Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 921-932
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5759

Pemetaan Topik Tugas Akhir Program Studi Ilmu Komputer Menggunakan Algoritma Latent Dirichlet Allocation

Roma Gabe Dalimunthe, Raissa Amanda Putri*

Sains dan Teknologi, Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
Email: [1]romagabedalimunthe@email.com, [2]raissa.ap@uinsu.ac.id
Email Penulis Korespondensi: raissa.ap@uinsu.ac.id
Submitted: 08/08/2024; Accepted: 15/08/2024; Published: 16/08/2024

Abstrak−Penelitian ini berfokus pada pemetaan topik tugas akhir mahasiswa dalam Program Studi Ilmu Komputer di Universitas Islam Negeri Sumatera Utara (UINSU) dengan menggunakan algoritma Latent Dirichlet Allocation (LDA). Latar belakang penelitian ini berangkat dari kebutuhan untuk memahami perkembangan dan tren riset dalam kumpulan tugas akhir yang diserahkan, yang dapat memberikan gambaran tentang kecenderungan akademik dan area penelitian yang sedang berkembang. Namun, pengelompokan manual topik-topik ini sering kali menjadi tantangan karena volume data yang besar dan kompleksitas isi. Algoritma Latent Dirichlet Allocation (LDA) menawarkan solusi dengan kemampuannya untuk mengidentifikasi struktur topik tersembunyi dalam dokumen teks secara otomatis. Tujuan dari penelitian ini adalah untuk mengungkap tema-tema dominan serta pola-pola topik yang muncul dalam kumpulan tugas akhir mahasiswa, sehingga dapat memberikan wawasan lebih mendalam mengenai area fokus penelitian. Metodologi penelitian mencakup pengumpulan data dari berbagai tugas akhir, preprocessing data untuk mengurangi noise dan redundansi, serta penerapan algoritma LDA untuk ekstraksi topik. Hasil penelitian menunjukkan bahwa algoritma LDA efektif dalam pemetaan topik judul tugas akhir mahasiswa di UINSU. Dengan menggunakan 1000 iterasi proses LDA pada 774 judul tugas akhir, ditemukan bahwa pembagian topik yang paling optimal adalah 7 topik dengan coherence score sebesar 0.4011. Topik-topik ini divisualisasikan melalui word cloud dan daftar kata, yang memudahkan pemahaman dan interpretasi tematik. Kesimpulan ini diharapkan dapat memberikan wawasan yang berguna tentang tren riset mahasiswa, memfasilitasi penilaian kualitas dan relevansi topik, serta mendukung pengembangan kurikulum akademik yang lebih baik di institusi pendidikan tinggi.

Kata Kunci: LDA; Pemetaan Topic; Tugas Akhir; Ilmu Komputer; UINSU

Abstract−This research focuses on mapping students' final assignment topics in the Computer Science Study Program at the North Sumatra State Islamic University (UINSU) using the Latent Dirichlet Allocation (LDA) algorithm. The background to this research stems from the need to understand research developments and trends in the collection of submitted final assignments, which can provide an overview of academic trends and developing research areas. However, manual clustering of these topics is often a challenge due to the large data volume and complexity of the content. The Latent Dirichlet Allocation (LDA) algorithm offers a solution with its ability to automatically identify hidden topic structures in text documents. The aim of this research is to reveal dominant themes and topic patterns that appear in students' final assignments, so as to provide deeper insight into the research focus area. The research methodology includes collecting data from various final projects, preprocessing the data to reduce noise and redundancy, and applying the LDA algorithm for topic extraction. The research results show that the LDA algorithm is effective in mapping the topics of students' final assignment titles at UINSU. By using 1000 iterations of the LDA process on 774 final assignment titles, it was found that the most optimal topic division was 7 topics with a coherence score of 0.4011. These topics are visualized through word clouds and word lists, which facilitate understanding and thematic interpretation. It is hoped that these conclusions will provide useful insights into student research trends, facilitate assessment of the quality and relevance of topics, and support the development of better academic curricula in higher education institutions.

Keywords: LDA; Topic Mapping; Thesis; Computer Science; UINSU

1. PENDAHULUAN

Mahasiswa di akhir masa studinya diberikan tanggung jawab untuk melakukan penelitian yang biasa disebut skripsi. Skripsi atau tugas akhir adalah karya ilmiah dari hasil penelitian mahasiswa program sarjana sebagai salah satu syarat mendapatkan gelar sarjana [1]. Tugas akhir adalah salah satu syarat yang harus ditempuh oleh mahasiswa disebuah perguruan tinggi untuk dapat menyelesaikan studi menjadi seorang sarjana [2]. Dalam penyusunan skripsi dibutuhkan sumber yang banyak. Dengan membaca mereka dapat mengetahui sesuatu dan menambah pengetahuan tentang apa yang akan mereka kerjakan [3]. Biasanya penelitian yang dilakukan berhubungan dengan program studi yang diambilnya dalam perkuliahan. Dalam pengerjaan tugas akhir ini melalui beberapa proses dan tahapan yang harus dilalui sebelum menghasilkan sebuah penelitian yang dapat dipertanggung jawabkan. Ilmu Komputer adalah salah satu program studi jenjang strata-1 (S1) di UINSU. Berdasarkan data mahasiswa masuk dan mahasiswa keluar setiap 4 tahun pada program studi tersebut terdapat ketidakseimbangan jumlah, yang artinya banyak mahasiswa yang tidak menyelesaikan studinya dengan tepat waktu. Setelah melakukan wawancara dengan beberapa mahasiswa yang kelulusannya tidak tepat waktu, diperoleh kesimpulan bahwa sebagian besar mahasiswa terkendala pada tugas akhir (skripsi). Salah satu masalah yang paling umum adalah kesulitan untuk menentukan judul atau topik tugas akhir yang ingin dilakukan, terlebih karena semakin banyaknya tugas akhir yang dihasilkan setiap tahun, membuat semakin sulit untuk melacak dan memahami tren penelitian yang berkembang. Hal ini dapat menjadi kendala bagi mahasiswa baru dalam memilih topik penelitian sehingga mengakibatkan timeline pembuatan tugas akhir terganggu.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 921-932
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5759

Untuk mengatasi permasalahan tersebut, dibutuhkan topic modeling pada pengambilan judul tugas akhir mahasiswa program studi ilmu komputer. *Topic modeling* yang dilkukan diharapkan dapat membantu dalam melakukan identifikasi dan menghasilkan tema-tema penting dalam koleksi data tekstual judul tugas akhir mahasiswa program studi Ilmu Komputer yang tidak terstruktur dan sulit diidentifikasikan oleh manusia secara manual untuk mendapatkan tema tersembunyi dan menemukan topik dari teks yang jumlahnya besar. Dengan lebih mengenali topik-topik tugas akhir berdasarkan referensi yang ada pada repository, mahasiswa dapat lebih terarah dalam menentukan tema penelitian yang akan dilakukan karena sudah memiliki gambaran referensi. Topik yang tepat akan memudahkan mahasiswa dalam mengerjakan tugas akhir dan dengan pemilihan topik yang tepat diharapkan dapat membantu mahasiswa dalam menyelesaikan tugas akhir.

Data Mining merupakan sebuah teknologi yang banyak dimanfaatkan untuk mendapatkan informasi yang penting dari sebuah data *warehouse*. *Text Mining* dapat di definisikan secara luas sebagai proses intensif pengetahuan dimana pengguna berinteraksi dengan kumpulan dokumen dari waktu ke waktu dengan menggunakan seperangkat alat analisis. Penambangan teks berusaha untuk mengekstrak informasi yang berguna dari sumber data melalui identifikasi dan eksplorasi pola yang menarik [4]. Salah satu penggunaan *data mining* adalah pemodelan. Pemodelan topik merupakan algoritma untuk mendapatkan topik-topik yang tidak terlihat melalui susunan kata di dalam dokumen yang tidak berstruktur. Pemodelan Topik berfokus untuk mendapatkan topik dalam kumpulan data dokumen.

Pemodelan topik termasuk pada proses pembelajaran mesin yang digunakan untuk menemukan topik-topik dalam kumpulan dokumen teks. Metode pemodelan topik dapat digunakan untuk berbagai macam kumpulan dokumen teks, termasuk kumpulan dokumen teks hasil penelitian tugas akhir. Langkah otomatisasi bisa dilakukan dengan pendekatan *topic modeling*. *Topic modeling* merupakan suatu pendekatan untuk menganalisis kumpulan dokumen berbentuk teks dan mengelompok kan menjadi beberapa topik. Metode yang digunakan untuk melakukan pendekatan topic modeling bermacam-macam, yakni *Latent Semantic Analysis* (LSA), *Probabilistic Latent Sematic Analysis* (pLSA), *Latent Dirichlet Allocation* (LDA), dan lain-lain [5].

Latent Dirichlet Allocation (LDA) merupakan salah satu teknik yang terdapat dalam metode Pemodelan Topik. Kemampuan dasar dari LDA adalah dimana sebuah dokumen mengandung bermacam topik. LDA merupakan bentuk pemodelan statistik dari koleksi dokumen untuk menemukan intuisi ini [6]. Penelitian sebelumnya yang melakukan topic modelling, yakni pada penelitian yang dilakukan oleh A. I. Alfanzar and I. S. Rozas [7] yang membuat pemodelan topik skripsi pada suatu program studi menggunakan teknik *Latent Dirichlet Allocation* (LDA) karena LDA dianggap mampu meringkas, mengklusterkan, menghubungkan, dan memproses data teks yang sangat besar. Dengan dilakukan banyak percobaan, hasil cluster tersebut telah diverifikasi oleh pihak stakeholder Program studi Sastra Inggris (UINSA) bahwa kata-kata yang ada pada topik cluster sesuai dengan menurut konsentrasi pada Program studi Sastra Inggris (UINSA). Selanjutnya penelitian yang dilakukan oleh Setijohatmo [8] yang berjudul Analisis Metoda Latent Dirichlet Allocation untuk Klasifikasi Dokumen Laporan Tugas Akhir Berdasarkan Pemodelan Topik, Hasil probabilitas sebuah kata dalam metode LDA dipengaruhi oleh dua faktor, yaitu jumlah topik dan jumlah dokumen. Jumlah topik yang terlalu banyak dapat menyebabkan hasil probabilitas kata menjadi tidak akurat, sedangkan jumlah dokumen yang terlalu sedikit dapat menyebabkan hasil probabilitas kata menjadi tidak stabil. Penelitian ini membandingkan nilai koherensi dari 4 topik. Penelitian yang akan dilakukan akan membandingkan nilai koherensi dari 10 topik. Kemudian penelitian yang dilakukan oleh Widodo [9] yang berjudul Analisis Tren Konten Pada Vtuber Indonesia Menggunakan Latent Dirichlet Allocation, Topik yang dihasilkan dari metode LDA tersebut kebanyakan menyebut Hololive karena vtuber teratas tersebut berasal dari agensi vtuber asal Jepang dengan nama yang sama. Topik dari konten yang ditayangkan jika diurut maka pertama topik yang populer mengenai gim minecraft dilanjutkan dengan reading donasi. Penelitian ini membandingkan nilai koherensi dari 5 topik. Penelitian yang akan dilakukan akan membandingkan nilai koherensi dari 10 topik. Selanjutnya penelitian dari Gustiara [10] yang berjudul Implementasi Latent Dirichlet Allocation Terhadap Data Kasus Tindak Pidana, kesimpulan dari penelitian ini adalah Jumlah kasus tindak pidana di Pengadilan Negeri Yogyakarta meningkat dari tahun ke tahun, dengan klasifikasi perkara tertinggi adalah narkotika, pencurian, dan kesehatan. Hukuman yang paling banyak dijatuhkan adalah penjara kurang dari satu tahun tanpa denda. Pada penelitian ini Preprocessing tidak menyertakan tahapan konversi slangwords. Penelitian yang akan dilakukan menyertakan proses konversi slangwords

Berdasarkan latar belakang masalah yang telah dipaparkan sebelumnya, maka dari itu penelitian ini akan menggunakan metode LDA dalam pembuatan *topic modelling*. Adapun judul penelitian yang diangkat adalah "Pemetaan Topik Tugas Akhir Menggunakan Algoritma *Latent Dirichlet Allocation*". Penelitian ini akan terkhusus memetakan topik-topik tugas akhir dari mahasiswa jurusan Ilmu Komputer di Universitas Islam Negeri Sumatera Utara

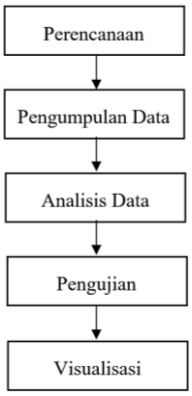
Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 921-932
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5759

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan jenis penelitian dengan kategori non-implementatif *analytic* dimana penggalian informasi diambil dari berbagai jenis kejadian yang kemudian dilakukan identifikasi dari setiap elemen-elemen yang penting untuk selanjutnya digali informasinya lebih dalam serta pengambilan keputusan untuk penelitian lebih lanjut [11]. Implementasi pengelompokan topik judul tugas akhir di Program Studi Ilmu Komputer UINSU dapat dilakukan dengan menerapkan cara kerja *text mining* menggunakan algoritma LDA (*Latent Dirichlet Allocation*).

Data yang digunakan untuk implementasi pengelompokan topik judul tugas akhir bersumber dari Prodi Studi Ilmu Komputer di UINSU. Data ini berisi informasi tentang seluruh tugas akhir sarjana yang telah diselesaikan oleh mahasiswa Program Studi Ilmu Komputer UINSU. Gambar 1 dibawah ini merupakan diagram alur proses penelitian yang akan dilakukan



Gambar 1. Kerangka Penelitian

a. Perencanaan
   Penelitian yang akan dilakukan berfokus untuk menganalisis topik tugas akhir pada program studi ilmu komputer di UINSU dengan mengumpulkan seluruh judul tugas akhir mahasiswa yang sudah diterima, baik dari mahasiswa yang sudah menyelesaikan perkuliahan maupun yang masih dalam proses penyusunan tugas akhir. Data judul tugas akhir diperoleh secara langsung dari pihak program studi terkait dan dikumpulkan dalam file .xlsx untuk kemudian diolah menggunakan bahasa pemrograman *python*.
b. Pengumpulan Data
   Proses pengumpulan data dengan teknik studi literatur dilakukan untuk mengungkap berbagai teori yang relevan dengan permasalahan yang sedang diteliti sebagai bahan rujukan dalam pembahasan hasil penelitian. Kegiatan studi literatur yang dilakukan adalah dengan metode pengumpulan data pustaka, membaca dan mencatat serta mengelolah bahan penelitian. Data yang akan digunakan pada penelitian ini bersumber dari prodi Program Studi Ilmu Komputer
c. Analisis Data
   Pada penelitian ini penulis akan menganalisis data judul tugas akhir mahasiswa program studi Ilmu Komputer di Universitas Islam Negeri Sumatera Utara sejak tahun 2019 hingga 2023. Data yang akan dijadikan bahan dalam penelitian ini berjumlah 774 judul tugas akhir, dimana keseluruhan judul tugas akhir ini ada judul-judul tugas akhir dari mahasiswa ilmu komputer di UINSU yang minimal sudah melakukan seminar proposal.
d. Pengujian
   Pengujian dalam penelitian ini terdiri dari beberapa tahapan yang dirancang untuk mengevaluasi efektivitas dan akurasi algoritma Latent Dirichlet Allocation (LDA) dalam pemetaan topik tugas akhir mahasiswa. Tahapan-tahapan tersebut adalah persiapan data, *preprocessing data*, penerapan algoritma LDA dan evaluasi hasil pemetaan.
e. Visualisasi
   Setelah melakukan pengujian penelitian, hasil dari proses pengujian yang berupa topik-topik tugas akhir akan divisualisasikan dalam bentuk *wordcloud*.

2.3 *Preprocessing* Data

Teknik Setelah data dikumpulkan, yang merupakan unstructured data (data tidak terstruktur), dataset tersebut harus menjalani proses text preprocessing sebelum dianalisis lebih lanjut. Proses ini bertujuan untuk membersihkan dan mengatasi data yang berisik agar hasil analisis dapat diperoleh secara optimal [12]. Tahapan dalam text preprocessing meliputi *case folding, filtering, tokenizing, dan stopword removal*. Setelah data dibersihkan, dilakukan pelabelan manual [13]. Diagram alur text *preprocessing* dapat dilihat pada Gambar 2.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 921-932
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5759

Gambar 2. *Preprocessing* Data [14]

Tahap 1: *Case folding*, tahapan ini akan merubah seluruh huruf kapital pada data menjadi huruf kecil
Tahap 2: *Filtering*, tahapan ini merupakan tahapan penghapusan karakter ataupun kata yang tidak mengandung informasi. Adapun karakter atau kata yang dihapus adalahseperti tanda baca, emoji, angka dan karakter non-huruf lainnya
Tahap 3: *Tokenizing*, tahapan ini merupakan tahapan pemecahan kalimat menjadi kata per kata. Sentimen utuh akan dipisah kata per kata dan setiap kata dibungkus dengan tanda kutip (‘”)
Tahap 4: Konversi *slangword*, *slangword* merupakan istilah untuk bahasa gaul yang umum digunakan, seperti singkatan. Pada tahap konversi *slangword* semua singkatan akan diubah menjadi bentuk awalnya
Tahap 5: *Stemming*, tahap ini merupakan tahap pengubahan seluruh kata menjadi kata dasar. Proses *stemming* akan menghapus seluruh imbuhan yang ada pada kata.

2.4 Vektorisasi TF-IDF

Penelitian ini menggunakan algoritma pembobotan kata TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk memberikan nilai pada kata-kata dalam suatu dokumen atau tweet. Proses pra-pemrosesan teks adalah langkah krusial dalam menghasilkan model sentimen yang akurat, karena komentar teks sering kali mengandung berbagai jenis gangguan seperti kata slang, tanda baca, kata tidak baku, singkatan, emotikon, dan sejenisnya. Oleh karena itu, metode pra-pemrosesan yang tepat sangat diperlukan. Rumus untuk algoritma TF-IDF dapat dilihat pada persamaan (1) berikut [15].

$$w_{ij} = tf_{ij}\ x\ \ln\left(\frac{D+1}{df_i+1}\right) + 1 \quad (1)$$

Dengan keterangan *tf* merupakan jumlah kemunculan kata, *D* merupakan jumlah data/judul tugas akhir yang dianalisis dan *df* merupakan jumlah dokumen/data yang mengandung kata tersebut. Setelah memperoleh nilai bobot kata, langkah selanjutnya adalah menormalisasi bobot menggunakan persamaan (2) berikut [16].

$$TF - IDF_{norm}(t,d) = \frac{TF(t,d)}{\sqrt{\sum_i (TF(t,d))^{\ 2}}} \quad (2)$$

2.5 Proses Algoritma *Latent Dirichlet Allocation*

Pemodelan topik adalah metode untuk mengekstrak dan merepresentasikan konteks dari kata-kata dengan menggunakan teknik komputasi statistik pada kumpulan teks yang besar. Tujuan dari pemodelan topik adalah untuk mengidentifikasi topik-topik serta kata-kata yang terdapat dalam korpus tersebut. Salah satu metode yang dapat digunakan dalam pemodelan topik adalah pengelompokan berdasarkan kemiripan data, seperti *Latent Dirichlet Allocation* (LDA). Teknik yang digunakan pada algoritma LDA bernama *Collapsed Gibbs Sampling.* Teknik ini digunakan mengestimasi probabilitas topik terhadap kata dan dokumen terhadap topik [17].

Nilai koherensi (*coherence score*) dalam konteks *Latent Dirichlet Allocation* (LDA) adalah ukuran yang digunakan untuk mengevaluasi kualitas topik yang dihasilkan oleh model. LDA adalah teknik pemodelan topik yang digunakan dalam analisis teks untuk mengidentifikasi struktur topik dalam kumpulan dokumen. Nilai koherensi membantu menilai seberapa baik topik-topik yang dihasilkan oleh model LDA dapat diinterpretasikan secara semantik [18] .

Prinsip dasar LDA adalah setiap dokumen dianggap sebagai kombinasi dari topik-topik tersembunyi yang belum diketahui, di mana setiap topik terdiri dari distribusi kata-kata tertentu [19]. Setelah tahap pra-pemrosesan, hasil data digunakan untuk membuat model topik dengan LDA. Sebelum proses pemodelan, dilakukan pembentukan dictionary dan corpus. Dictionary berfungsi sebagai data yang mengandung himpunan kata unik beserta nomor indeksnya, sedangkan corpus adalah data berbentuk bag-of-words yang akan digunakan dalam model. Proses pemodelan topik dengan LDA melibatkan penentuan jumlah topik sebagai kelompok klaster kata dan jumlah passes yang merupakan jumlah iterasi dalam proses pelatihan model. Validasi hasil pemodelan topik dilakukan dengan menggunakan topic coherence untuk memastikan bahwa model yang terbentuk memiliki nilai probabilitas tertinggi pada dokumen yang dihasilkan. Hasil dari pemodelan topik LDA mencerminkan jumlah topik yang sesuai berdasarkan nilai topic coherence setiap model[20].

3. HASIL DAN PEMBAHASAN

3.1 Analisis Data

Pada penelitian ini penulis akan menganalisis data judul tugas akhir mahasiswa program studi Ilmu Komputer di Universitas Islam Negeri Sumatera Utara sejak tahun 2019 hingga 2023. Tahap pertama pada penelitian adalah

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 921-932
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5759

melakukan pengumpulan data. Proses pengumpulan data judul tugas akhir dilakukan dengan cara meminta data yang dibutuhkan secara langsung ke pihak program studi Ilmu Komputer di Universitas Islam Negeri Sumatera Utara. Data yang akan dijadikan bahan dalam penelitian ini berjumlah 774 judul tugas akhir, dimana keseluruhan judul tugas akhir ini ada judul-judul tugas akhir dari mahasiswa yang sudah melakukan seminar proposal. Penelitian ini bertujuan untuk mengidentifikasi topik utama yang muncul dalam tugas akhir dan memahami distribusi topik-topik tersebut dalam berbagai dokumen. Berikut adalah gambaran dari data yang akan digunakan.

Tabel 1. Representase Data

No	Nama	Judul Penelitian
1	Indri Gusmita Br Rambe	Pemanfaatan Smart Trash Can dengan Metode Logika Fuzzy Berbasis Nodemcu Pada Smartphone
2	Irma Yunita Nasution	Penerapan Teknologi Augmented Reality Untuk Pengenalan Gerakan Sholat Berdasarkan 4 Mazhab Menggunakan Metode Markerless Augmented Reality Berbasis Android
3	Salmah Simanjuntak	Sistem Pendukung Keputusan Pemilihan Siswa Kelas Unggulan Menggunakan Metode Simple Additive Weighting (SAW) (Studi Kasus pada MTS Negeri 2 Medan)
4	Razzaq H.Nur Wijaya	Perbandingan Algoritma Dijkstra dan Algoritma Steepest Ascent Hill Climbing dalam Menentukan Rute Terpendek (Studi Kasus: Antar Lokasi Objek Wisata di Kabupaten Humbang Hasundutan)
5	Lailan Sofinah Harahap	KLASIFIKASI TANAMAN BUGENVIL BERDASARKAN TEKSTUR DAUN MENGGUNAKAN GRAY LEVEL CO-OCCURRENCE MATRIX (GLCM) DAN K-NEAREST NEIGHBOR (KNN)
6	Indah Eka Yulia Sari	Segmentasi Citra Dengan Menggunakan Metode OTSU pada Citra Naskah Arab (Studi Kasus: Museum Negeri Provinsi Sumatera Utara)
7	Yuli Kartika Siregar	Analisis Perbandingan Algoritma Contraharmonic Mean Filter dan Arithmetic Mean Filter untuk reduksi Noise Eksponensial pada Citra Digital
8	Rizky Sundari Tampubolon	Implementasi Metode Canny untuk Deteksi Tepi Pola Tulisan Arab pada Nisan Kuno Peninggalan Sejarah di Sumatera Utara (Studi Kasus: Museum Negeri Provinsi Sumatera Utara)
9	Siti Sarah Harahap	Alat Pemandu Jalan Untuk Penyandang Tunanetra Menggunakan Metode Fuzzy Berbasis Mikrokontroller
10	Lili Suriani	Penerapan Data Mining Menggunakan Algoritma Apriori pada Sistem Persediaan Produk Kosmetik
..	..	..
773	Nurainun Br Nainggolan	Algoritma K-Means Clustering Untuk Pengelompokan Penjualan Sembako Di Grosir Yuda
774	Wahyuda Pratama	Algoritma Naïve Bayes Classifier Pada Klasifikasi Review Aplikasi X (Twitter) Di Play Store

3.2 *Preprocessing* Data

Setelah data judul tugas akhir berhasil dikumpulkan, langkah pertama adalah melakukan *text preprocessing*. Proses ini penting karena dataset yang digunakan merupakan data tidak terstruktur (unstructured data). Tahapan text preprocessing yang dilakukan dalam penelitian ini meliputi: a. *Case Folding*, yaitu proses mengubah semua huruf, baik yang menggunakan huruf besar maupun kecil, menjadi huruf kecil (*lowercase*). b. *Filtering*, yaitu tahap pembersihan data dari tanda baca, simbol, dan elemen lain yang tidak diperlukan, seperti *URL*. c. *Tokenizing*, yaitu proses membagi teks menjadi kata-kata terpisah, contohnya "saya pergi ke sekolah" dipecah menjadi "saya", "pergi", "ke", "sekolah". d. Stopword Removal, yaitu tahap menghapus kata-kata sambung seperti "ke", "di", "dan", "dia", "kami", "aku", "saya". Tabel 2. menunjukkan contoh sentimen sebelum dan setelah dilakukan *text preprocessing*.

Tabel 2. *Preprocessing* Data

Nama	Judul Penelitian
Input	Pemanfaatan Smart Trash Can dengan Metode Logika Fuzzy Berbasis Nodemcu Pada Smartphone
Output	['manfaat', 'smart', 'trash', 'logika', 'fuzzy', 'bas', 'nodemcu', 'smartphone']

3.3 Vektorisasi TF-IDF

Tahap selanjutnya setelah melewati tahapan pelabelan dan pembersihan sentimen adalah tahap pembobotan TF-IDF (*Term Frequency-Inverse Document Frequency*), dimana pada tahapan ini menggunakan teknik perhitungan

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 921-932
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5759

setiap pembobotan kata (*term*) yang ada didalam data dokumen dihitung dari setiap kata dan setiap kata kemudian akan dikalikan idf. Berikut merupakan sampel perhitungan nilai TF dan nilai DF dari 3 buah data.

Tabel 3. Sampel Data

Judul Tugas Akhir
['analisis', 'sentimen', 'opini', 'masyarakat', 'indonesia', 'covid', 'media', 'sosial', 'twitter', 'naive', 'bayes']
['analisis', 'sentimen', 'covid', 'dasar', 'opini', 'masyarakat', 'indonesia', 'media', 'sosial', 'twitter']
['analisis', 'sentimen', 'masyarakat', 'testimoni', 'wiralaba', 'minum', 'naive', 'bayes']

Dengan menggunakan persamaan (1), maka akan diperoleh nilai bobot dari setiap kata. Berikut adalah contoh perhitungan untuk data pertama

$$IDF = \ln\left(\frac{5+1}{3+1}\right) + 1 = \ln(1.5) + 1 = 0.405 + 1 = 1.405$$

Dengan menggunakan cara perhitungan yang sama, maka akan diperoleh nilai dari TF-IDF dari seluruh data seperti pada tabel 4 berikut.

Tabel 4. Hasil vektorisasi TF-IDF

Term	TF-IDF		
	D1	D2	D3
analisis	1.405	1.405	1.405
sentimen	1.405	1.405	1.405
opini	1.693	1.693	0
masyarakat	1.405	1.405	1.405
indonesia	1.693	1.693	0
covid	1.693	1.693	0
media	1.693	1.693	0
sosial	1.693	1.693	0
twitter	1.693	1.693	0
naive	1.693	0	1.693
bayes	1.693	0	1.693
dasar	0	2.098	0
testimoni	0	0	2.098
wiralaba	0	0	2.098
minum	0	0	2.098

Selanjutnya nilai TF-IDF dinormalisasikan untuk menyamakan interval dari setiap data. Dengan menggunakan persamaan (2), maka diperoleh nilai akhir dari proses vektorisasi TF-IDF, nilai tersebut selanjutnya akan diproses menggunakan algoritma *Latent Dirichlet Allocation* (LDA). Berikut adalah contoh perhitungan normalisasi untuk data pertama

$$Term\ 1 - \text{D1} = \frac{1.405}{9.012} = 0.155$$

Dengan menggunakan cara yang sama maka diterapkan perhitungan seperti diatas pada setiap data. Berikut adalah hasil normalisasi dari vektorisasi TF-IDF dari keseluruhan data.

Tabel 5. Normalisasi hasil vektorisasi TF-IDF

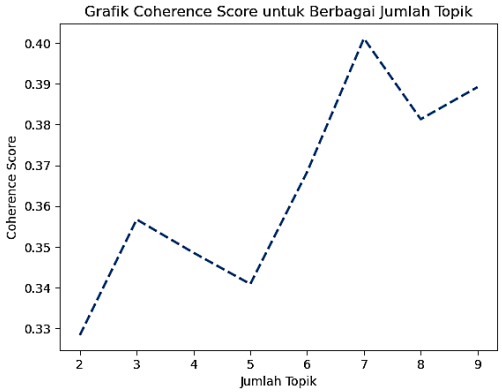
Term	TF-IDF		
	D1	D2	D3
analisis	0.155	0.155	0.155
sentimen	0.155	0.155	0.155
opini	0.187	0.187	0
masyarakat	0.155	0.155	0.155
indonesia	0.187	0.187	0
covid	0.187	0.187	0
media	0.187	0.187	0
sosial	0.187	0.187	0
twitter	0.187	0.187	0
naive	0.187	0	0.187
bayes	0.187	0	0.187
dasar	0	0.232	0
testimoni	0	0	0.232
wiralaba	0	0	0.232

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 921-932
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5759

Term	TF-IDF		
	D1	D2	D3
minum	0	0	0.232

3.4 Proses Algoritma *Latent Dirichlet Allocation*

Untuk menentukan hasil pemodelan, salah satu cara adalah dengan memeriksa visualisasi grafik coherence score. Coherence score adalah ukuran yang digunakan untuk mengevaluasi hasil Topic Modeling, di mana semakin tinggi nilai coherence score, semakin baik model yang dihasilkan. Grafik coherence score biasanya menunjukkan fluktuasi, dengan nilai yang naik turun. Peneliti akan membandingkan beberapa grafik, yang menunjukkan variasi topik mulai dari 2 hingga 9, sebagaimana dapat dilihat pada gambar 3 berikut.



Gambar 3. Grafik *cohoerence score*

Berdasarkan gambar 3 diatas didapatkan informasi bahwa pada grafik *coherence score* diatas memiliki pola yang berulang dan semakin banyak limit topiknya maka semakin tinggi nilai *coherence* yang dihasilkan. Berdasarkan grafik tersebut maka peneliti memutuskan untuk menggunakan topik sebanyak 7 karna memiliki nilai koherensi tertinggi, dengan begitu jumlah topik tersebut yang akan menjadi acuan untuk membuat model selanjutnya. Dengan menggunakan bahasa pemrograman pyhton dalam menerapkan algoritma LDA dengan 1000 kali iterasi maka diperoleh *coherence score* dari setiap jumlah topik yagng diuji. Tabel 6 berikut menampilkan nilai *coherence score* yang dihasilkan dari bahasa pemrograman python:

Tabel 6. *Preprocessing* Data

Jumlah Topik	*Coherence Score*
2	0.324
3	0.3283
4	0.3485
5	0.3408
6	0.3683
7	0.4011
8	0.3813
9	0.3891

3.5 Visualisasi

Dengan menggunakan python, data dianalisi dan memoeroleh hasil bahwa terdapat 7 topik penelitian populer pada kumpulan data judul mahasiswa di program studi ilmu komputer di UINSU yang diteliti. Berikut adalah visualisasi dari hasil pemetaan topik.

a. Topik 1
   Berikut adalah beberapa kata yang mewakili topik 1 dari hasil penelitian. Pada topik 1 yang ditemukan memiliki 5 kata sebagai kata kunci untuk mewakilkan topik tersebut, yaitu fuzzy, putus, dukung, logika dan tentu.

```
Topik 1:
----------------------
        0.048*"fuzzy"
        0.030*"putus"
        0.027*"dukung"
        0.025*"logika"
        0.024*"tentu"
======================
```

Gambar 4. Perwakilan kata pada Topik 1

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 921-932
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5759

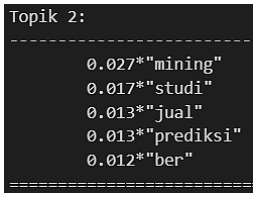
Selain perwakilan kata, hasil analisis LDA juga menghasilkan visualisasi kata pada topik dalam bentuk *wordcloud*. Berikut adalah visualiasi kata dalam bentuk *wordcloud* untuk topik 1.

Gambar 5. *Wordcloud* topik 1

Berdasarkan hasil analisis LDA pada topik 1, dapat dilihat berdasarkan gambar 4. dan gambar 5 terdapat beberapa kata seperti “fuzzy”, “putus”, “dukung”, “logika” dan “tentu”, dari beberapa perwakilan kata yang ada pada *wordcloud* dapat disimpulkan bahwa topik 1 mengarah pada tema sistem pendukung keputusan.

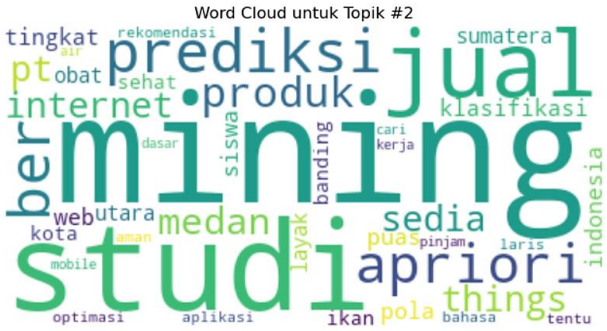
b. Topik 2

Berikut adalah beberapa kata yang mewakili topik 2 hasil penelitian. Pada topik 2 yang ditemukan memiliki 5 kata sebagai kata kunci untuk mewakilkan topik tersebut, yaitu mining, studi, jual, prediksi dan ber



Gambar 6. Perwakilan kata pada Topik 2

Selain perwakilan kata, hasil analisis LDA juga menghasilkan visualisasi kata pada topik dalam bentuk *wordcloud*. Berikut adalah visualiasi kata dalam bentuk *wordcloud* untuk topik 2.

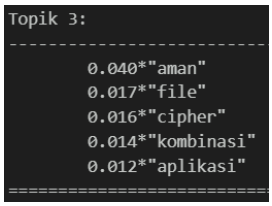


Gambar 7. *Wordcloud* topik 2

Berdasarkan hasil analisis LDA pada topik 2, dapat dilihat pada gambar 6. Dan gambar 7. terdapat beberapa kata seperti “*mining*”, “studi”, “prediksi”, “jual” dan “prediksi”, dari beberapa perwakilan kata yang ada pada *wordcloud* dapat disimpulkan bahwa topik 2 mengarah pada tema data mining seperti prediksi.

c. Topik 3

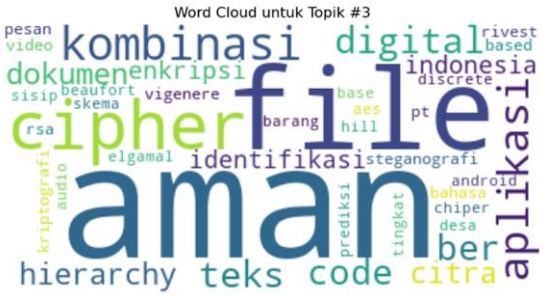
Berikut adalah beberapa kata yang mewakili topik 3 hasil penelitian. Pada topik 3 yang ditemukan memiliki 5 kata sebagai kata kunci untuk mewakilkan topik tersebut, yaitu aman, file, chiper, kombinasi dan aplikasi



Gambar 8. Perwakilan kata pada Topik 3

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 921-932
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5759

Selain perwakilan kata, hasil analisis LDA juga menghasilkan visualisasi kata pada topik dalam bentuk *wordcloud*. Berikut adalah visualiasi kata dalam bentuk *wordcloud* untuk topik 3.



Gambar 9. *Wordcloud* topik 3

Berdasarkan hasil analisis LDA pada topik 3, berdasarkan gambar 8. Dan gambar 9. dapat dilihat bahwa terdapat beberapa kata seperti “aman”, “file”, “chiper”, “kombinasi” dan “aplikasi”, dari beberapa perwakilan kata yang ada pada *wordcloud* dapat disimpulkan bahwa topik 3 mengarah pada tema keamanan data atau kriptografi.

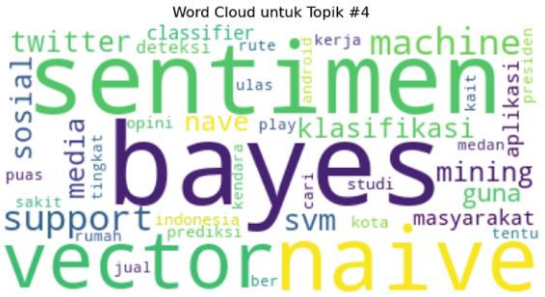
d. Topik 4

Berikut adalah beberapa kata yang mewakili topik 4 hasil penelitian. Pada topik 4 yang ditemukan memiliki 5 kata sebagai kata kunci untuk mewakilkan topik tersebut, yaitu bayes, sentimen, naive, vector dan support

```
Topik 4:
--------------------------
        0.055*"bayes"
        0.042*"sentimen"
        0.038*"naive"
        0.031*"vector"
        0.030*"support"
==========================
```

Gambar 10. Perwakilan kata pada Topik 4

Selain perwakilan kata, hasil analisis LDA juga menghasilkan visualisasi kata pada topik dalam bentuk *wordcloud*. Berikut adalah visualiasi kata dalam bentuk *wordcloud* untuk topik 4.



Gambar 11. *Wordcloud* topik 4

Berdasarkan hasil analisis LDA pada topik 4, dapat dilihat pada gambar 10. Dan gambar 11. bahwa terdapat beberapa kata seperti “bayes”, “sentimen”, “naive”, “vector” dan “support”, dari beberapa perwakilan kata yang ada pada *wordcloud* dapat disimpulkan bahwa topik 4 mengarah pada tema analisis sentimen.

e. Topik 5

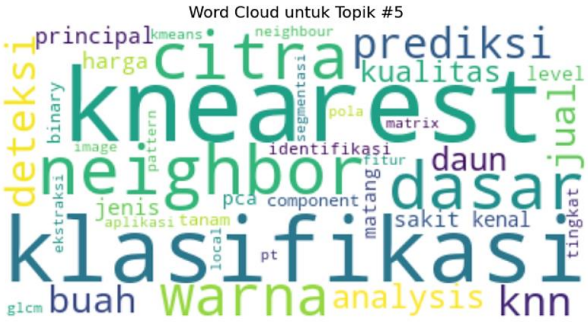
Berikut adalah beberapa kata yang mewakili topik 5 hasil penelitian. Pada topik 5 yang ditemukan memiliki 5 kata sebagai kata kunci untuk mewakilkan topik tersebut, yaitu knearest, klasifikasi, neighbour, dasar dan citra

```
Topik 5:
----------------------------
        0.025*"knearest"
        0.023*"klasifikasi"
        0.022*"neighbor"
        0.021*"dasar"
        0.021*"citra"
============================
```

Gambar 12. Perwakilan kata pada Topik 5

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 921-932
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5759

Selain perwakilan kata, hasil analisis LDA juga menghasilkan visualisasi kata pada topik dalam bentuk *wordcloud*. Berikut adalah visualiasi kata dalam bentuk *wordcloud* untuk topik 5.

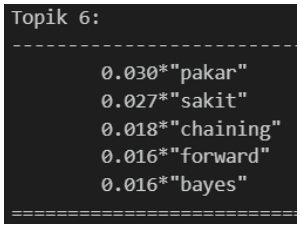


Gambar 13. *Wordcloud* topik 5

Berdasarkan hasil analisis LDA pada topik 5, dapat dilihat pada gambar 12. Dan gambar 13. bahwa terdapat beberapa kata seperti “knearest”, “klasifikasi”, “neighbour”, “dasar” dan “citra”, dari beberapa perwakilan kata yang ada pada *wordcloud* dapat disimpulkan bahwa topik 5 mengarah pada tema klasifikasi.

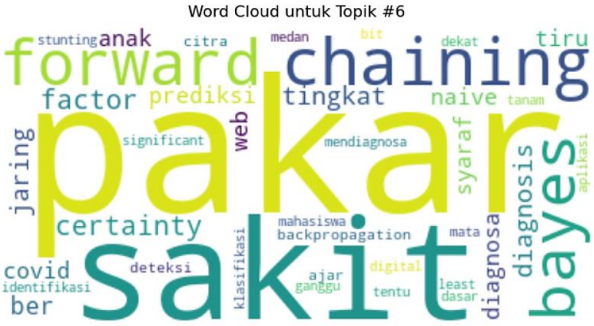
f. Topik 6

Berikut adalah beberapa kata yang mewakili topik 6 hasil penelitian. Pada topik 6 yang ditemukan memiliki 5 kata sebagai kata kunci untuk mewakilkan topik tersebut, yaitu pakar, sakit, chaining, forward dan bayes.



Gambar 14. Perwakilan kata pada Topik 6

Selain perwakilan kata, hasil analisis LDA juga menghasilkan visualisasi kata pada topik dalam bentuk *wordcloud*. Berikut adalah visualiasi kata dalam bentuk *wordcloud* untuk topik 6.

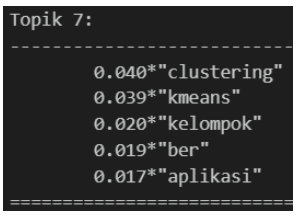


Gambar 15. *Wordcloud* topik 6

Berdasarkan hasil analisis LDA pada topik 6, dapat dilihat pada gambar 14. Dan 15. Bahwa terdapat beberapa kata seperti “pakar”, “sakit”, “chaining”, “forward” dan “bayes”, dari beberapa perwakilan kata yang ada pada *wordcloud* dapat disimpulkan bahwa topik 6 mengarah pada tema sistem pakar.

g. Topik 7

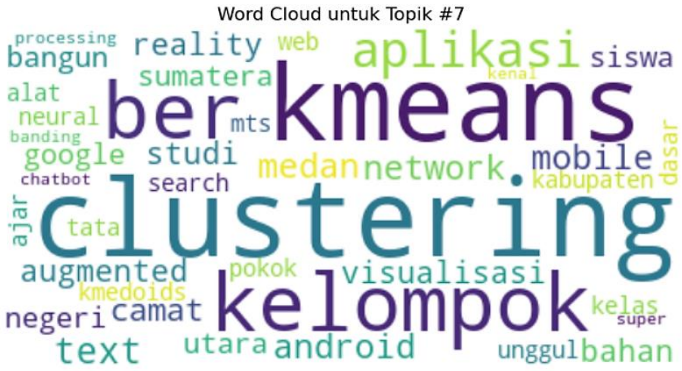
Berikut adalah beberapa kata yang mewakili topik 7 hasil penelitian. Pada topik 7 yang ditemukan memiliki 5 kata sebagai kata kunci untuk mewakilkan topik tersebut, yaitu clustering, kmeans, kelompok, ber dan aplikasi.



Gambar 16. Perwakilan kata pada Topik 7

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 921-932
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5759

Selain perwakilan kata, hasil analisis LDA juga menghasilkan visualisasi kata pada topik dalam bentuk *wordcloud*. Berikut adalah visualiasi kata dalam bentuk *wordcloud* untuk topik 7.



Gambar 17. *Wordcloud* topik 7

Berdasarkan hasil analisis LDA pada topik 7, dapat dilihat pada gambar 16. Dan 17. Bahwa terdapat beberapa kata seperti "clustering", "kmeans", "kelompok" dan "aplikasi", dari beberapa perwakilan kata yang ada pada *wordcloud* dapat disimpulkan bahwa topik 7 mengarah pada tema pengelompokan.

Setelah menetapkan tema topik berdasarkan kata-kata yang terkandung pada topik tersebut, maka dapat disimpulkan bahwa dari 774 data judul mahasiswa UINSU di program studi Ilmu Komputer dapat dipetakan menjadi 7 topik utama yaitu analisis sentimen, pengolahan citra, klasifikasi, mikrokontroller, kriptografi, pengelompokan dan sistem pendukung keputusan.

4. KESIMPULAN

Berdasarkan dari proses pemetaan topik yang telah dilakukan, maka dapat disimpulkan bahwa penggunakan dari algoritma *Latent Dirichlet Allocation* (LDA) cukup efektif dalam memetakan topik dari judul tugas akhir mahasiswa program studi Ilmu Komputer di Universitas Islam Negeri Sumatera Utara (UINSU). Dengan melakukan 1000 iterasi pada algoritma LDA, proses pemetaan yang dilakukan menemukan bahwa jumlah topik yang paling optimal yang mewakilkan 774 data judul tugas akhir berjumlah sebanyak 7 topik, dengan mendapatkan *coherence score* sebesar 0.4011. Adapun topik-topik utama pada program studi Ilmu Komputer di UINSU yang ditemukan pada hasil penelitian yang dilakukan adalah mengenai sistem pendukung keputusan, *data mining* , keamanan data atau kriptografi, analisis sentimen *atau text mining*, sistem pakar dan pengelompokan atau *clustering*. Hasil topik divisualisasikan dalam bentuk *word cloud* serta daftar kata kunci yang utama yang mewakili dari setiap topik yang ada untuk membantu dalam memahami dan menganalisis tema utama dari tugas akhir yang diangkat oleh mahasiswa program studi ilmu komputer di UINSU.

REFERENCES

[1] D. A. C. Rachman, R. Goejantoro, and F. D. T. Amijaya, "Implementasi Text Mining Pengelompokkan Dokumen Skripsi Menggunakan Metode K-Means Clustering," *Eksponensial*, vol. 11, no. 2, p. 167, 2021, doi: 10.30872/eksponensial.v11i2.660.

[2] Y. Yulindawati, S. Lailiyah, A. Yusnita, and A. Hafifah, "Rekomendasi Pemilihan Judul Tugas Akhir Menggunakan Metode Naïve Bayes," *J. Inf. Syst. Manag.*, vol. 5, no. 2, pp. 171–175, 2024, doi: 10.24076/joism.2024v5i2.1383.

[3] Y. Rahmi and I. Marnola, "Peningkatan Kemampuan Membaca Pemahaman Siswa Melalui Model Pembelajaran Cooperative Integrated Reading and Compotion (Circ)," *J. Basicedu*, vol. 4, no. 3, pp. 662–672, 2020, doi: 10.31004/basicedu.v4i3.406.

[4] Y. Findawati and M. Alfan Rosid, *BUKU AJAR TEXT MINING*. 2020.

[5] M. R. Firdaus, F. M. Rizki, F. M. Gaus, and I. K. Susanto, "Analisis Sentimen Dan Topic Modelling Dalam Aplikasi Ruangguru," *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 4, no. 1, p. 66, 2020, doi: 10.30645/j-sakti.v4i1.188.

[6] R. Kurniawan, I. Zufria, and R. P. Putra, *Penerapan Text Mining Pada Sistem Penyeleksian Judul Skripsi Mahasiswa Menggunakan Algoritma Latent Dirichlet Allocation Di Program Studi Ilmu Komputer Uin Sumatera Utara Medan*, vol. 4, no. 11. 2022.

[7] A. I. Alfanzar and I. S. Rozas, "Topic Modelling Skripsi Menggunakan Metode Latent," *JSiI (Jurnal Sist. Informasi)*, vol. 7, no. 1, pp. 7–13, 2020.

[8] U. T. Setijohatmo, S. Rachmat, T. Susilawati, Y. Rahman, and K. Kunci, "Analisis Metoda Latent Dirichlet Allocation untuk Klasifikasi Dokumen Laporan Tugas Akhir Berdasarkan Pemodelan Topik," *Pros. 11th Ind. Res. Work. Natl. Semin.*, pp. 402–408, 2020.

[9] Akdeas Oktanae Widodo, F. Septiadi, and Nur Aini Rakhmawati, "Analisis Tren Konten Pada Vtuber Indonesia Menggunakan Latent Dirichlet Allocation," *J. Inform. dan Rekayasa Elektron.*, vol. 6, no. 1, pp. 56–63, 2023, doi: 10.36595/jire.v6i1.718.

[10] D. Gustiara, "Implementasi Latent Dirichlet Allocation Terhadap Data Kasus Tindak Pidana," 2023, [Online].

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 921-932
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5759

Available: https://dspace.uii.ac.id/handle/123456789/45167

[11] A. A. Nabhan, B. Rahayudi, and D. E. Ratnawati, "Klasifikasi Tweets Masyarakat yang Membicarakan Layanan GoFood dan GoRide pada GoJek Dimedia Sosial Twitter Selama Masa Kenormalan Baru (New Normal) dengan Metode Naïve Bayes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 7, pp. 3018–3025, 2021, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/9454

[12] A. R. Iqbal and Y. Miftahuddin, "Implementasi SVM Untuk Deteksi Komentar Negatif Berbahasa Indonesia di Twitter," *Fti*, vol. X, no. X, 2022, [Online]. Available: https://eproceeding.itenas.ac.id/index.php/fti/article/view/966%0Ahttps://eproceeding.itenas.ac.id/index.php/fti/article/download/966/942

[13] A. H. Hasugian, M. Fakhriza, and D. Zukhoiriyah, "Analisis Sentimen Pada Review Pengguna E-Commerce Menggunakan Algoritma Naïve Bayes," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 6, no. 1, p. 98, 2023, doi: 10.53513/jsk.v6i1.7400.

[14] N. A. Pasaribu and Sriani, "The Shopee Application User Reviews Sentiment Analysis Employing Naïve Bayes Algorithm," *Int. J. Softw. Eng. Comput. Sci.*, vol. 3, no. 3, pp. 194–204, 2023, doi: 10.35870/ijsecs.v3i3.1699.

[15] A. Handayani and I. Zufria, "Analisis Sentimen Terhadap Bakal Capres RI 2024 di Twitter Menggunakan Algoritma SVM," *J. Inf. Syst. Res.*, vol. 5, no. 1, pp. 53–63, 2023, doi: 10.47065/josh.v5i1.4379.

[16] Sriani, A. H. Lubis, and L. P. A. Lubis, "Sentiment analysis on twitter about the death penalty using the support vector machine method," *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 11, no. 2, pp. 312–321, 2024, doi: 10.37373/tekno.v11i2.1096.

[17] I. F. Akbar, T. G. Laksana, A. B. Arifa, and M. R. Silalahi, "Pengelompokan Teks Berita Utama dengan Metode LDA (Latent Dirichlet Allocation) melalui Pemahaman Pemodelan Topik," pp. 475–484, 2023.

[18] E. P. Putri, "Implementasi Latent Dirichlet Allocation (Lda) Untuk Pemodelan Topik Faktor Perceraian," *Angew. Chemie Int. Ed. 6(11), 951–952.*, no. Mi, pp. 5–24, 2020.

[19] K. F. Nurdin, T. E. Sutanto, and A. Santoso, "Analisa Pemodelan Topik Berita Daring Menggunakan Semi-Supervised dan Fully Unsupervised Latent Dirichlet Allocation," *J. Pendidik. Tambusai*, vol. 7, no. 2, pp. 8043–8055, 2023, [Online]. Available: https://www.palupos.com

[20] P. S. Nautika and W. Yustanti, "Analisis Pinjaman Online Pada Sosial Media Twitter Menggunakan Latent Direchlet Allocation (LDA)," vol. 06, pp. 427–436, 2024.