

Representasi Pengetahuan Ontologi untuk Klasifikasi Topik Penelitian pada Bidang Ilmu Informatika

Desty Rodiah*¹, Kanda Januar Miraswan², Junia Kurniati³, Dellin Irawan⁴, Vanya Terra Ardani⁵

^{1,2,3,4,5}Jurusan Teknik Informatika, Universitas Sriwijaya, Palembang, Indonesia

*Email Penulis Korespondensi: destyrodiah@ilkom.unsri.ac.id

Abstrak

Penelitian di bidang informatika yang sering melibatkan beberapa subdisiplin dalam satu penelitian mengakibatkan kesulitan dalam melakukan klasifikasi dari topik penelitian. Biasanya klasifikasi teks pada bidang pemrosesan bahasa alami melalui proses pelatihan dan pengujian. Namun klasifikasi yang menggunakan metode ontologi yaitu klasifikasi tanpa melakukan pelatihan, sehingga mengatasi permasalahan akan kebutuhan data latih. Pada klasifikasi dengan ontologi terdapat beberapa masalah yaitu pencarian term yang tidak memiliki kesamaan antar kata dengan ontologi dan ketepatan dalam pengukuran kemiripan data dengan representasi pengetahuan. Untuk mengatasi hal tersebut digunakan metode *fastText* yang melakukan pencarian term agar memiliki relasi kesamaan antar kata dengan ontologi. Dan metode Wu-Palmer adalah metode untuk mengukur kemiripan dan keterkaitan makna semantik kata dengan representasi pengetahuan dalam ontologi. Tahapan dalam penelitian ini dimulai dari *preprocessing* (*Casefolding*, *Tokenizing*, *Stopword Removal*, *Lemmatization*), *Query Processing* (*Query Reduction* dan *Remove Duplicate Query*), *Word Embedding* dengan *FastText* dan *Semantic Similarity* dengan metode Wu-Palmer. Data yang digunakan dalam penelitian ini berasal dari penelitian hasil dari tugas akhir mahasiswa teknik informatika Fasilkom Unsri dengan jumlah data sebanyak 200 penelitian. Hasil dari klasifikasi ini didapatkan 178 dari 200 topik yang diklasifikasi sesuai dengan data aktual dengan nilai akurasi sebesar 89,5% yang menunjukkan bahwa sistem secara umum mampu melakukan klasifikasi dengan baik.

Abstract

Research in informatics often involves multiple subdisciplines, making topic classification challenging. Typically, text classification in natural language processing requires training and testing. However, ontology-based classification eliminates the need for training data. Challenges in ontology-based classification include finding terms that lack similarity with ontology and ensuring accuracy in measuring data similarity with knowledge representation. To address this, the *fastText* method identifies term similarities between words and ontology, while the Wu-Palmer method measures semantic similarity and relationships within ontology. The research process includes *preprocessing* (*Casefolding*, *Tokenizing*, *Stopword Removal*, *Lemmatization*), *Query Processing* (*Query Reduction*, *Duplicate Removal*), *Word Embedding* with *fastText*, and *Semantic Similarity* measurement using Wu-Palmer. The dataset consists of 200 research studies from Fasilkom Unsri informatics students' final projects. The classification results show that 178 out of 200 topics were correctly classified, achieving an accuracy of 89.5%, demonstrating the system's effectiveness.

Keywords: Classification, Final Project Research, Ontologi, Wu-Palmer, Natural Language Processing

1. PENDAHULUAN

Penelitian di bidang informatika sering kali melibatkan beberapa subdisiplin, karena banyak masalah yang kompleks memerlukan pendekatan multidisiplin [1]. Beberapa contoh penelitian informatika yang melibatkan banyak subdisiplin seperti

kecerdasan buatan untuk keamanan siber dimana subdisiplin yang terkait yaitu kecerdasan buatan, pembelajaran mesin, keamanan siber dan analisis data besar. Semakin banyak penelitian yang multi disiplin, maka akan semakin sulit pula untuk untuk mengkategorikan penelitian tersebut jika harus dilakukan kategori secara manual. Kesulitan dalam kategori penelitian misalnya penelitian berupa skripsi yang banyak membahas mengenai topik yang serupa, baik teori maupun metode yang digunakan [2]. Bahkan kadang juga menggunakan data yang sama dan tujuan yang serupa sehingga tidak jarang hasil penelitian menjadi mirip. Selain itu ada juga penelitian yang mengangkat topik yang sama namun menggunakan istilah dan metode yang berbeda sehingga kesulitan juga dalam menentukan apakah topik tersebut sudah pernah digunakan atau belum [2]. Sehingga permasalahan klasifikasi topik penelitian harus segera diselesaikan

Klasifikasi teks pada bidang pemrosesan bahasa alami biasanya dimulai dengan proses pelatihan dan dilanjutkan dengan proses pengujian [3]. Metode ontologi merupakan metode untuk melakukan klasifikasi tanpa melalui proses pelatihan. Hal ini dapat mengatasi permasalahan akan kebutuhan data latih pada proses klasifikasi. Sebuah ontologi menyatakan secara formal bidang topik diskusi [4]. Sebuah ontologi terdiri dari hubungan antara berbagai istilah yang berada pada daftar istilah terbatas. Penelitian yang dilakukan oleh [5] melakukan pencarian tugas akhir dengan ontologi dan Booyer Moore. Pada penelitian tersebut membangun struktur ontologi RepositoryTA untuk Jurusan Teknik Informatika UNSRI. Sistem pencarian tersebut memanfaatkan ontologi dan algoritma booyer moore untuk perhitungan nilai *similarity* antar data yang didapatkan dengan *keyword* yang dimasukkan. Hasil penelitian tersebut mendapatkan akurasi sebesar 99,84% untuk 16 kali percobaan. Penelitian yang dilakukan oleh [4] mengenai klasifikasi dokumen berita teks berbahasa Indonesia menggunakan ontologi, pada penelitian tersebut menggunakan data dari google. Sistem dibuat dengan tahap *preprocessing* dan perancangan ontologi. Struktur ontologi dibuat berdasarkan undang - undang No 24 Tahun 2007. Sistem peramban ontologi untuk mengambil data berdasarkan ontologi. Keluaran dari program adalah halaman yang mengandung kata kunci yang tersimpan dari *file* ontologi. Penelitian lain adalah mengenai klasifikasi topik skripsi berdasarkan makna dengan pendekatan semantik web. Pada penelitian yang dilakukan oleh [2] melakukan pemetaan bidang minat skripsi mahasiswa S1 Teknik Informatika Unpad. Klasifikasi aktual menunjukkan bahwa skripsi mahasiswa didominasi oleh topik Sistem Informasi dan Multimedia sebesar 50,23% dan klasifikasi prediktif dilakukan dengan algoritma KNN dan menghasilkan persentase sebesar 47,88%. Penelitian yang dilakukan oleh [6] melakukan penelitian mengenai klasifikasi teks berbasis ontologi untuk dokumen tugas akhir berbahasa indonesia. Data yang digunakan sebanyak 500 data abstrak dari skripsi. Ontologi dibangun berdasarkan *Association for Computing Machinery (ACM) IEEE Computer Society*. Untuk menghitung kemiripan diantara dokumen dari sebuah node ontologi digunakan metode dao. Hasilnya menunjukkan akurasi sebesar 87%. Namun hasil yang diberikan belum maksimal yaitu pada pengujian terdapat beberapa abstrak yang tidak ditemukan topik yang sesuai dan ada beberapa abstrak yang terdeteksi lebih dari satu klasifikasi. Masalah tersebut terjadi karena pada abstrak tersebut terdapat *term* yang tidak memiliki relasi kesamaan antar kata dengan istilah pada ontologi dan terdapat juga faktor dari nilai dao yang sama pada ontologi.

Pada penelitian ini akan mencoba untuk mengatasi masalah pencarian term agar memiliki relasi kesamaan antar kata dengan ontologi. Pencarian term itu berkaitan dengan metode ekstraksi fitur [7]. Pada bidang pemrosesan teks, ekstraksi fitur dari karakteristik teks yang tidak terstruktur menjadi tantangan tersendiri [8]. Pada penelitian yang dilakukan oleh [8] yang melakukan perbandingan untuk 3 metode *word embedding* yaitu Word2Vec, GloVe dan *fastText* lalu dilakukan klasifikasi dengan menggunakan algoritma *Convolutional Neural Network*. Dataset yang digunakan adalah 20 *newsgroup* dan *Reuters Newswire*. Berdasarkan penelitian tersebut didapatkan bahwa *fastText* unggul dengan nilai *F-Measure* sebesar 0,979 untuk dataset *newsgroup* dan 0,715 untuk *reuters*. Penelitian yang dilakukan oleh [9] melakukan klasifikasi *stance* dengan menggunakan *fastText* pada proses *embedding* dan pendekatan *deep learning*. Dataset yang digunakan adalah *stance* pada judul post kesehatan di *facebook* dengan jumlah data 3500 judul post. Model *fastText* menghasilkan nilai *F1 macro score* sebesar 53,8% dan 52,7% untuk model *word2Vec*. Penelitian yang dilakukan oleh [10] melakukan klasifikasi sentimen twitter mengenai Program Vaksinasi Covid-19. Metode yang digunakan dalam penelitian tersebut adalah *Support Vector Machine* (SVM) untuk melakukan klasifikasi dan fitur *fastText language mode* untuk merepresentasikan kata-kata pada *dataset*. Nilai *F1-Scores* yang dihasilkan oleh metode SVM dan fitur *fastText* lebih baik yaitu sebesar 59% dibandingkan dengan Naive Bayes sebesar 57% dan SVM fitur TF-IDF 56% dan LSTM fitur Word2Vec 54%.

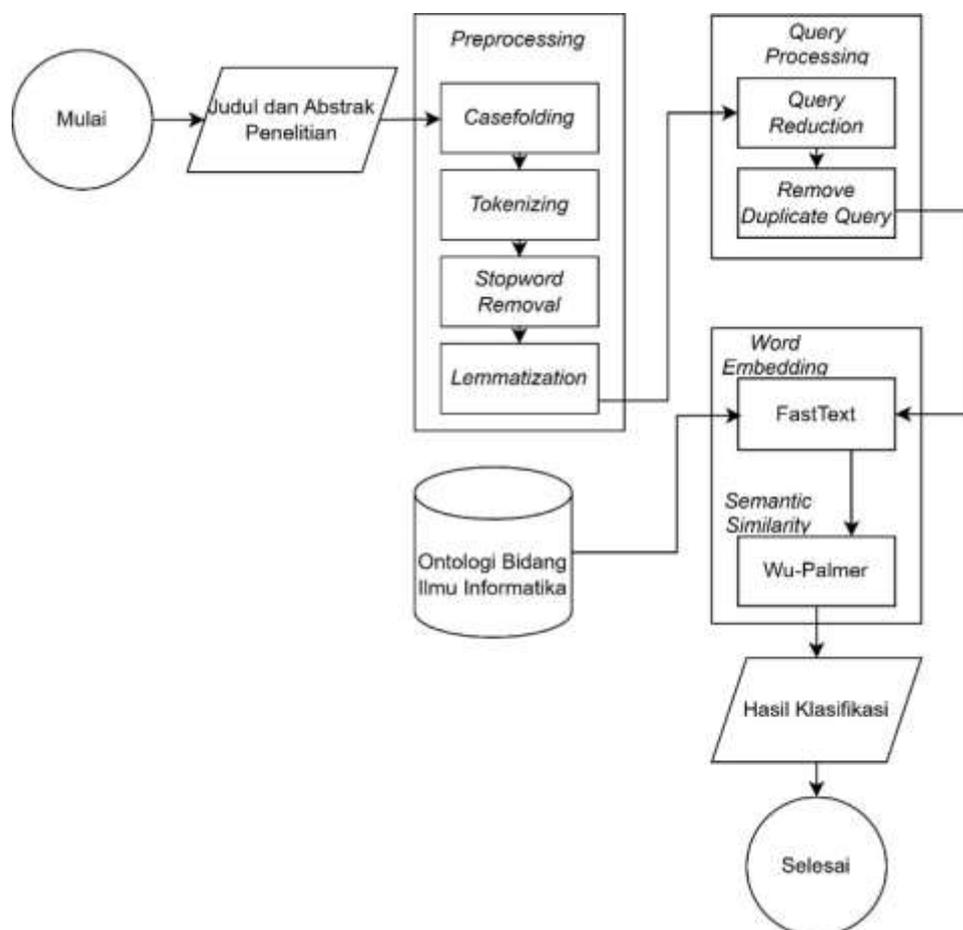
Untuk mengimplementasikan penggunaan ontologi dalam kasus klasifikasi topik penelitian, tantangan lain adalah pengukuran kemiripan data dengan representasi pengetahuan dalam ontologi. Seperti permasalahan yang terjadi pada penelitian [6] yang menyebutkan bahwa terdapat beberapa abstrak yang terdeteksi lebih dari satu klasifikasi, hal ini disebabkan oleh faktor dari nilai *dao* yang sama pada ontologi. Dengan menggunakan algoritma kemiripan semantik, Wu-Palmer dapat mengukur seberapa relevan atau terkait setiap dokumen atau term dan mampu mengukur derajat kemiripan makna semantik antar kata [11]. Penelitian yang dilakukan oleh [12] yang melakukan perbandingan algoritma *cosinus* dan *wu-palmer* dalam mencari kemiripan kata. Kasus yang digunakan adalah pada *plagiasim checker*. Untuk menguji penggunaan algoritma Wu-Palmer dalam studi kasus mendeteksi plagiasi, peneliti menggunakan situs ws4jdemo.appspot.com. Hasilnya menunjukkan bahwa Wu-Palmer lebih mudah menemukan kata-kata dalam *wordnet* daripada kalimat. Penelitian yang dilakukan oleh [13] yang melakukan perbandingan dalam menghitung kesamaan konsep secara semantic pada *Latent Semantic* terhadap pendekatan *WordNet* pendekatan *Latent Semantic* terhadap *WordNet* untuk *Semantic Similarity*. Pendekatan *WordNet* yang digunakan adalah Wu Palmer, Jiang Conrath. *Latent semantic* menggunakan *corpus* dari *General Reading up to 1st year collage* dan dari *Encyclopedia*. Hasil penelitian tersebut berdasarkan nilai *F-measure* yang menunjukkan total kinerja, *WordNet* dengan Wu-Palmer menunjukkan kecenderungan yang lebih baik dari pada metode lain.

Berdasarkan penelitian terkait maka penelitian ini melakukan klasifikasi topik penelitian pada bidang ilmu informatika berdasarkan judul dan abstrak dengan memanfaatkan representasi pengetahuan ontologi. Untuk mendukung kinerja ontologi dalam melakukan klasifikasi, akan menggunakan metode *fastText* dan metode Wu-Palmer. Metode *fastText* adalah metode *word embedding* yang melakukan pencarian *term*

agar memiliki relasi kesamaan antar kata dengan ontologi. Sedangkan metode Wu-Palmer adalah metode untuk mengukur kemiripan dan keterkaitan makna semantik kata dengan representasi pengetahuan dalam ontologi.

2. METODOLOGI PENELITIAN

Sistem klasifikasi yang dikembangkan adalah untuk menentukan topik penelitian di bidang ilmu informatika berdasarkan judul dan abstrak. Pada kerangka fikir penelitian yang terlihat pada gambar 1, terdapat tahapan perancangan sistem. Arsitektur sistem klasifikasi yang dikembangkan diharapkan menjadi solusi dalam permasalahan melakukan klasifikasi.



Gambar 1. Arsitektur Representasi Pengetahuan Ontologi untuk Klasifikasi Topik Penelitian

Berikut adalah penjelasan Gambar 1 tahapan arsitektur representasi pengetahuan ontologi untuk klasifikasi topik penelitian

2.1. Dataset

Data yang diperlukan untuk membangun sistem klasifikasi topik penelitian pada bidang ilmu informatika adalah judul dan abstrak tugas akhir yang didapatkan dari tugas

akhir mahasiswa teknik informatika Fasilkom Unsri. Bidang Ilmu yang digunakan pada penelitian ini ada empat yaitu Data science dan pengenalan pola, Sistem Terdistribusi, Grafika Komputer dan Visualisasi serta pemrosesan Bahasa Alami. Data tersebut diambil dari kumpulan tugas akhir yang telah dikumpulkan diperpustakaan. Tugas akhir yang digunakan ada 200 tugas akhir.

2.2.Pre-processing

Tahap pertama judul dan abstrak penelitian di lakukan preprocessing yaitu *case folding*, tokenizing, stopword removal dan lemmatization. *Case folding* mengubah semua huruf dalam dokumen menjadi huruf kecil, atau huruf bawah [14]. Tokenisasi adalah proses memecah teks menjadi bagian yang lebih kecil [15]. Proses menghapus kata-kata dari stoplist disebut *stopword removal* [16]. Stoplist adalah daftar kata yang tidak relevan yang muncul dalam dokumen. Proses mengurangi sebuah kata menjadi bentuk dasar, atau bentuk kata yang ditemukan dalam kamus, disebut *lemmatization* [17].

2.3.Query Processing

Pada tahap *query processing* dilakukan proses *query reduction* dan *remove duplicate query*. *Query reduction* adalah proses pengurangan jumlah informasi yang perlu diproses untuk mengevaluasi query pengguna [18]. Jadi pada *query reduction* mengambil query yang lebih spesifik dibandingkan query yang dimasukkan. Setelah dilakukan *query reduction* dilanjutkan dengan proses *remove duplicate query* dimana proses ini menghapus query yang duplikat.

2.4.Word Embedding

Sebelum masuk ke proses *word embedding*, query yang dihasilkan akan dilakukan proses pencarian dari setiap kata hasil query ke ontologi. Ontologi yang digunakan adalah ontologi Bidang Ilmu Informatika yang dikembangkan oleh [19]. Proses pencarian query dilakukan untuk setiap domain dengan menggunakan sparql query [20]. Setelah proses pencarian dengan ontologi, maka semua kata yang ditemukan pada hasil pencarian akan dilakukan seperti proses *preprocessing* dan *query processing* juga. Lalu proses selanjutnya adalah proses *word embedding*. Setiap kata query maupun hasil dari pencarian pada sparql query diubah ke dalam bentuk vektor dengan menggunakan metode fastText. Metode fastText adalah metode word embedding yang merupakan pengembangan dari word2vec [8], [9], [21]. Metode ini melihat representasi kata dengan melihat informasi subword.

2.5.Sementic Similarity

Hasil dari perhitungan vektor dengan metode fastText akan dihitung *similarity* dari kedua vektor kata dengan menggunakan metode wu-palmer. Algoritma wu-palmer adalah algoritma kemiripan semantik yang dapat mengukur tingkat keterkaitan atau relevansi antar dokumen dan term [11], [12], [22]. Untuk menghitung nilai kemiripan kata pada metode Wu-Palmer menggunakan Pers (1). Hasil dari perhitungan nilai kemiripan kata dari setiap domain akan dihitung rata-ratanya dan dijadikan sebagai nilai *similarity* dari masing - masing domain bidang ilmu informatika. Nilai *similarity* yang tertinggi akan menjadi hasil klasifikasi topik dari query yang dimasukkan.

$$\text{sim}_{\text{wup}}(s1, s2) = \frac{2 \times \text{depth (LCS)}}{(\text{depth}(s1)+\text{depth}(s2))} \quad (1)$$

2.6. Model Evaluasi

Evaluasi dilakukan untuk mengetahui keakuratan dan ketepatan model yang telah dibuat dalam melakukan klasifikasi, evaluasi akan dilakukan dengan menggunakan data tes. Confusion matrix adalah tabel silang yang mencatat jumlah kejadian antara dua penilai, klasifikasi benar / aktual dan klasifikasi yang diprediksi [23]. Metriks yang dilakukan adalah metriks akurasi, precision, recall dan F1-Score. Akurasi adalah salah satu metrik yang paling populer dalam klasifikasi multi-kelas dan secara langsung dihitung dari confusion matriks [24]. Presisi merupakan rasio antara *True Positive* (TP) dengan keseluruhan data yang diprediksi positif. Sehingga, precision berusaha memperkecil terjadinya *False Positive* (FP). Recall merupakan rasio antara *True Positive* (TP) dengan keseluruhan datayang kenyataannya bernilai positif. Sehingga, recall berusaha memperkecil terjadinya *False Negative* (FN). F1-score merupakan harmonic mean dari *precision* dan *recall* [25].

3. HASIL DAN ANALISIS

Data yang diperlukan untuk membangun sistem klasifikasi topik penelitian pada bidang ilmu informatika adalah judul dan abstrak tugas akhir yang didapatkan dari tugas akhir mahasiswa teknik informatika Fasilkom Unsri yang terdapat pada repository unsri. Data yang di gunakan ada 200 penelitian.

Tahap pertama judul dan abstrak penelitian di lakukan *preprocessing* yaitu *case folding*, *tokenizing*, *stopword removal* dan *lemmatization*. Setelah dilakukan *preprocessing* maka setiap kata hasil dari *preprocessing* dilakukan query *processing* dengan teknik query *reduction*, dimana setiap kata akan di cari sinonim yang lebih spesifik dari kata tersebut. Setelah itu dilakukan proses penghapusan kata yang duplikat. Setelah proses query *reduction* maka dilakukan proses pencarian dari setiap kata hasil ke ontologi dengan menggunakan sparql query. Dilakukan proses pencarian kata untuk setiap domain dengan menggunakan sparql query. Setelah dilakukan proses pencarian dengan ontologi, maka semua kata yang ditemukan pada hasil pencarian akan dilakukan seperti proses *preprocessing* juga, agar pada saat dilakukan pencocokan kata data yang digunakan sudah bersih. Lalu proses selanjutnya adalah proses perbandingan kata. Kata yang akan dibandingkan adalah kata dari hasil *query reduction* pada judul dan abstrak penelitian dibandingkan dengan kata yang didapatkan dari hasil sparql query setiap domain. Setiap kata yang akan dibandingkan akan dilakukan diubah dalam bentuk vektor dengan menggunakan metode *fastText*. Hasil dari perhitungan vektor dengan metode *fastText* akan dihitung *similarity* dari kedua vektor kata dengan menggunakan metode wu-palmer. Hasil dari bobot kata dari setiap domain akan dihitung rata-ratanya dan dijadikan sebagai nilai *similarity* dari masing - masing domain bidang ilmu informatika. Setelah mendapatkan nilai *similarity* dari setiap domain, maka nilai *similarity* yang paling besar akan kelas dari judul dan abstrak tugas akhir tersebut.

Telah dilakukan pengujian untuk 200 data tugas akhir dengan hasil ditunjukkan pada Tabel 2, dan keterangan penomoran domain ditunjukkan pada Tabel 1.

Tabel 1. Keterangan Domain

No	Domain
1	Data Science dan Pengenalan Pola
2	Sistem Terdistribusi
3	Pemrosesan Bahasa Alami
4	Grafika Komputer dan Visualisasi

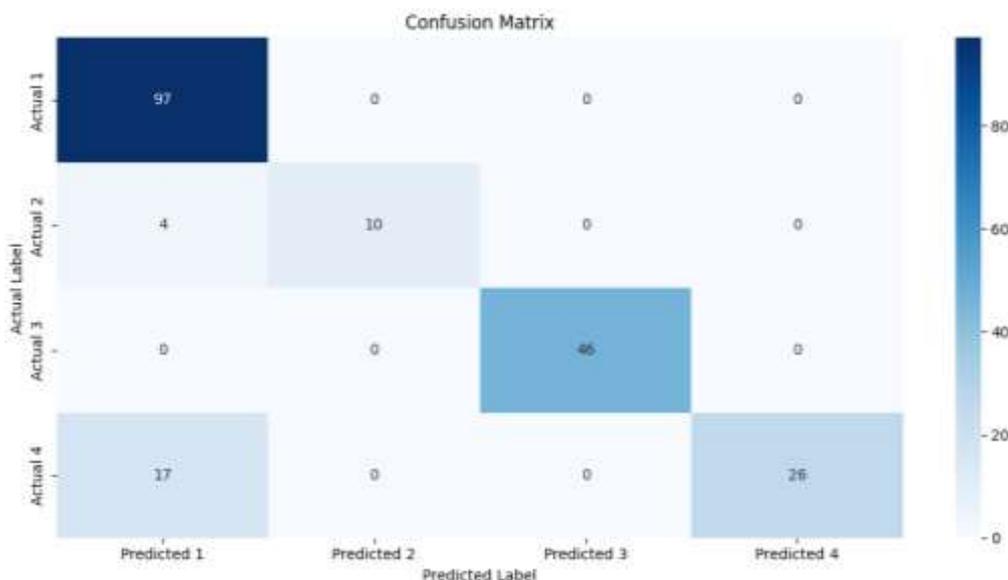
Tabel 2. Hasil Pengujian

No	Nilai Similarity	Actual	Predict	Ket
1	[(1, 0.33564250268607276), (3, 0.19928848236895186), (4, 0.183088649827911), (2, 0.1593004126100916)]	1	1	Sesuai
2	[(1, 0.2974927030487694), (2, 0.2773069928727061), (3, 0.27083740272726264), (4, 0.24076689557457961)]	2	1	Tidak Sesuai
3	[(1, 0.3524697451871226), (3, 0.23901830781557465), (4, 0.19924241071763782), (2, 0.17187962287249775)]	1	1	Sesuai
4	[(1, 0.3022186356356153), (3, 0.19915891892096732), (2, 0.14171024943621552), (4, 0.1346049268604826)]	1	1	Sesuai
5	[(1, 0.3410767896093489), (3, 0.20971303843881695), (4, 0.15913940796241816), (2, 0.14243911989574598)]	1	1	Sesuai
6	[(1, 0.325247124744993), (3, 0.1872081403093777), (4, 0.17091308274351258), (2, 0.1478716038451897)]	1	1	Sesuai
7	[(2, 0.27266606356355516), (1, 0.2539805319529774), (3, 0.22613367225929798), (4, 0.1980069236706135)]	2	2	Sesuai
8	[(1, 0.3110698391109303), (3, 0.1967392878445668), (4, 0.1482814312321947), (2, 0.14740897570866282)]	1	1	Sesuai
9	[(2, 0.2751290833124576), (1, 0.25923613472536317), (3, 0.24334610100033632), (4, 0.21773138496711728)]	2	2	Sesuai

10	[(3, 0.3320461336343777), (1, 0.26976725740133806), (4, 0.18704120594346113), (2, 0.14920783352710298)]	3	3	Sesuai
...
200	[(3, 0.30419217876817234), (1, 0.24987199817216954), (4, 0.19402105672785225), (2, 0.18415856403737416)]	3	3	Sesuai

Berdasarkan hasil pengujian didapatkan 178 dari 200 topik yang di klasifikasikan sesuai dengan data aktual dan 21 topik yang diklasifikasikan tidak sesuai. Hampir semua topik yang tidak sesuai masuk ke dalam kelas 1 (Data science dan Pengenalan Pola). Hal ini terjadi karena pada domain data science dan pengenalan pola memiliki struktur pengetahuan yang luas hampir ke seluruh domain informatika, sehingga semua hasil klasifikasi lebih condong ke domain tersebut. Kelas yang cukup banyak terklasifikasi salah adalah kelas 4 (Grafika Komputer dan Visualisasi). Hal ini terjadi karena struktur pengetahuan dari Grafika Komputer dan Visualisasi termasuk dalam rumpun pengenalan pola sehingga banyak hasil klasifikasi lebih condong ke domain Data Science dan Pengenalan Pola.

Dari hasil pengujian, dilakukan perhitungan *confusion matriks* didapatkan hasil seperti ditunjukkan pada Gambar 2 sebagai berikut:



Gambar 2. Confusion Matriks

Dari hasil confusion matriks maka dapat dihitung nilai akurasi, presisi, recall dan F1-score, dengan hasil dilihat pada Gambar 3.

	precision	recall	f1-score	support
1	0.822	1.000	0.902	97
2	1.000	0.714	0.833	14
3	1.000	1.000	1.000	46
4	1.000	0.605	0.754	43
accuracy			0.895	200
macro avg	0.956	0.830	0.872	200
weighted avg	0.914	0.895	0.888	200

Gambar 3. Hasil perhitungan *precision*, *recall*, *f1-score* dan *accuracy*

Berdasarkan nilai yang didapatkan, maka dapat disimpulkan bahwa akurasi dari klasifikasi topik penelitian yaitu 89,5% dari total prediksi menunjukkan bahwa sistem secara umum melakukan prediksi yang baik. Presisi sebesar 95,6% yang artinya sistem sangat salah memberikan prediksi positif yang salah (FP rendah). *Recall* sebesar 83% menunjukkan bahwa ada beberapa kasus positif yang terlewat oleh sistem. sebesar 0,87 menunjukkan keseimbangan antara presisi yang tinggi (0,95) dan recall yang lebih rendah (0,83).

4. KESIMPULAN

Berdasarkan hasil yang didapatkan maka dapat disimpulkan bahwa sistem klasifikasi topik tugas akhir dengan menggunakan ontologi telah berhasil di kembangkan dan didapatkan hasil 178 dari 200 topik dapat diklasifikasikan sesuai dengan data aktual. Nilai akurasi yang didapatkan sebesar 89.5% yang menunjukkan bahwa sistem secara umum mampu melakukan klasifikasi dengan baik. Hampir semua topik yang terklasifikasi tidak sesuai dengan data aktual itu ter-klasifikasi ke dalam kelas 1 (Data science dan Pengenalan Pola). Hal ini terjadi karena pada domain data science dan pengenalan pola memiliki struktur pengetahuan yang luas hampir ke seluruh domain informatika, sehingga semua hasil klasifikasi lebih condong ke domain tersebut. Saran-saran untuk untuk penelitian lebih lanjut yaitu butuh penelitian lebih lanjut untuk mengembangkan struktur ontologi agar hasil klasifikasi tidak berfokus hanya ke satu kelas saja. Mengembangkan kembali query sparql yang digunakan agar memberikan hasil yang lebih spesifik. Menggunakan teknik pemrosesan query yang lain agar hasil pencarian menjadi lebih spesifik. Menggunakan menggunakan metode *similarity* yang lebih detail dalam menghitung tingkat kesamaan kata.

5. UCAPAN TERIMA KASIH

Penelitian/Publikasi artikel ini dibiayai oleh Anggaran DIPA (Daftar Isian Pelaksanaan Anggaran) dengan Nomor : SP DIPA-023.1.7.2.677515/2024 tanggal 24 November 2023 Sesuai Surat Perjanjian Pelaksanaan Pekerjaan Hibah Penelitian Sateks Bagi Dosen Fakultas Ilmu Komputer Universitas Sriwijaya dengan Nomor: 31206/UN9.FIK/TU.SB5/2024 tanggal 22 Juli 2024.

DAFTAR PUSTAKA

- [1] F. Sulianta, *Metode Riset Informatika*. 25AD.
- [2] A. Pradana and R. Ridwansyah, “Klasifikasi Topik Skripsi Berdasarkan Makna dengan Pendekatan Semantik Web,” 2021. [Online]. Available: <https://jurnal.pcr.ac.id/index.php/jkt/>
- [3] K. Arifin and S. I. A. I. Idrus, “Klasifikasi Emosi Pengguna Twitter Terhadap Bakal Calon Presiden Pada Pemilu 2024 Menggunakan Algoritma Naïve Bayes,” *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)*, vol. 23, pp. 37–45, Feb. 2024, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jis/index>
- [4] H. Februariyanti and E. Zuliarso, “Klasifikasi Dokumen Berita Teks Bahasa Indonesia menggunakan Ontologi,” *Jurnal Teknologi Informasi DINAMIK*, vol. 17, no. 1, pp. 14–23, 2012, [Online]. Available: <http://www.google.com>
- [5] D. Rodiah, Yunita, and N. Yusliani, “Pencarian Tugas Akhir dengan Ontologi dan Boyer-Moore (Studi Kasus: Jurusan Teknik Informatika UNSRI),” *Generic*, vol. 15, no. 1, 2023.
- [6] A. P. Lestari, Maskur, and N. Hayatin, “Klasifikasi Teks Berbasis Ontologi Untuk Dokumen Tugas Akhir Berbahasa Indonesia,” *REPOSITOR*, vol. 1, no. 2, pp. 79–86, 2019.
- [7] S. Norindah Sari, M. Reza Faisal, D. Kartini, I. Budiman, and T. Hamonangan Saragih, “Perbandingan Ekstraksi Fitur dengan Pembobotan Supervised dan Unsupervised pada Algoritma Random Forest untuk Pemantauan Laporan Penderita COVID-19 di Twitter,” *Jurnal Komputasi*, vol. 11, no. 1, 2023.
- [8] A. Nurdin, B. Anggo, S. Aji, A. Bustamin, and Z. Abidin, “PERBANDINGAN KINERJA WORD EMBEDDING WORD2VEC, GLOVE, DAN FASTTEXT PADA KLASIFIKASI TEKS,” *Jurnal TEKNOKOMPAK*, vol. 14, no. 2, p. 74, 2020.
- [9] E. Lim, E. I. Setiawan, and J. Santoso, “Stance Classification Post Kesehatan di Media Sosial Dengan FastText Embedding dan Deep Learning,” *JOURNAL OF INTELLIGENT SYSTEMS AND COMPUTATION*, vol. 1, no. 2, pp. 65–73, 2019, doi: 10.52985/insyst.v1i2.86.
- [10] M. M. Kusairi and S. Agustian, “SVM Method with FastText Representation Feature for Classification of Twitter Sentiments Regarding the Covid-19 Vaccination Program,” *Jurnal Teknologi Informasi dan Komunikasi*, vol. 13, no. 2, 2022.
- [11] G. U. Abriani and M. A. Yaqin, “Implementasi Metode Semantic Similarity untuk Pengukuran Kemiripan Makna antar Kalimat,” *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, vol. 1, no. 2, pp. 47–57, Dec. 2019, doi: 10.28926/ilkomnika.v1i2.15.
- [12] A. Millah *et al.*, “Perbandingan Penggunaan Algoritma Cosinus dan Wu Palmer untuk Mencari Kemiripan Kata dalam Plagiarism Checker,” 2017.
- [13] I. W. S. Wicaksana, “Membandingkan Pendekatan Latent Semantic terhadap WordNet untuk Semantic Similarity,” 2006. [Online]. Available: <http://lsa.colorado.edu/>.

- [14] O. Chathalr and E. Cljatli, "Query Space Reduction in Information Retrieval D ublI F c it y I TNIVERSITY," 1997.
- [15] L. A. Mullen, K. Benoit, O. Keyes, D. Selivanov, and J. Arnold, "Fast, Consistent Tokenization of Natural Language Text," *J Open Source Softw*, vol. 3, no. 23, p. 655, Mar. 2018, doi: 10.21105/joss.00655.
- [16] J. Kaur and P. Kaur Buttar, "A Systematic Review on Stopword Removal Algorithms," 2018, [Online]. Available: <http://www.ijfrcsce.org>
- [17] S. A. Kristiani and E. Rainarli, "KLASIFIKASI DOKUMEN MENGGUNAKAN SUPPORT VECTOR MACHINE DAN MUTUAL INFORMATION," Universitas Komputer Indonesia, 2019.
- [18] K. Fergus, "Query space reduction in information retrieval," 1997.
- [19] D. Rodiah, K. J. Miraswan, J. Kurniati, D. Irawan, and V. T. Ardani, "Pengembangan Representasi Pengetahuan Ontologi Domain Bidang Ilmu Informatika," *Jurnal PROCESSOR*, vol. 19, no. 2, Oct. 2024, doi: 10.33998/processor.2024.19.2.1905.
- [20] A. M. Sinaga, R. J. Sipahutar, and D. I. P. Hutasoit, "Penerapan Ontology Web Language pada Domain Ulos Batak Toba," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 4, p. 493, Oct. 2018, doi: 10.25126/jtiik.201854903.
- [21] F. F. Yusron and A. Komarudin, "Chatbot Informasi Penerimaan Mahasiswa Baru Menggunakan Metode FastText dan LSTM," *Journal of Applied Computer Science and Technology(JACOST)*, vol. 5, 2024, doi: 10.52158/jacost.648.
- [22] Y. Caterina, M. A. Yaqin, and S. Zaman, "Pengukuran Kemiripan Makna Kalimat dalam Bahasa Indonesia Menggunakan Metode Path," *Fountain of Informatics Journal*, vol. 6, no. 2, p. 45, Mar. 2021, doi: 10.21111/fij.v6i2.4844.
- [23] M. Grandini, E. Bagli, and G. Visani, "Metrics for Multi-Class Classification: an Overview," Aug. 2020, [Online]. Available: <http://arxiv.org/abs/2008.05756>
- [24] T. B. Sasongko, "Komparasi dan Analisis Kinerja Model Algoritma SVM dan PSO-SVM (Studi Kasus Klasifikasi Jalur Minat SMA)," 2016.
- [25] M. Luthfi Bangun Permadi and R. Gumilang, "PENERAPAN ALGORITMA CNN (CONVOLUTIONAL NEURAL NETWORK) UNTUK DETEKSI DAN KLASIFIKASI TARGET MILITER BERDASARKAN CITRA SATELIT," *Jurnal Sosial dan Teknologi (Sostech)*, Feb. 2024.