOTAN FASTTEXT**

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 pada
Jurusan Teknik Informatika



Oleh:

Dellin Irawan

NIM : 09021282126080

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA**

2024

LEMBAR PENGESAHAN SKRIPSI

**KLASIFIKASI BERBASIS ONTOLOGI UNTUK DOKUMEN
TUGAS AKHIR DENGAN METODE PEMBOBOTAN FASTTEXT**

Oleh:

Dellin Irawan

NIM: 09021282126080

Indralaya, 31 Desember 2024

Pembimbing I



**Desty Rodiah, S.Kom., M.T.
NIP: 198912212020122011**

**Mengetahui,
Ketua Jurusan Teknik Informatika**



**Hadipurnawan Satria, Ph.D.
NIP: 198004182020121001**

TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari Selasa tanggal 31 Desember 2024 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Dellin Irawan
NIM : 09021282126080
Judul : Klasifikasi Berbasis Ontologi untuk Dokumen Tugas Akhir dengan Metode Pembobotan FastText.

Dan dinyatakan **LULUS**

1. Ketua Penguji

Kanda Januar Miraswan, S.Kom., M.T.
NIP: 199001092019031012



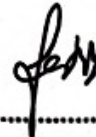
2. Penguji

Osvari Arsalan, S.Kom., M.T.
NIP: 198806282018031001



3. Pembimbing I

Desty Rodiah, S.Kom., M.T.
NIP: 198912212020122011



HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Dellin Irawan
NIM : 09021282126080
Program Studi : Teknik Informatika
Judul Skripsi : Klasifikasi Berbasis Ontologi untuk Dokumen Tugas Akhir
dengan Metode Pembobotan FastText

Hasil pengecekan *Software iThenticate/Turnitin* : 8%

Menyatakan bahwa laporan penelitian saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan penelitian ini, maka saya bersedia menerima sanksi akademik Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, 31 Desember 2024



Dellin Irawan
NIM 09021282126080

MOTTO DAN PERSEMBAHAN

“Yang membuat hidup ini menarik adalah kemungkinan untuk mewujudkan mimpi menjadi kenyataan.”

— Paulo Coelho, *The Alchemist*.

“Jangan pernah membaca karena ingin dianggap pintar; bacalah karena kamu mau membaca, dan dengan sendirinya kamu akan jadi pintar.”

— Ziggy Zezsyzazeoviennazabrizkie, *Jakarta Sebelum Pagi*.

Ku persembahkan karya tulis ini kepada:

- Allah Swt
- Nabi Muhammad saw
- Kedua orang tua dan saudari saya
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

ABSTRACT

The classification of undergraduate thesis documents is a crucial process for organizing and categorizing student research into appropriate fields of study. This research develops an ontology-based document classification system for final projects by integrating the FastText word embedding method to improve classification accuracy. The system utilizes BidangIlmuInformatika.owl ontology as a knowledge base to determine the hierarchy and relationships between concepts in the field of informatics. FastText method is implemented to generate word vector representations (word embeddings) capable of capturing semantic context from document text. The classification process is performed by calculating the similarity between document vectors weighted with FastText and concept representations in the ontology. The dataset consists of 160 informatics student undergraduate thesis documents divided into four main areas: Data Science and Pattern Recognition, Distributed Systems, Computer Graphics and Visualization, and Natural Language Processing. Evaluation results show that the proposed system achieves an accuracy of 25%. This is because the concepts within the Data Science and Pattern Recognition domain have a broad knowledge structure that spans nearly all areas of Informatics.

Keywords: Document Classification, Undergraduate Thesis, Ontology, FastText, Semantic Similarity.

ABSTRAK

Klasifikasi dokumen tugas akhir merupakan proses penting untuk mengorganisir dan mengelompokkan penelitian mahasiswa ke dalam bidang keilmuan yang sesuai. Penelitian ini mengembangkan sistem klasifikasi dokumen tugas akhir berbasis ontologi dengan mengintegrasikan metode pembobotan FastText untuk meningkatkan akurasi klasifikasi. Sistem yang dibangun menggunakan ontologi BidangIlmuInformatika.owl sebagai basis pengetahuan untuk menentukan hierarki dan relasi antar konsep dalam bidang informatika. Metode FastText diimplementasikan untuk menghasilkan representasi vektor kata (*word embedding*) yang mampu menangkap konteks semantik dari teks dokumen. Proses klasifikasi dilakukan dengan menghitung *similarity* antara vektor dokumen yang telah dibobotkan dengan FastText dan representasi konsep dalam ontologi. *Dataset* yang digunakan terdiri dari 160 dokumen tugas akhir mahasiswa informatika yang terbagi dalam empat bidang utama: *Data Science* dan Pengenalan Pola, Sistem Terdistribusi, Grafika Komputer dan Visualisasi, serta Pemrosesan Bahasa Alami. Hasil evaluasi menunjukkan bahwa sistem yang diusulkan mencapai akurasi sebesar 25%. Hal ini dikarenakan konsep pada domain *Data Science* dan Pengenalan Pola memiliki struktur pengetahuan yang luas hampir ke seluruh domain Informatika.

Kata Kunci: Klasifikasi Dokumen, Tugas Akhir, Ontologi, FastText, Kesamaan Semantik.

KATA PENGANTAR

Puji dan syukur kehadirat Allah SWT atas segala nikmat, rahmat dan karunia Nya, sehingga penulis dapat menyelesaikan Skripsi yang berjudul “Klasifikasi Berbasis Ontologi untuk Dokumen Tugas Akhir dengan Metode Pembobotan FastText”. Skripsi ini disusun untuk memenuhi salah satu persyaratan kelulusan program sarjana pada Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Dalam menyelesaikan skripsi ini, penulis menerima bantuan, bimbingan dan dukungan dari banyak pihak, baik secara langsung maupun tidak langsung. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada:

1. Allah SWT atas rahmat dan nikmat-Nya penulis dapat menyelesaikan skripsi ini dengan baik.
2. Ayah Dedi Irawan, Ibu Nelly Farlina, dan Adik penulis Yollin Irawan dan Fellin Irawan yang selalu mendoakan, mendukung, memberikan semangat, dan memotivasi untuk menyelesaikan skripsi ini.
3. Bapak Prof. Dr. Erwin, S.Si., M.Si. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Hadipurnawan Satria, Ph.D. selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Ibu Desty Rodiah, S.Kom., M.T. selaku dosen pembimbing skripsi yang telah membimbing dan memberikan arahan kepada penulis dalam pengerjaan skripsi.

6. Bapak Kanda Januar Miraswan, S.Kom., M.T. selaku dosen pembimbing akademik yang telah memberikan bimbingan selama proses kegiatan perkuliahan.
7. Seluruh Dosen, Admin, dan Staff Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
8. Teman – teman Project Penting Ahmad Azhari, Hanif Syahri Ramadhani, Risky Armansyah, Vanya Terra Ardani, dan Yolendri Anisyahfitri yang telah menemani, memberikan motivasi, dan dukungan kepada penulis.
9. Kak Alif Toriq Alkausar yang telah memberikan motivasi, saran, dan masukan kepada penulis.
10. Teman – teman Teknik Informatika angkatan 2021 yang telah menemani, memberikan motivasi, dan semangat selama pengerjaan skripsi ini.
11. Pihak – pihak lain yang tidak dapat penulis sebutkan satu-persatu.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih banyak sekali kekurangan dikarenakan kurangnya pengalaman dan pengetahuan penulis. Oleh karena itu penulis mengharapkan kritik dan saran yang membangun guna kemajuan penelitian selanjutnya. Semoga skripsi ini dapat bermanfaat. Terima kasih.

Indralaya, 31 Desember 2024

Penulis

Dellin Irawan

DAFTAR ISI

halaman

LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang	I-1
1.3 Rumusan Masalah	I-3
1.4 Tujuan Penelitian	I-4
1.5 Manfaat Penelitian	I-4
1.6 Batasan Masalah	I-5
1.7 Sistematika Penulisan	I-5
1.8 Kesimpulan	I-6
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1
2.2.1 Klasifikasi	II-1
2.2.2 Klasifikasi Dokumen	II-2
2.2.3 <i>Preprocessing</i>	II-2
2.2.4 <i>Query Processing</i>	II-5
2.2.5 FastText	II-7
2.2.6 Ontologi	II-9
2.2.7 <i>Cosine Similarity</i>	II-10
2.2.8 <i>Confusion Matrix</i>	II-11
2.2.9 <i>Rational Unified Process</i>	II-14
2.3 Penelitian Lain yang Relevan	II-16
2.4 Kesimpulan	II-18
BAB III METODE PENELITIAN	III-1
3.1 Pendahuluan	III-1
3.2 Pengumpulan Data	III-1

3.2.1.	Jenis dan Sumber Data	III-1
3.2.2.	Metode Pengumpulan Data	III-1
3.3	Tahapan Penelitian.....	III-2
3.3.1.	Mengumpulkan Data	III-2
3.3.2.	Menentukan Kerangka Kerja Penelitian	III-3
3.3.3.	Menentukan Kriteria Pengujian	III-7
3.3.4.	Menentukan Format Data Pengujian.....	III-7
3.3.5.	Menentukan Alat Bantu Penelitian.....	III-8
3.3.6.	Melakukan Pengujian Penelitian.....	III-9
3.3.7.	Melakukan Analisis dan Menarik Kesimpulan Penelitian	III-9
3.4	Kesimpulan.....	III-10
BAB IV PENGEMBANGAN PERANGKAT LUNAK		IV-1
4.1	Pendahuluan	IV-1
4.2	Fase Insepsi	IV-1
4.2.1	Pemodelan Bisnis	IV-1
4.2.2	Analisis Kebutuhan Sistem	IV-2
4.2.3	Analisis dan Desain.....	IV-4
4.2.3.1	Analisis Kebutuhan Perangkat Lunak	IV-5
4.2.3.2	Analisis Data	IV-5
4.2.3.3	Analisis Teks <i>Pre-processing</i>	IV-6
4.2.3.4	Analisis <i>Query Processing</i>	IV-10
4.2.3.5	Analisis Ontologi.....	IV-12
4.2.3.6	Analisis Pembobotan FastText	IV-14
4.2.3.7	Analisis <i>Cosine Similarity</i>	IV-16
4.2.3.8	Analisis Hasil Klasifikasi	IV-16
4.2.3.9	Analisis <i>Confusion Matrix</i>	IV-17
4.2.3.10	Desain Perangkat Lunak	IV-17
4.3	Fase Elaborasi.....	IV-24
4.3.1.	Pemodelan Bisnis	IV-25
4.3.1.1	Data yang Digunakan	IV-25
4.3.2.	Kebutuhan Sistem	IV-25
4.3.3.	Analisis dan Perancangan	IV-26
4.3.3.1	<i>Activity Diagram</i>	IV-26
4.3.3.2	<i>Sequence Diagram</i>	IV-30
4.3.3.3	Perancangan Antarmuka	IV-35
4.4	Fase Konstruksi	IV-37
4.4.1	Kebutuhan Sistem	IV-37
4.4.2	<i>Class Diagram</i>	IV-38
4.4.3	Implementasi	IV-39
4.4.2.1	Implementasi Kelas	IV-39
4.4.2.2	Implementasi Antarmuka	IV-42
4.5	Fase Transisi	IV-46
4.5.1	Pemodelan Bisnis	IV-46
4.5.2	Rencana Pengujian	IV-46
4.5.3	Implementasi	IV-48

4.6	Kesimpulan.....	IV-52
BAB V HASIL DAN ANALISIS PENELITIAN V-1		
5.1	Pendahuluan	V-1
5.2	Hasil Percobaan Penelitian	V-1
5.2.1	Konfigurasi Percobaan	V-1
5.2.1.1	Data Hasil Konfigurasi	V-2
5.3	Analisis Hasil Penelitian	V-12
5.4	Kesimpulan.....	V-14
BAB VI KESIMPULAN DAN SARAN VI-1		
6.1	Pendahuluan	VI-1
6.2	Kesimpulan.....	VI-1
6.3	Saran.....	VI-2
DAFTAR PUSTAKA		xvi
LAMPIRAN.....		xxi

DAFTAR TABEL

halaman

Tabel II-1. Contoh Case Folding	II-3
Tabel II-2. Contoh Tokenizing	II-3
Tabel II-3. Contoh Stopword Removal	II-4
Tabel II-4. Contoh Stemming.....	II-5
Tabel II-5. Sinonim WordNet.....	II-5
Tabel II-6. Hapus Duplikat.....	II-6
Tabel III-1 Tabel Confusion Matrix (Proboningrum & Sidauruk, 2021).....	II-11
Tabel III-2. Rancangan Confusion Matrix untuk 4 kelas.....	III-8
Tabel III-3. Hasil Pengujian.....	III-9
Tabel IV-1. Kebutuhan Fungsional	IV-3
Tabel IV-2. Kebutuhan Non-fungsional	IV-4
Tabel IV-3. Sampel Data	IV-6
Tabel IV-4. Teks Dataset	IV-7
Tabel IV-5. Hasil Case Folding	IV-8
Tabel IV-6. Hasil Tokenizing.....	IV-8
Tabel IV-7. Hasil Stopword Removal	IV-9
Tabel IV-8. Hasil Stemming.....	IV-10
Tabel IV-9. Sinonim di WordNet.....	IV-11
Tabel IV-10. Hasil Hapus Duplikat	IV-11
Tabel IV-11. Konsep Ontologi yang Sesuai dengan Teks Dataset	IV-12
Tabel IV-12. Vektor FastText Taks Dataset	IV-14
Tabel IV- 13. Vektor FastText Konsep Ontologi yang Sesuai Teks Dataset ...	IV-15
Tabel IV-14. Nilai <i>Cosine</i> Similarity.....	IV-16
Tabel IV-15. Hasil Klasifikasi	IV-17
Tabel IV-16. Definisi Aktor.....	IV-18
Tabel IV-17. Definisi Use Case	IV-19
Tabel IV-18. Skenario Use Case Upload Dataset.....	IV-20
Tabel IV-19. Skenario Use Case Klasifikasi	IV-22

Tabel IV-20. Pengujian dengan input teks dari pengguna.....	IV-23
Tabel IV-21. Implementasi Kelas	IV-40
Tabel IV-22. Rencana Pengujian Use Case Upload Dataset	IV-47
Tabel IV-23. Rencana Pengujian Use Case Klasifikasi.....	IV-47
Tabel IV-24. Rencana Pengujian Use Case Pengujian Input Teks dari Pengguna	IV-48
Tabel IV-25. Hasil pengujian Use Case Upload Dataset.....	IV-49
Tabel IV-26. Hasil Pengujian Use Case Klasifikasi	IV-50
Tabel IV-27. Hasil Pengujian Use Case Pengujian dengan Input Teks dari Pengguna.....	IV-51
Tabel V-1. Hasil konsep ontologi yang sesuai.....	V-2
Tabel V-2. Hasil perhitungan similarity.....	V-6
Tabel V-3. Confusion Matrix Data Uji	V-12
Tabel V-4. Hasil evaluasi confusion matrix seluruh kelas.....	V-13
Tabel V-5. Hasil evaluasi confusion matrix keseluruhan sistem	V-13

DAFTAR GAMBAR

	halaman
Gambar II-1. Perbedaan CBOW dan Skipgram(Liebenlito et al., 2024)	II-7
Gambar II-2. Arsitektur N-gram FastText (Nanda et al., 2024).....	II-8
Gambar II-3. Metode FastText (window size=5) (Choi & Lee, 2020)	II-9
Gambar II-4. Model Arsitektur RUP (IBM, 1998).....	II-14
Gambar III-1. Tahapan Penelitian	III-2
Gambar III-2. Kerangka Kerja Penelitian	III-3
Gambar III-3. Implementasi ontologi dengan Protégé (Rodiah, 2024).	III-5
Gambar III-4. Implementasi kelas dan subclass ontologi domain informatika	III-6
Gambar IV-1. Use Case Diagram	IV-18
Gambar IV-2. Activity Diagram Upload Dataset	IV-27
Gambar IV-3. Activity Diagram Melakukan Klasifikasi Berbasis Ontologi dengan Metode Pembobotan FastText	IV-28
Gambar IV-4. Activity Diagram Pengujian Input Teks dari Pengguna.	IV-29
Gambar IV-5. Sequence Diagram Upload Dataset.....	IV-30
Gambar IV-6. Sequence Diagram Melakukan Klasifikasi Berbasis Ontologi dengan Metode Pembobotan FastText	IV-33
Gambar IV-7. Sequence Diagram Pengujian Input Teks dari Pengguna.....	IV-34
Gambar IV-8. Rancangan Antarmuka Halaman Dataset Classification.....	IV-35
Gambar IV-9. Rancangan Antarmuka ketika Dataset telah di Upload.....	IV-35
Gambar IV-10. Rancangan Antarmuka setelah proses klasifikasi dataset	IV-36
Gambar IV-11. Rancangan Antarmuka Hamalam Manual Input	IV-36
Gambar IV-12. Rancangan Antarmuka Setelah Klasifikasi Input Pengguna.	IV-37
Gambar IV-13. Class Diagram	IV-39
Gambar IV-14. Antarmuka Halaman Dataset Classification.....	IV-43
Gambar IV-15. Antarmuka setelah upload dataset (1)	IV-43
Gambar IV-16. Antarmuka setelah upload dataset (2)	IV-44
Gambar IV-17. Antarmuka setelah proses klasifikasi (1).....	IV-44
Gambar IV-18. Antarmuka setelah proses klasifikasi (2).....	IV-44
Gambar IV-19. Antarmuka Halaman Manual Input	IV-45

Gambar IV-20. Antarmuka Setelah Input Teks	IV-45
Gambar IV-21. Antarmuka Setelah Proses Klasifikasi.....	IV-46

DAFTAR LAMPIRAN

1. Kode Program
2. Sample Data
3. Teks Dataset
4. Hasil Case Folding
5. Hasil Tokenizing
6. Hasil Stopword Removal
7. Hasil Stemming
8. Hasil Sinonim WordNet
9. Hasil Hapus Duplikat
10. Hasil Pemetaan Konsep Ontologi
11. Hasil Pembobotan FastText
12. Hasil Konsep Ontologi yang Sesuai

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini diuraikan gagasan yang menjadi latar belakang penulisan skripsi. Gagasan tersebut meliputi latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, sistematika penulisan, dan kesimpulan.

1.2 Latar Belakang

Pertumbuhan pesat penelitian akademik di Perguruan Tinggi dan Universitas menuntut cara yang efektif dan tepat untuk mengklasifikasikan Tugas Akhir yang telah selesai. Meningkatnya jumlah makalah penelitian, Tugas Akhir, dan disertasi di berbagai disiplin ilmu setiap tahun telah menjadikan tugas mengategorikan upaya intelektual ini menjadi suatu proses yang rumit, berkepanjangan, dan mahal (Montes et al., 2024). Salah satunya pada penelitian bidang Informatika. Penelitian bidang Informatika yang kompleks seringkali melibatkan multidisiplin ilmu. Semakin banyak subdisiplin yang terintegrasi, akan semakin sulit untuk menempatkan penelitian tersebut dalam satu kategori tunggal (Rodiah, 2024). Pada jurusan Teknik Informatika Universitas Sriwijaya, dosen-dosen memiliki bidang keilmuan masing-masing. Mahasiswa perlu memilih dosen pembimbing sesuai dengan bidang keilmuan dosen. Terkadang mahasiswa merasa kesulitan dalam menentukan topik dan melakukan kesalahan dengan memilih dosen pembimbing yang tidak sesuai dengan bidang keilmuannya sehingga dibutuhkan

suatu sistem yang dapat mengklasifikasikan dokumen tugas akhir tersebut sesuai dengan bidang yang telah ditentukan.

Klasifikasi dokumen bertujuan untuk mengelompokkan dokumen berdasarkan kategori dokumen yang telah ditentukan (Lestari et al., 2019). Ontologi merupakan istilah yang digunakan dalam berbagai bidang ilmu dengan tujuan menggambarkan sekumpulan objek yang terlibat serta hubungan di antara objek-objek tersebut. Hubungan ini biasanya direpresentasikan dalam kosakata dan diproses oleh perangkat lunak untuk menghasilkan pengetahuan (Cahyono et al., 2016). Klasifikasi teks berbasis ontologi merupakan proses klasifikasi teks dengan mengintegrasikan konsep ontologi dengan pendekatan struktur formal ontologi untuk mengorganisir dan memahami informasi yang ada di dalam teks.

Penelitian yang dilakukan oleh (Lestari et al., 2019) mengenai klasifikasi teks berbasis ontologi untuk dokumen tugas akhir berbahasa Indonesia menggunakan sebanyak 500 data abstrak tugas akhir. Ontologi yang dibangun berdasarkan *Association for Computing Machinery (ACM) IEEE Computer Society*. Metode pembobotan yang digunakan yaitu TF-IDF dengan menghitung nilai TF dan *inverse* frekuensi dokumen yang mengandung kata (IDF). Hasil akurasi yang diperoleh sebesar 87%. Namun hasil tersebut belum maksimal karena pada pengujian terdapat beberapa data abstrak yang tidak memiliki kecocokan dengan topik yang ada dan beberapa data abstrak terdeteksi memiliki lebih dari satu klasifikasi. Masalah tersebut muncul karena pada data abstrak terdapat *term* yang tidak memiliki relasi kesamaan antar kata dengan istilah yang ada pada ontologi.

Berdasarkan hasil penelitian tersebut, pada penelitian ini akan menggunakan metode pembobotan FastText untuk klasifikasi dokumen tugas akhir.

FastText merupakan salah satu metode pembobotan kata serta merupakan pengembangan lebih lanjut dari metode *embedding* Word2Vec, di mana n-gram sebuah karakter juga digunakan pada representasi kata (Fadhil & S, 2022). Dengan menggabungkan teknik continuous bag-of-words (CBOW) dan skip-gram, FastText dapat membentuk representasi vektor untuk setiap n-gram dan memperhitungkan hubungan antar kata atau kalimat. Penelitian sebelumnya oleh (Nurdin et al., 2020) mengenai perbandingan kinerja *word embedding* Word2Vec, GloVe, dan FastText didapatkan hasil bahwa kinerja terbaik dari penelitian diperoleh dengan menggunakan *word embedding* FastText.

Pada penelitian ini akan menggunakan metode pembobotan FastText untuk klasifikasi dokumen tugas akhir berbasis ontologi. FastText memiliki keunggulan dalam menghasilkan representasi vektor kata yang akurat, dengan mempertimbangkan struktur *sub-word* atau bagian-bagian kecil dari kata. Melalui penelitian ini diharapkan dapat memberikan solusi yang efektif dalam mengklasifikasikan dokumen tugas akhir berbasis ontologi.

1.3 Rumusan Masalah

Berdasarkan penjelasan pada latar belakang sebelumnya, diperoleh rumusan masalah sebagai berikut:

1. Bagaimana mengembangkan sistem klasifikasi berbasis ontologi untuk dokumen tugas akhir dengan metode pembobotan FastText?

2. Bagaimana kinerja klasifikasi berbasis ontologi untuk dokumen tugas akhir dengan pembobotan FastText?

1.4 Tujuan Penelitian

Tujuan penelitian ini berdasarkan rumusan masalah di atas adalah sebagai berikut:

1. Menghasilkan sistem yang dapat melakukan klasifikasi berbasis ontologi untuk dokumen tugas akhir menggunakan metode pembobotan FastText.
2. Mengetahui kinerja pembobotan FastText dalam mengklasifikasi dokumen tugas akhir berbasis ontologi.

1.5 Manfaat Penelitian

Adapun manfaat yang diperoleh dari penelitian ini adalah sebagai berikut:

1. Memperoleh informasi mengenai kinerja pembobotan FastText dalam klasifikasi berbasis ontologi untuk dokumen tugas akhir.
2. Hasil Penelitian dapat membantu pihak terkait dalam mengklasifikasi atau mengelompokkan tugas akhir sesuai dengan bidang-bidang yang telah ditetapkan dan dapat meningkatkan efisiensi pengelolaan informasi di lingkungan akademik.

1.6 Batasan Masalah

Batasan masalah yang telah ditetapkan dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini merupakan data abstrak dokumen tugas akhir mahasiswa Teknik Informatika Universitas Sriwijaya.
2. Data abstrak yang digunakan berbahasa Indonesia.
3. Data penelitian berjumlah 160 data.
4. Data diklasifikasikan menjadi empat kelas, yaitu Pemrosesan Bahasa Alami, Grafika Komputer dan Visualisasi, *Data Science* dan Pengenalan Pola, dan Sistem Terdistribusi.
5. Menggunakan metode *Cosine Similarity* untuk menghitung kesamaan makna antara dua kata.

1.7 Sistematika Penulisan

Berikut diuraikan sistematika dalam penulisan skripsi ini:

BAB I. PENDAHULUAN

Pada bab ini diuraikan gagasan yang menjadi latar belakang penulisan skripsi. Gagasan tersebut meliputi latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, sistematika penulisan, dan kesimpulan penelitian.

BAB II. KAJIAN LITERATUR

Pada bab ini berisi penjelasan secara rinci mengenai dasar-dasar teori yang akan digunakan pada penelitian ini, meliputi definisi tugas akhir, klasifikasi, ontologi, algoritma FastText, *Cosine similarity*, serta literatur ilmiah yang menjadi referensi dalam penelitian ini.

BAB III. METODE PENELITIAN

Pada bab ini akan dijelaskan tahap-tahap dalam penelitian, mulai dari pengumpulan data, analisis data, mekanisme pengujian, dan desain sistem yang akan dikembangkan. Setiap proses akan diuraikan secara rinci berdasarkan kerangka kerja yang telah ditentukan.

1.8 Kesimpulan

Berdasarkan penjelasan yang telah diuraikan sebelumnya, penelitian ini akan melakukan klasifikasi berbasis ontologi untuk dokumen tugas akhir dengan menggunakan metode pembobotan FastText.

DAFTAR PUSTAKA

- Asril, H., Mustakim, & Kamila, I. (2019). Klasifikasi Dokumen Tugas Akhir Berbasis Text Mining menggunakan Metode Naïve Bayes Classifier dan K-Nearest Neighbor. *Klasifikasi Dokumen Tugas Akhir Berbasis Text Mining Menggunakan Metode Naïve Bayes Classifier Dan K-Nearest Neighbor*.
- Cahyono, N., Utami, E., & Amborowati, A. (2016). Konseptualisasi Ontologi Iklan Website. *Jurnal Buana Informatika*, 7. <https://doi.org/https://doi.org/10.24002/jbi.v7i4.768>
- Choi, J., & Lee, S. W. (2020). Improving FastText with inverse document frequency of subwords. *Pattern Recognition Letters*, 133, 165–172. <https://doi.org/10.1016/j.patrec.2020.03.003>
- Dharma, E. M., Lumban Gaol, F., Leslie, H., Warnars, H. S., & Soewito, B. (2022). THE ACCURACY COMPARISON AMONG WORD2VEC, GLOVE, AND FASTTEXT TOWARDS CONVOLUTION NEURAL NETWORK (CNN) TEXT CLASSIFICATION. *Journal of Theoretical and Applied Information Technology*, 31(2). www.jatit.org
- Fadhil, I. M., & S, Y. S. (2022). Klasifikasi Topik Pada Tweet Berbahasa Indonesia menggunakan Ekspansi fitur Fast-Text dengan Metode Support Vector Machine (SVM). *E-Proceeding of Engineering*, 9(Vol.9, No.3 Juni 2022), 1970–1977.

- Farhan AlShammari, A. (2023). Implementation of Text Similarity using Word Frequency and Cosine Similarity in Python. In *International Journal of Computer Applications* (Vol. 185, Issue 36).
- Gasparetto, A., Marcuzzo, M., Zangari, A., & Albarelli, A. (2022). Survey on Text Classification Algorithms: From Text to Predictions. *Information (Switzerland)*, 13(2). <https://doi.org/10.3390/info13020083>
- Hossain, M. R., Hoque, M. M., & Sarker, I. H. (2021). Text Classification Using Convolution Neural Networks with FastText Embedding. *Advances in Intelligent Systems and Computing*, 1375 AIST, 103–113. https://doi.org/10.1007/978-3-030-73050-5_11
- Kristiani, S. A. (2019). *KLASIFIKASI DOKUMEN MENGGUNAKAN SUPPORT VECTOR MACHINE DAN MUTUAL INFORMATION*. Universitas Komputer Indonesia.
- Lestari, A. P., Maskur, & Hayatin, N. (2019). Klasifikasi Teks Berbasis Ontologi Untuk Dokumen Tugas Akhir Berbahasa Indonesia. *Repositor*, 1, 79–86. <https://doi.org/10.22219/repositor.v1i2.23>
- Liebenlito, M., Yesinta, A. A., & Musti, M. I. S. (2024). Deteksi Clickbait pada Judul Berita Online Berbahasa Indonesia Menggunakan FastText. *Journal of Applied Computer Science and Technology*, 5(1), 56–62. <https://doi.org/10.52158/jacost.v5i1.655>

- Malik, S., & Jain, S. (2021). Semantic Ontology-Based Approach to Enhance Text Classification. *ISIC'21: International Semantic Intelligence Conference*, 25–27.
- Montes, C. D., Silvosa, J. V., Abalorio, C. C., & Nakazato, R. B. (2024). Application of BERT Model for Unsupervised Text Classification using Hierarchical Clustering for Automatic Classification of Thesis Manuscript. *5th International Conference on Electronics and Sustainable Communication Systems, ICESC 2024 - Proceedings*, 278–284. <https://doi.org/10.1109/ICESC60852.2024.10690039>
- Mustofa, H., & Mahfudh, A. A. (2019). Klasifikasi Berita Hoax Dengan Menggunakan Metode Naive Bayes. *Walisongo Journal of Information Technology*, 1(1), 1. <https://doi.org/10.21580/wjit.2019.1.1.3915>
- Nanda, A. E., Sihananto, A. N., & Rizki, A. M. (2024). Analisis Sentimen Pada Pembatalan Tuan Rumah Indonesia Di Piala Dunia U-20 Menggunakan Fasttext Embeddings Dan Algoritma Recurrent Neural Network. 2(2), 246–257. <https://doi.org/10.59841/saber.v2i2.1000>
- Nurdin, A., Seno aji, B., Bustamin, A., & Abidin, Z. (2020). PERBANDINGAN KINERJA WORD EMBEDDING WORD2VEC, GLOVE, DAN FASTTEXT PADA KLASIFIKASI TEKS. *Jurnal Tekno Kompak*, 14, 74. <https://doi.org/10.33365/jtk.v14i2.732>

- Park, K., Hong, J. S., & Kim, W. (2020). A Methodology Combining Cosine Similarity with Classifier for Text Classification. *Applied Artificial Intelligence*, 34(5), 396–411. <https://doi.org/10.1080/08839514.2020.1723868>
- Pawestri, S., & Suyanto, Y. (2024). Analisis Perbandingan Metode Similarity untuk Kemiripan Dokumen Bahasa Indonesia pada Deteksi Kemiripan Teks Bahasa Indonesia. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 8(3), 1440. <https://doi.org/10.30865/mib.v8i3.7648>
- Pribadi, N. H., Sarno, R., Ahmadiyah, A. S., & Sungkono, K. R. (2021). Semantic Recommender System Based on Semantic Similarity Using FastText and Word Mover's Distance. *International Journal of Intelligent Engineering and Systems*, 14(2), 377–385. <https://doi.org/10.22266/ijies2021.0430.34>
- Proboningrum, S., & Sidauruk, A. (2021). SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN SUPPLIER KAIN DENGAN METODE MOORA. *Sistem Informasi* |, 8(1), 43–48.
- Rodiah, D. (2024). Pengembangan Representasi Pengetahuan Ontologi Domain Bidang Ilmu Informatika. *Jurnal PROCESSOR*, 19(2). <https://doi.org/10.33998/processor.2024.19.2.1905>
- Siregar, S. W., & Putri, R. A. (2024). Information System for Data Collection on Building Tools and Materials Using the Rational Unified Process (RUP)

Method. *PIKSEL : Penelitian Ilmu Komputer Sistem Embedded and Logic*, 12(2), 455–466. <https://doi.org/10.33558/piksel.v12i2.9951>

Widianto, A., & Pebriyanto, E. (2024). Document Similarity using Term Frequency-Inverse Document Frequency Representation and Cosine Similarity. *Journal of Dinda Data Science, Information Technology, and Data Analytics*, 4(2), 149–153. <http://journal.ittelkom-pwt.ac.id/index.php/dinda>

Zulbichaq, A., Musrsityo, Y. T., & Saputra, M. C. (2019). Pengembangan Sistem Informasi Penggajian Pada Tugu Hotel Lombok Menggunakan Metode Rational Unified Process. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(3), 2490–2498. <http://j-ptiik.ub.ac.id>