Jurnal INTEK (Informatika dan Teknologi Informasi)
Volume 8 Nomor 2, November 2025
eISSN: 2620 – 4924 | pISSN: 2620 – 4843 | Halaman: 35 – 44

Klasifikasi Judul Skripsi Berbasis Naive Bayes Classifier dengan Analisis Tren Topik

Ike Yunia Pasa[1], Nur Wachid Adi Prasetya[2*]
[1]Teknologi Informasi, Universitas Muhammadiyah Purworejo, Purworejo 54111, Indonesia
[2]Teknologi Rekayasa Multimedia, Politeknik Negeri Cilacap, Cilacap 53212, Indonesia
ikeypasa@umpwr.ac.id, nwap.pnc@pnc.ac.id

ABSTRAK
Proses penentuan judul skripsi merupakan langkah awal yang sangat penting dalam penyusunan tugas akhir mahasiswa. Namun, banyak mahasiswa mengalami kesulitan dalam memilih judul yang sesuai dengan minat dan kompetensinya, serta dalam memastikan orisinalitas judul yang diajukan. Hal ini mengakibatkan kurangnya keberagaman topik dan potensi terjadinya pengulangan penelitian. Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi judul skripsi menggunakan algoritma *Naive Bayes Classifier* (NBC) dan menambahkan dimensi analisis tren topik untuk mengidentifikasi perkembangan isu penelitian dari tahun ke tahun.
Metode yang digunakan mencakup tahapan pengumpulan data, pra-pemrosesan teks (case folding, tokenizing, dan stopword removal), ekstraksi fitur menggunakan metode TF-IDF, pelatihan model NBC, serta evaluasi performa model menggunakan confusion matrix dengan metrik akurasi, precision, recall, dan F1-score. Data yang digunakan terdiri dari 81 judul skripsi mahasiswa Program Studi Teknologi Informasi Universitas Muhammadiyah Purworejo, yang dikategorikan ke dalam lima bidang: sistem informasi/website, multimedia, jaringan, IoT, dan sistem cerdas.
Hasil pengujian menunjukkan bahwa model mencapai akurasi 76,5%, precision 77,3%, recall 76,5%, dan F1-score 72,5%, yang menandakan performa model yang cukup baik. Selain itu, melalui analisis frekuensi kata (*Word Frequency Analysis*), diperoleh gambaran topik dominan pada setiap tahun, yang divisualisasikan menggunakan diagram batang. Penelitian ini diharapkan dapat membantu mahasiswa dan institusi dalam memilih judul skripsi secara lebih terarah dan berdasarkan tren yang berkembang.
Kata kunci: Judul skripsi, klasifikasi, *Naïve Bayes Classifier*, tren topik, *Word Frequency Analysis*

1. PENDAHULUAN

Hampir semua perguruan tinggi di Indonesia mewajibkan mahasiswa menyelesaikan skripsi sebagai syarat utama sebelum mereka dapat lulus dan resmi menyandang gelar sarjana [1]. Menurut Rismen (2015), skripsi merupakan tulisan ilmiah yang menjadi kewajiban mahasiswa dalam ketentuan akademik di setiap lembaga pendidikan tinggi. Mahasiswa menyusun laporan hasil riset terhadap suatu isu dalam bidang ilmu tertentu, dengan mengacu pada teori dan rumpun keilmuan yang sesuai di masing-masing perguruan tinggi [2]. Dalam menyusun skripsi, mahasiswa diharapkan memiliki dasar yang kuat dalam statistika, metodologi penelitian, strategi pembelajaran, atau evaluasi belajar, disertai pemahaman pada bidang keilmuan yang sesuai dengan jurusan yang diambil [2]. Mahasiswa yang sedang menempuh proses penyusunan skripsi dituntut untuk menyelesaikannya tepat waktu. Isi dari skripsi biasanya merupakan hasil penelitian yang mengangkat suatu fenomena, dengan tetap mengikuti aturan dan pedoman akademik yang telah ditetapkan [3].

Langkah pertama yang sangat penting dalam menulis skripsi adalah memilih topik atau judul yang sesuai. Jika topik atau judul tersebut tidak selaras dengan minat dan kemampuan mahasiswa, bukan tidak mungkin proses penulisannya akan mengalami hambatan, bahkan

Jurnal INTEK (Informatika dan Teknologi Informasi)
Volume 8 Nomor 2, November 2025
eISSN: 2620 – 4924 | pISSN: 2620 – 4843 | Halaman: 35 – 44

dapat gagal diselesaikan. Tidak sedikit mahasiswa yang kebingungan saat harus memilih topik skripsi yang benar-benar sesuai dengan kemampuan mereka. Menentukan topik yang selaras dengan keahlian memang bukan hal yang sederhana dan sering kali menjadi tantangan tersendiri di awal penyusunan skripsi. [4].

Seiring makin banyaknya mahasiswa dari tahun ke tahun, jumlah judul skripsi yang diajukan pun terus bertambah. Tidak jarang mahasiswa kesulitan memastikan apakah judul yang mereka pilih benar-benar baru atau sudah pernah digunakan sebelumnya. Situasi ini menimbulkan adanya penelitian yang sebenarnya telah dilakukan, namun kembali diteliti ulang tanpa adanya pengembangan dari penelitian terdahulu. Selain itu, dikhawatirkan beberapa karya tulis ilmiah memiliki judul serupa yang telah dibuat lebih dulu, sehingga berpotensi dianggap sebagai plagiarisme [5]. Hal ini juga akan menimbulkan permasalahan pada pihak jurusan, karena mengakibatkan keberagaman judul skripsi yang sedikit.

Salah satu pendekatan yang dapat dilakukan adalah melakukan klasifikasi judul skripsi. Melalui pendekatan klasifikasi, mahasiswa dapat lebih mudah menemukan topik skripsi yang sesuai dengan minatnya. Cara ini tidak hanya mempermudah proses pemilihan judul, tapi juga membantu agar penelitian yang mereka lakukan jadi lebih tepat sasaran dan relevan [6]. Proses klasifikasi judul skripsi dapat mempermudah mahasiswa dalam mencari referensi yang relevan, sekaligus memfasilitasi dosen untuk memberikan bimbingan sesuai kompetensi mereka. Sistem klasifikasi yang tersusun rapi juga mempercepat pencarian karya terdahulu dan membantu perguruan tinggi dalam menata data skripsi secara lebih sistematik [7]. Selain itu, dengan melakukan klasifikasi judul skripsi, mahasiswa atau jurusan dapat menangkap perubahan tren penelitian serta menemukan peluang kajian baru yang sesuai perkembangan teknologi dan kebutuhan masa depan. Kajian terhadap jalur riset para mahasiswa terdahulu juga dapat memberi gambaran bagaimana efektivitas kurikulum dalam membekali mahasiswa menghadapi dunia profesional [8].

Salah satu metode yang dapat diaplikasikan guna melakukan proses klasifikasi antara lain *Naive Bayes Classifier* (NBC). NBC adalah metode klasifikasi yang bekerja dengan memanfaatkan prinsip dasar Bayes. Dengan menggunakan analisis peluang dan data yang pernah dikumpulkan sebelumnya, metode ini membantu memprediksi kemungkinan yang akan terjadi di masa depan. Pendekatan ini juga berguna dalam mengelompokkan data agar lebih mudah dianalisis [9].

Penelitian-penelitian sebelumnya telah banyak menerapkan algoritma *Naive Bayes Classifier* (NBC) dalam klasifikasi judul skripsi atau karya ilmiah mahasiswa, khususnya dalam bidang Teknik Informatika. Nurdin et al. (2021) berhasil mengklasifikasikan 170 judul karya ilmiah ke dalam lima bidang dengan rata-rata akurasi mencapai 86,68% [9]. Dania et al. (2024) mengembangkan sistem klasifikasi berdasarkan konsentrasi studi dan memperoleh akurasi sebesar 80% [10]. Sukriadi et al. (2023) memanfaatkan pendekatan text mining untuk klasifikasi skripsi dan mengelompokkan judul ke dalam kategori sesuai (Y) dan tidak sesuai (T), dengan hasil implementasi menunjukkan 65% judul sesuai [11]. Penelitian oleh Zaza et al. (2024) di Universitas Stella Maris mencatat akurasi tertinggi, yaitu 100% dalam klasifikasi judul berdasarkan minat, dengan kategori dominan pada Rekayasa Perangkat Lunak dan Expert Decision System [12]. Sementara itu, Yumami et al. (2025) menekankan pentingnya klasifikasi otomatis sebagai solusi terhadap lonjakan jumlah skripsi, dengan model yang berhasil mencapai akurasi klasifikasi sebesar 65% [13]. Secara keseluruhan, kelima studi ini menegaskan bahwa Naive Bayes adalah metode yang

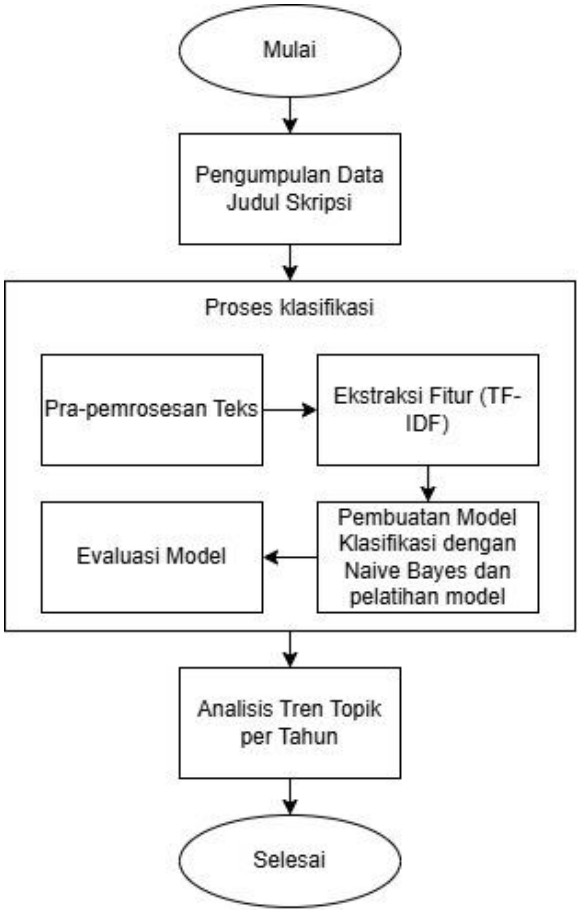
sederhana namun efektif dalam klasifikasi teks judul skripsi, namun masih terbuka peluang untuk optimalisasi dan pengembangan lebih lanjut.

Penelitian ini ditujukan untuk mengklasifikasikan judul skripsi mahasiswa dengan bantuan algoritma Naive Bayes Classifier. Tidak hanya sebatas pengelompokan berdasarkan bidang ilmu, penelitian ini memberikan gambaran mengenai tren topik yang berkembang dari tahun ke tahun. Berbeda dengan penelitian sebelumnya yang hanya fokus pada seberapa akurat sistem dalam mengklasifikasi, penelitian ini menambahkan elemen waktu guna melihat bagaimana pola dan popularitas topik berubah.

2. METODE

2.1 Pengumpulan Data

Data penelitian yang digunakan adalah data judul skripsi di Program Studi Teknologi Informasi, Jurusan Teknik, Universitas Muhammadiyah Purworejo. Data penelitian yang diperoleh sebanyak 81 judul skripsi dari tahun 2021 sampai 2025. Data judul skripsi ini akan dibagi menjadi lima kategori, yaitu sistem informasi/website, multimedia, jaringan, IoT, dan sistem cerdas. Metode penelitian yang digunakan adalah sesuai Gambar 1.



Gambar 1. Metode penelitian

Jurnal INTEK (Informatika dan Teknologi Informasi)
Volume 8 Nomor 2, November 2025
eISSN: 2620 – 4924 | pISSN: 2620 – 4843 | Halaman: 35 – 44

2.2 Proses Klasifikasi

2.2.1 Pra Pemrosesan Teks

Oleh karena data penelitian ini adalah berupa judul skripsi, maka data tersebut adalah data tidak terstruktur berupa teks, sehingga pemrosesan data dilakukan dengan metode text mining. Text Mining merupakan pendekatan yang dimanfaatkan untuk menggali informasi dari data dalam jumlah besar, terutama data berbasis teks. Proses ini dilakukan dengan menggabungkan teknologi seperti machine learning, pengolahan bahasa alami (NLP), manajemen pengetahuan, hingga sistem pencarian informasi. Dalam penerapannya, text mining mencakup langkah-langkah awal seperti pembersihan data, pengelompokan teks, dan penarikan informasi penting dari dokumen. Teknik ini sangat berguna untuk menemukan pola tersembunyi atau informasi berharga dari kumpulan data teks yang kompleks. Tak hanya itu, metode ini juga dapat digunakan untuk menyelesaikan persoalan seperti klasifikasi dokumen, pencarian informasi yang relevan, penarikan data penting, hingga pengelompokan topik [14].

Pada penelitian ini, text mining digunakan pada pra pemrosesan teks, antara lain case folding, tokenizing, dan stopwords. Dalam proses pra-pemrosesan teks, case folding dilakukan dengan mengonversi semua huruf dalam teks menjadi huruf kecil. Setelah itu, teks dipecah menjadi unit-unit kata melalui proses tokenizing. Kata-kata umum yang tidak memberikan makna spesifik terhadap isi dokumen, atau disebut stopwords, akan dihapus dari data [14].

2.2.2 Ekstraksi Fitur

Ekstraksi fitur pada penelitian ini menggunakan TF-IDF. TF-IDF adalah metode pembobotan yang sering digunakan untuk mengetahui seberapa penting sebuah kata dalam sebuah dokumen dan juga dalam keseluruhan kumpulan dokumen. Ketika metode ini diterapkan, hasilnya berupa vektor yang memuat berbagai istilah, di mana setiap kata diperlakukan sebagai fitur yang dapat dikenali sistem. Nilai bobot dari tiap kata kemudian dihitung menggunakan rumus tertentu, sehingga sistem dapat mengenali mana kata yang paling relevan dalam konteks dokumen tersebut [15].

Dalam proses ini, nilai bobot untuk tiap judul skripsi dihitung dengan mengalikan frekuensi istilah dari tiap kata penyusun judul dengan nilai IDF. Tahapan pembobotan dilakukan melalui [7]:

1. Penentuan jumlah dokumen yang mengandung kata (Frekuensi Dokumen)
2. Perhitungan IDF (Invers dari Frekuensi Dokumen)
3. Perkalian antara Frekuensi Dokumen dengan IDF untuk mendapatkan bobot

Adapun rumus dari TF-IDF adalah sebagai berikut [15][16]:

$$W_{dt} = tf_{df} \, x \, Id_{ft} \qquad (1)$$

Di mana:
W_{dt} : nilai pembobotan dokumen ke-d terhadap istilah ke-t
tf_{dt} : frekuensi kemunculan istilah yang dicari dalam satu dokumen
Id_{ft} : *Inversed Document Frequency* (log (N/df))
N : total jumlah dokumen keseluruhan
df : jumlah dokumen yang memuat istilah yang dicari

2.2.3 Pembuatan Model Klasifikasi

.

Jurnal INTEK (Informatika dan Teknologi Informasi)
Volume 8 Nomor 2, November 2025
eISSN: 2620 – 4924 | pISSN: 2620 – 4843 | Halaman: 35 – 44

Pada penelitian ini, pembuatan atau pembangunan model klasifikasi menggunakan metode *Naive Bayes Classifier* (NBC). Konsep peluang bersyarat Teorema Bayes dijelaskan menggunakan persamaan matematis (2) [12], [10].

$$\boldsymbol{P(h|D) = \frac{P(D|h).P(h)}{P(D)}} \qquad (2)$$

Di mana:
h : dugaan atau asumsi mengenai jenis atau kelas suatu data.
D : data yang belum dikelompokkan ke dalam kategori tertentu.
P(h) : seberapa besar kemungkinan sebuah dugaan sebelum data baru dianalisis (ini dikenal sebagai probabilitas awal).
P(D) : probabilitas munculnya data D secara umum.
P(h|D): peluang bahwa hipotesis h benar jika data D telah diketahui, sering disebut sebagai probabilitas setelah pengamatan.
P(D|h): seberapa besar peluang data D muncul jika hipotesis h diasumsikan benar.

Pada tahap ini juga dilakukan pelabelan dan pemisahan dataset (data judul skripsi) untuk keperluan training dan testing. Label dataset berdasarkan kategori dari judul skripsi yaitu sebanyak 5 label, antara lain sistem informasi/website, multimedia, jaringan, IoT, dan sistem cerdas. Adapun pembagian dataset untuk training sebanyak 80% dari jumlah dataset dan 20% untuk testing model klasifikasi.

2.2.4 Evaluasi Model Klasifikasi

Guna mengetahui seberapa baik model klasifikasi bekerja, penelitian ini menggunakan *Confusion Matrix*. Metode ini menyajikan hasil pengujian yang menunjukkan ukuran penting seperti akurasi, presisi, *recall*, dan F-*measure*. Akurasi menunjukkan berapa banyak data yang berhasil diprediksi dengan tepat, sementara presisi menggambarkan proporsi prediksi positif yang benar. *Recall* digunakan untuk menilai seberapa baik model menangkap kasus positif yang sebenarnya ada. F-*measure* sendiri memberikan gambaran keseimbangan antara presisi dan recall. Dengan metrik-metrik ini, peneliti bisa menilai performa model secara lebih menyeluruh [17]. Rumus perhitungan Akurasi, Presisi, *Recall*, dan F-*Measure* ditunjukkan persamaa (3) hingga (6) [17].

$$\boldsymbol{Akurasi = \frac{TP + TN}{TP + TN + FP + FN}} \qquad (3)$$

$$\boldsymbol{Presisi = \frac{TP}{FP + TP}} \qquad (4)$$

$$\boldsymbol{Recall = \frac{TP}{FN + TP}} \qquad (5)$$

$$\boldsymbol{F - measure = 2\ x\ \frac{(Recall\ x\ Presisi)}{(Recall + Presisi)}} \qquad (6)$$

Di mana:

TP : *True Positive*, menggambarkan situasi di mana data yang memang tergolong positif berhasil dikenali dengan tepat oleh sistem.
TN : *True Negative*, kondisi ketika sistem berhasil mengidentifikasi data negatif sesuai dengan kenyataannya.
FP : *False Positive*, terjadi saat data yang seharusnya negatif malah dianggap sebagai positif oleh sistem, sehingga menimbulkan kesalahan klasifikasi.
FN : *False Negative*, menunjukkan kesalahan sistem dalam mengenali data positif, yang malah dikategorikan sebagai negatif.

2.2.5 Analisis Tren Topik

Analisis tren topik pada penelitian ini menggunakan *Word Frequency Analysis*. *Word Frequency Analysis* digunakan dalam text mining untuk mengetahui seberapa sering suatu kata muncul dalam sekumpulan data. Teknik ini bermanfaat untuk mengidentifikasi kata-kata yang dominan dalam teks, memberikan gambaran awal terkait topik yang sering dibahas, serta menyajikan hasilnya dalam bentuk tabel berisi daftar kata dan jumlah kemunculannya [18].

Guna memvisualisasikan frekuensi kata yang sering muncul, digunakan visualisasi menggunakan Bar Chart, untuk menampilkan sepuluh kata yang paling sering muncul pada judul skripsi tiap tahunnya.

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan bahasa pemrograman python untuk membantu pembentukan model klasifikasi. Berdasarkan pembagian dataset untuk training dan testing, diperoleh data sebanyak 64 dataset (80%) sebagai data training dan 17 dataset sebagai data testing (20%).

```
Jumlah data training: 64
Jumlah data testing: 17
```

Gambar 2. Jumlah data training dan testing

Dari 17 dataset testing, terdapat kecocokan antara data testing dan data prediksi, di mana terdapat 13 dataset yang cocok, dan 4 dataset yang tidak cocok.

```
Hasil testing:
[2, 1, 5, 1, 1, 1, 1, 2, 2, 1, 1, 1, 1, 2, 2, 1, 2]
Hasil prediksi:
[2, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 2, 1, 1, 1]
```

Gambar 3. Hasil testing dan hasil prediksi data testing

Evaluasi model klasifikasi menggunakan Naïve Bayes Classifier menunjukkan tingkat akurasi, presisi, recall, atau F-measure ditunjukkan Gambar 4.

Jurnal INTEK (Informatika dan Teknologi Informasi)
Volume 8 Nomor 2, November 2025
eISSN: 2620 – 4924 | pISSN: 2620 – 4843 | Halaman: 35 – 44

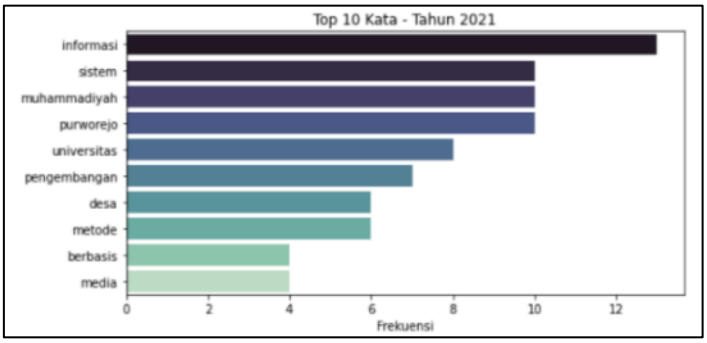
Classification Report:

	precision	recall	f1-score	support
1	0.71	1.00	0.83	10
2	1.00	0.50	0.67	6
5	0.00	0.00	0.00	1
accuracy			0.76	17
macro avg	0.57	0.50	0.50	17
weighted avg	0.77	0.76	0.73	17

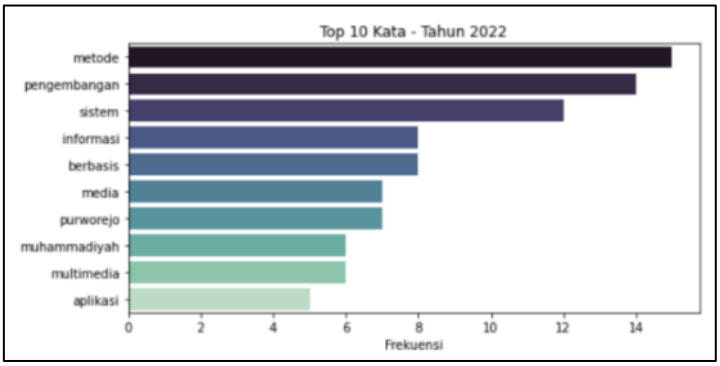
Gambar 4. Hasil Classification Report dari model klasifikasi Naïve Bayes Classifier

Pada gambar 4, akurasi model senilai sekitar 76,5%, yang berarti dari sebagian besar prediksi model adalah bernilai benar. Nilai presisi sebesar 77%, yang berarti dari semua prediksi positif yang dilakukan oleh model, sekitar 77% benar-benar positif. Ini menunjukkan model cukup baik dalam menghindari kesalahan positif (*False Positive*). Nilai *recall* sebesar 76%, yang berarti dari semua data yang sebenarnya positif, model berhasil menangkap sekitar 76% di antaranya. Ini menunjukkan model cukup baik dalam menemukan semua kasus positif. Selain itu, nilai F1-*measure* atau F1-*score* sebesar 73%, yang menandakan keseimbangan antara kemampuan model mengidentifikasi data positif dan menghindari prediksi positif yang salah.

Kemudian, dengan menggunakan *Word Frequency Analysis*, dapat diketahui sepuluh kata terbanyak yang ada pada judul-judul skripsi pada tahun 2021 sampai 2025. Visualisasi data *Word Frequency Analysis* menggunakan diagram Bar Chart.

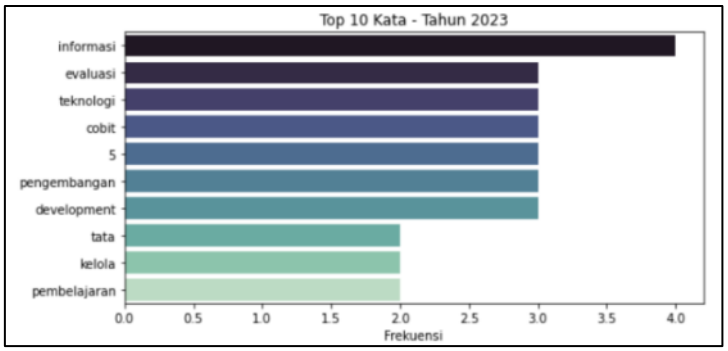


Gambar 5. Hasil *Word Frequency Analysis* judul skripsi tahun 2021

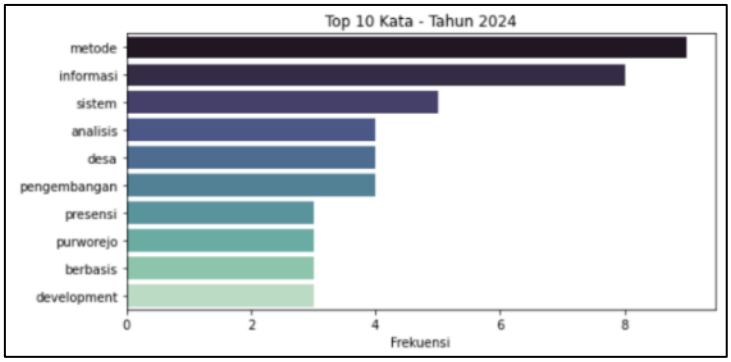


Gambar 6. Hasil *Word Frequency Analysis* judul skripsi tahun 2022

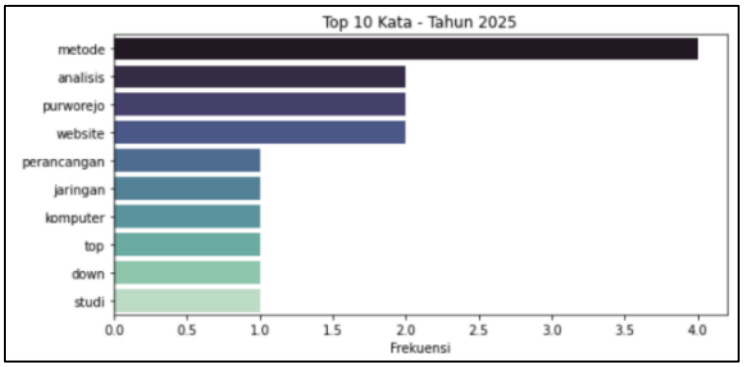
Jurnal INTEK (Informatika dan Teknologi Informasi)
Volume 8 Nomor 2, November 2025
eISSN: 2620 – 4924 | pISSN: 2620 – 4843 | Halaman: 35 – 44



Gambar 7. Hasil *Word Frequency Analysis* judul skripsi tahun 2023



Gambar 8. Hasil *Word Frequency Analysis* judul skripsi tahun 2024



Gambar 9. Hasil *Word Frequency Analysis* judul skripsi tahun 2025

4. KESIMPULAN

Penerapan algoritma *Naive Bayes Classifier* berhasil mengklasifikasikan judul skripsi ke dalam lima kategori bidang ilmu dengan tingkat akurasi 76,5%. Evaluasi menggunakan precision, recall, dan F1-score juga menunjukkan performa model yang baik dan seimbang. Selain klasifikasi, analisis tren topik melalui word frequency analysis memberikan gambaran tentang tema-tema yang populer dari tahun ke tahun. Pendekatan ini tidak hanya berguna bagi mahasiswa dalam menentukan judul skripsi, tetapi juga bagi institusi dalam memetakan arah dan keberagaman topik penelitian mahasiswa.

REKOMENDASI

Penelitian ke depan dapat melibatkan lebih banyak program studi atau universitas agar datanya makin beragam. Selain itu, dapat menerapkan algoritma lain seperti SVM atau Random Forest, untuk membandingkan algoritma mana yang performanya paling baik. Visualisasi dapat ditingkatkan dengan WordCloud atau teknik seperti LDA Topic Modeling, untuk memperkaya pemahaman terhadap topik-topik yang muncul

DAFTAR PUSTAKA

[1] R. Nuraeni, A. Sudiarjo, and R. Rizal, "Perbandingan Algoritma Naïve Bayes Classifier dan Algoritma Decision Tree untuk Analisa Sistem Klasifikasi Judul Skripsi," *Innov. Res. Informatics*, vol. 3, no. 1, pp. 26–31, 2021, doi: 10.37058/innovatics.v3i1.2976.

[2] M. Wangge, "Penerapan Metode Principal Component Analysis (PCA) Terhadap Faktor-faktor yang Mempengaruhi Lamanya Penyelesaian Skripsi Mahasiswa Program Studi Pendidikan Matematika FKIP UNDANA," *J. Cendekia J. Pendidik. Mat.*, vol. 5, no. 2, pp. 974–988, 2021, doi: 10.31004/cendekia.v5i2.465.

[3] M. M. Da'awi and W. I. Nisa, "PSIKODINAMIKA : JURNAL LITERASI PSIKOLOGI Pengaruh Dukungan Sosial terhadap Tingkat Stres dalam Penyusunan Tugas Akhir Skripsi," *J. Literasi Psikol.*, vol. 1, no. 1, pp. 67–75, 2021.

[4] R. Sulaiman, R. Artiono, and B. Rahajeng, "Menentukan Topik Skripsi Mahasiswa Dengan Menggunakan Relasi Fuzzy Intuisionistik," *Sainmatika J. Ilm. Mat. dan Ilmu Pengetah. Alam*, vol. 19, no. 1, p. 8, 2022, doi: 10.31851/sainmatika.v19i1.7927.

[5] B. H. Irawan, M. S. H. Simarangkir, and E. Erlinna, "Deteksi Kemiripan Judul Skripsi Menggunakan Algoritma Levenshtein Distance Pada Kampus Stmik Mic Cikarang," *Edutic - Sci. J. Informatics Educ.*, vol. 7, no. 2, pp. 143–149, 2021, doi: 10.21107/edutic.v7i2.10051.

[6] D. M. U. Atmaja and R. Mandala, "Analisa Judul Skripsi untuk Menentukan Peminatan Mahasiswa Menggunakan Vector Space Model dan Metode K-Nearest Neighbor," *IT Soc.*, vol. 4, no. 2, pp. 1–6, 2020, doi: 10.33021/itfs.v4i2.1182.

[7] A. Supriatman, "Pembobotan TF-IDF pada Judul Penelitian Dosen Sebagai Dasar Klasifikasi Menggunakan Algoritma K-NN (Studi Kasus: Universitas Siliwangi)," *J. Serambi Eng.*, vol. 6, no. 1, pp. 1573–1579, 2021, doi: 10.32672/jse.v6i1.2645.

[8] A. Erna, A. K. Jailani, S. Informasi, and U. Dipa, "Identifikasi Potensi Topik Penelitian Baru : Analisis Judul Skripsi Teknik Informatika (2022-2024) menggunakan Latent Dirichlet Allocation (LDA)," vol. XIII, no. 2, pp. 159–167, 2024.

[9] N. Nurdin, M. Suhendri, Y. Afrilia, and R. Rizal, "Klasifikasi Karya Ilmiah (Tugas Akhir) Mahasiswa Menggunakan Metode Naive Bayes Classifier (NBC)," *Sistemasi*, vol. 10, no. 2, p. 268, 2021, doi: 10.32520/stmsi.v10i2.1193.

[10] S. Dania, R. Ishak, M. Kom, and H. Dalai, "Penerapan Algoritma Naive Bayes Classifier Untuk Klasifikasi Judul Skripsi Berdasarkan Konsentrasi," *J. Ilm. Ilmu Komput. Banthayo Lo Komput.*, vol. 3, no. 1, pp. 15–22, 2024.

[11] S. Sukriadi, I. Ismail, and A. M. Andzar, "Penerapan Text Mining Dalam Klasifikasi Judul Skripsi Yang Diusulkan Mahasiswa Menggunakan Metode Naïve Bayes," *J. Ilm. Sist. Inf. dan Tek. Inform.*, vol. 6, no. 2, pp. 184–196, 2023, doi: 10.57093/jisti.v6i2.174.

[12] D. A. U. Zaza, E. Umar, F. Ema, and O. Sanga, "Penerapan Text Mining Dalam Klasifikasi Judul Skripsi Mahasiswa Pada Universitas Stella Maris Sumba Menggunakan Metode Naïve Bayes," vol. 02, no. 03, pp. 327–338, 2024.

[13] E. Yumami, M. Jannah, K. O. Putra, I. Teknologi, and M. Gama, "KONSENTRASI PROGRAM STUDI MENGGUNAKAN," vol. 5, no. 1, pp. 24–28, 2025.

[14] T. Ridwansyah, "Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 5, pp. 178–185, 2022, doi: 10.30865/klik.v2i5.362.

[15] A. D. D. Wibiyanto and A. Wibowo, "Penerapan Algoritma Multiclass Support Vector Machine Dan Tf-Idf Untuk Klasifikasi Topik Tugas Akhir," *Skanika*, vol. 6, no. 1, pp. 42–50, 2023, doi: 10.36080/skanika.v6i1.2999.

[16] N. Andriani and A. Wibowo, "Implementasi Text Mining Klasifikasi Topik Tugas Akhir Mahasiswa Teknik Informatika Menggunakan Pembobotan TF-IDF dan Metode Cosine Similarity Berbasis Web," *Senamika*, no. September, pp. 130–137, 2021.

[17] D. Putra and A. Wibowo, "Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes," *Pros. Semin. Nas. Ris. Dan Inf. Sci.*, vol. 2, pp. 84–92, 2020.

[18] S. Rosalin and B. F. Supriyanto, "Analisis Sentimen Program Merdeka Belajar dengan Text Analysis Wordcloud & Word Frequency," *J. Minfo Polgan*, vol. 12, no. 1, pp. 25–32, 2023, doi: 10.33395/jmp.v12i1.12312.