

PENERAPAN SISTEM REKOMENDASI PENELITI BERBASIS *HYBIRD* DENGAN MODEL *LDA* DAN *NAÏVE BAYES*

1st Alfonsus Antero Arnayusrendito

Sistem Informasi

Telkom University

Surabaya, Indonesia

alfonsusa@student.telkomuniversity.ac.id

2nd Mochamad Nizar Palefi Ma'ady

Sistem Informasi

Telkom University

Surabaya, Indonesia

mnizarpm@telkomuniversity.ac.id

3rd Purnama Anaking

Sistem Informasi

Telkom University

Surabaya, Indonesia

purnamaanaking@telkomuniversity.ac.id

Abstrak — Kolaborasi penelitian sangat penting untuk meningkatkan kualitas dan produktivitas penelitian, sehingga dapat memperkuat peran inovasi dalam pengembangan ilmu pengetahuan dan teknologi. Namun, banyak peneliti di ruang lingkup Telkom University Surabaya yang kesulitan menemukan dosen atau rekan peneliti yang sesuai bidang keahliannya, yang menjadi suatu tantangan besar dalam membangun kerjasama penelitian yang efektif. Penelitian ini bertujuan untuk meningkatkan kolaborasi multidisipliner dan memfasilitasi pencarian rekan peneliti di lingkungan akademik Telkom University Surabaya, melalui pengembangan sistem rekomendasi dosen peneliti yang inovatif dengan memanfaatkan data publikasi dari *Google Scholar*. Metode yang digunakan adalah model *Latent Dirichlet Allocation (LDA)* sebagai identifikasi topik, sementara *Naïve Bayes* digunakan untuk klasifikasi data kata kunci guna memprediksi dosen peneliti yang relevan. Hasil penelitian ini berupa sistem rekomendasi yang dapat merekomendasikan dosen sesuai dengan bidang keahlian dan pengalaman penelitian mereka. Sistem ini diharapkan mendukung peneliti dalam menemukan mitra kolaborasi dan membantu mahasiswa dalam memilih dosen pembimbing. Evaluasi model menunjukkan performa baik dalam membantu pengguna untuk merekomendasikan dosen peneliti dengan akurasi 88%, *precision* 86%, *recall* 85%, dan *F1-score* 85%. Sistem ini sejalan dengan tujuan *Sustainable Development Goals (SDGs)* dalam mendukung inovasi yang modern dan kolaborasi penelitian.

Kata kunci— Sistem Rekomendasi, *Latent Dirichlet Allocation*, *Naïve Bayes*.

I. PENDAHULUAN

Publikasi merupakan salah satu jenis jurnal atau *paper* yang berisikan mengenai artikel penelitian ilmiah. Pentingnya publikasi ilmiah sebagai tanda bukti utama orisinalitas penelitian, hal ini menjadikannya sebagai modal pengalaman penelitian sebagai akademisi [1].

Jumlah penelitian di Indonesia, berdasarkan data dari Kemendikbudristek, terus meningkat seiring dengan kemajuan dalam ilmu pengetahuan, sains, dan teknologi [2]. Reputasi akademik Indonesia di kancah internasional semakin diakui, membuka peluang untuk kolaborasi penelitian. Hal ini sangat penting dalam menyebarkan pengetahuan baru, mendorong inovasi, dan membuka peluang untuk melakukan penelitian lanjutan [3].

Namun, hal ini dapat menjadi tantangan baru di lingkungan akademik, terutama dalam upaya menemukan *partner* penelitian yang sesuai dengan bidang keahlian. Maka dari itu diperlukannya sistem rekomendasi yang dapat memberikan fasilitas pencarian *partner* penelitian secara tepat [4]. Sistem rekomendasi adalah suatu sistem pengambilan informasi (*Information Retrieval*), yang dapat dikatakan sebuah perangkat lunak yang dapat memberikan sebuah saran dalam bentuk item atau file yang diberikan kepada pengguna sesuai dengan kebutuhan yang diinginkan serta dapat membantu pengguna dalam mengambil suatu keputusan [5], [6].

Didalam suatu sistem rekomendasi, akan menghasilkan suatu saran yang bermanfaat bagi pengguna [7]. Dalam penelitian ini bertujuan untuk membentuk sebuah produk berupa sistem rekomendasi dosen peneliti dalam ruang lingkup Telkom University Surabaya. Sistem rekomendasi ini akan memanfaatkan data publikasi dari seluruh dosen Telkom University Surabaya yang berasal dari berbagai program studi. Data terintegrasi dengan *Google Scholar* sebagai platform jurnal ilmiah untuk mengidentifikasi topik penelitian yang relevan dan memberikan rekomendasi dosen secara optimal.

Penelitian ini akan menggunakan metode *Latent Dirichlet Allocation* dan klasifikasi *Naive Bayes*, *Latent Dirichlet Allocation* salah satu teknik *topic modelling* paling populer yang dapat mengidentifikasi pola topik dalam suatu kumpulan teks [8], Tujuan penggunaan metode *Latent Dirichlet Allocation* dapat mengidentifikasi suatu topik dengan akurat pada suatu kumpulan data maupun dokumen yang cukup besar dan kompleks [9], [10] Kemudian metode *Naive Bayes* merupakan metode *data mining* dengan menggunakan penerapan probabilistik dan statistik untuk melakukan klasifikasi data [11], yang digunakan untuk memprediksi suatu kemungkinan yang akan datang berdasarkan dari suatu data yang telah dimanfaatkan [12], [13]

Penelitian ini diharapkan dapat membantu para peneliti, dosen, dan mahasiswa dalam memberikan saran serta rekomendasi dosen peneliti Telkom University Surabaya yang tepat, agar dapat mempermudah dalam menemukan mitra kolaborasi dalam bidang penelitian tertentu. Lalu, penelitian ini juga diharapkan dapat menjadi

rekomenadasi bagi mahasiswa yang akan menempuh tugas akhir dalam menentukan dosen pembimbing berdasarkan bidang dan pengalaman penelitian. Hasil penelitian yang didapatkan berupa aplikasi berbasis website yang terintegrasi dengan *Google Scholar*. Sehingga penelitian yang diharapkan mampu memberikan saran dan tanggapan untuk mencari *partner*/dosen peneliti dalam ruang lingkup Telkom University Surabaya.

II. KAJIAN TEORI

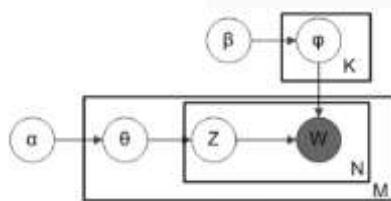
A. Preprocessing Text

Text Preprocessing merupakan salah satu tahap pembersihan data, menyusun, dan merapikan data dalam bentuk teks yang awal mula tidak terstruktur hingga terstruktur sesuai format asli sehingga dengan adanya pembersihan data dalam teks ini mampu dapat dianalisis lebih efektif dan akurat [14], [15]

- *Case Folding* : Mengubah huruf kapital menjadi huruf kecil dan menghapus angka maupun tanda baca.
- *Normalization* : Mengubah kata informal menjadi kata yang lebih formal.
- *Lemmatization* : Mengubah kata imbuhan menjadi kata pada aslinya.
- *Stopwords Removal* : Menghapus sebuah kata yang tidak memiliki makna dan inti yang signifikan

B. Latent Dirichlet Allocation

Latent Dirichlet Allocation merupakan salah satu model generatif melalui pendekatan *conditional probability* dalam memodelkan suatu kumpulan dokumen. Pada dasarnya setiap dokumen dapat disimpulkan sebagai campuran dari berbagai struktur pola yang tersembunyi dan tidak dapat diketahui [15], [16]. Metode ini dikembangkan sejak tahun 2003, metode ini merupakan salah satu metode terpopuler dalam melakukan pemodelan topik. Sebagai contoh proses *LDA* disajikan pada gambar 1.



GAMBAR 1
(REPRESENTASI MODEL LDA)

Dalam penelitian ini, *Latent Dirichlet Allocation (LDA)* diusulkan untuk mengekstraksi kata kunci dari dokumen abstrak. evaluasi dilakukan menggunakan skor koherensi untuk menentukan topik yang paling relevan.

C. Naïve Bayes

Naïve Bayes merupakan salah satu metode pembelajaran mesin yang bertugas untuk mengklasifikasi data dengan perhitungan probabilitas dan statistik yang sederhana [17]. Peran *Naïve Bayes* dalam *text mining* pada sistem rekomendasi dosen peneliti mampu mendapatkan hasil tingkat akurasi yang tinggi dan memiliki prediksi dan prakiraan yang kuat di segala kondisi, serta memiliki algoritma yang mampu mengandalkan pendekatan

probabilitas. Sehingga *Naïve Bayes* mampu menangani dataset dalam jumlah yang besar dengan efisien. Strategi ini mengasumsikan bahwa semua atribut bersifat independen satu sama lain dalam hubungannya dengan variabel kelas [18]. Dalam hal ini proses klasifikasi metode *Naïve Bayes* dapat dirumuskan pada rumus 1.

$$P(H|X) = \frac{P(X|H) P(H)}{P(X)} \quad (1)$$

Berdasarkan rumus diatas dapat diasumsikan sebagai berikut [19]:

- X : Data yang belum diketahui kelasnya
- H : Hipotesis data X dari suatu kelas
- $P(H|X)$: Probabilitas hipotesis H pada kondisi X
- $P(H)$: Probabilitas hipotesis H
- $P(X|H)$: Probabilitas X pada kondisi hipotesis H
- $P(X)$: Probabilitas X

D. Cross-Validation

Cross Validation adalah metode evaluasi evaluasi performa dalam *data mining* yang menguji kinerja model dengan membagi dataset menjadi beberapa bagian. Secara bergiliran, satu bagian digunakan untuk pengujian dan sisanya untuk pelatihan, lalu proses diulang sesuai jumlah lipatan (K), [12], [13]. Metode ini membantu memperoleh akurasi model yang lebih valid, sebagai contoh pada tabel 1 membagikannya (K) = 5 dengan 100 dataset.

TABEL 1
(CROSS-VALIDATION)

Jumlah	F-1	F-2	F-3	F-4	F-5
1-20	Test	Train	Train	Train	Train
21-40	Train	Test	Train	Train	Train
41-60	Train	Train	Test	Train	Train
61-80	Train	Train	Train	Test	Train
81-100	Train	Train	Train	Train	Test

E. Confusion Matrix

Confusion Matrix merupakan metode evaluasi performa yang dapat digunakan untuk memberikan tingkat akurasi maupun menghitung dari kinerja model sendiri untuk mengetahui tingkat kebenaran dari proses klasifikasi [12].

TABEL 2
(CONFUSION MATRIX)

Actual	Prediction	
	Positive	Negative
Positive	TP	TN
Negative	FP	FN

Metode evaluasi yang terdiri dari akurasi, *precision*, *recall*, dan *F1-Score* dapat dihitung sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$Recall = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

F. *Flask*

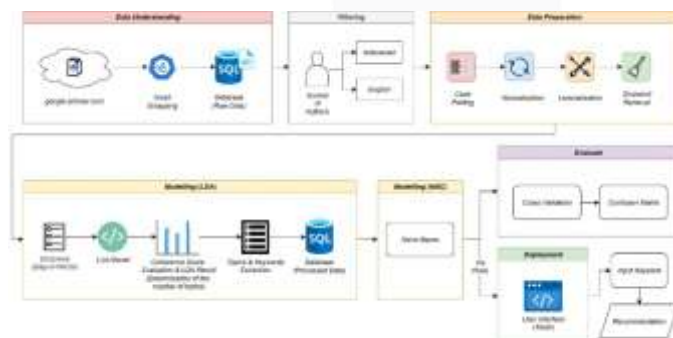
Flask merupakan salah satu *web framework* dengan menggunakan bahasa pemrograman *Python*. *Flask* merupakan salah satu *framework back-end* berbasis *micro framework* karena *Flask* memiliki kemampuan *simplicity*, *flexibility*, dan ringan digunakan, sehingga *framework flask* dapat mampu dikembangkan dan menyesuaikan dengan kebutuhan didukung adanya ekstensi yang diperlukan[20].

G. Sistem Rekomendasi

Sistem rekomendasi merupakan sebuah sistem perangkat lunak yang menyediakan berupa saran atau usulan informasi dalam bentuk file atau item untuk membantu dan mempermudah pencarian pengguna yang diberikan dalam mendukung pengambilan keputusan sesuai dengan pilihan, minat, dan kebutuhan pengguna[5][6]. Selain hal tersebut, dalam penelitian ini adalah sistem rekomendasi yang mampu memberikan rekomendasi serta nilai relevansi dosen sesuai kecocokan dengan kata kunci pencarian dari pengguna[18].

III. METODE PENELITIAN

Berdasarkan penelitian yang dilakukan, metodologi yang diterapkan adalah *CRISP-DM* yang merupakan salah satu metodologi penelitian dalam penerapan *data mining* yang terbagi menjadi 6 tahap yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluate*, dan *Deployment*. Sebagai gambarannya, berikut merupakan arsitektur sistem yang diusulkan peneliti pada gambar 2.



GAMBAR 2
(ARSITEKTUR SISTEM)

B. Data Understanding

Pada tahap *Data Understanding* ini merupakan tahap pengumpulan data yang meliputi data program studi, data dosen, dan scraping data publikasi per dosennya. Dalam pengumpulan data publikasi per dosennya dilakukan dengan teknik *scraping* menggunakan *HTML DOM Parser*. yang dapat mengekstrasi data ini berupa data judul, data dosen penulis, data abstrak publikasi, dan *link* publikasi. Berikut ini merupakan proses *scraping* disajikan pada gambar 4.



GAMBAR 4
(PROSES SCRAPING DATA PUBLIKASI)

Proses *scraping* ini dilakukan dengan memanfaatkan *ID* unik dosen yang terdapat pada tautan di halaman profil *author Google Scholar*. *ID* tersebut dimasukkan ke dalam input label pada halaman *web scraping* yang ditunjukkan pada Gambar 4. Selanjutnya, data yang diperoleh melalui teknik *scraping* akan disimpan secara otomatis ke dalam *database MySQL*, dan dilanjutkan dengan *preprocessing text* pada tahap *Data Preparation*.

C. Data Preparation

Pada tahap *Data Preparation* ini merupakan tahap *preprocessing text* pada data abstrak. Pada penelitian ini, tahap *preprocessing text* ini akan dilakukan dengan 4 tahap yaitu *case folding*, *normalization*, *lemmatization*, dan *stopwords removal*. Hasil *preprocessing text* dapat disajikan pada tabel 3.

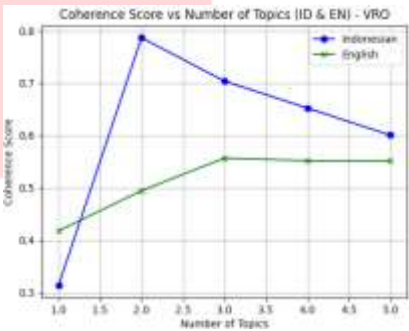
TABEL 3
(PREPROCESSING TEXT)

Tahap	Hasil
Abstrak Mentah	The COVID-19 pandemic has caused significant loss of human life worldwide and presented unprecedented challenges to public health, food systems...
Case Folding	the covid pandemic has caused significant loss of human life worldwide and presented unprecedented challenges to public health food systems...
Normalization	the covid-19 pandemic has caused significant loss of human life worldwide and presented unprecedented challenges to public health food systems...
Lemmatization	the covid-19 pandemic ha caused significant loss of human life worldwide and presented unprecedented challenge to public health food system

Stopwords Removal	covid-19 pandemic caused significant loss human life worldwide presented unprecedented challenge public health food system
-------------------	--

D. Modeling

Tahap *Modeling* ini merupakan tahap eksekusi data setelah melakukan *preprocessing text*, yaitu melakukan pemodelan topik dengan model *Latent Dirichlet Allocation*. Proses ini diawali dengan ekstrasi fitur menggunakan *Bag-of-Word*. Kemudian dilanjutkan dengan pemodelan topik. Berikut merupakan hasil dari pemodelan topik dari salah satu data abstrak yang dimiliki oleh peneliti dengan melakukan evaluasi skor koherensi terlebih dahulu.



GAMBAR 5
(EVALUASI COHERENCE SCORE)

Berdasarkan gambar 5 hasil evaluasi *coherence score* menunjukkan jumlah topik ideal adalah 2 topik bahasa Indonesia dengan nilai koherensi sebesar 0.7868 dan 3 topik bahasa Inggris dengan nilai koherensi sebesar 0.5568. Berdasarkan hasil evaluasi dengan *coherence score* tersebut, Berikut ini merupakan hasil pemodelan topik setelah melakukan evaluasi *Coherence Score* yang disajikan pada tabel 4.

TABEL 4
(HASIL PEMODELAN TOPIK)

Topics	keyword (id)
1	serangan, dns, botnet, deteksi, metode, server, model, jaringan, mesin, penelitian, simulasi, alamat, program, virtual, pembelajaran
2	klasifikasi, twitter, tweet, akurasi, buah, kemacetan, kecelakaan, parameter, sistem, fitur, kernel, uji, coba, metode, upa
Topics	keyword (en)
1	answer, vehicle, system, headlight, tree, student, operator, detection, image, accuracy, traffic, research, math, question, electronic
2	system, data, user, student, design, prediction, paper, model, efficiently, trial, implementation, information, mubaarok, ustadz, managing
3	word, system, answer, automatic, assessment, automatically, generate, issue, application, user, operator, accuracy, math, based, type

Setelah melakukan pemodelan topik menggunakan *LDA*, langkah selanjutnya adalah menyimpan data. Kata kunci yang dihasilkan dari pemodelan topik tersebut akan digunakan sebagai fitur untuk proses klasifikasi menggunakan algoritma *Naïve Bayes*.

E. Evaluate

Proses Evaluasi dalam tahap penelitian ini dilakukan dengan menggunakan 2 metode yaitu *Cross-Validation* dan *Confusion Matrix*. Tujuan evaluasi performa ini digunakan untuk mengetahui seberapa baik model mempelajari sebuah *dataset* melalui proses pelatihan menggunakan model *Naïve Bayes*.

- *Cross-Validation*

Evaluasi menggunakan *Cross-Validation* akan terbagi menjadi 5 bagian data (*k*). Berikut disajikan hasil *Cross-Validation* pada tabel 5.

TABEL 5
(HASIL *CROSS-VALIDATION*)

Fold	Cross-Validation Score
F-1	0.8974
F-2	0.8860
F-3	0.8746
F-4	0.8581
F-5	0.8662
Accuracy	0.8765

Berdasarkan hasil *Cross-Validation* menunjukkan hasil rata-rata data uji per bagian data sebesar 0.8974, 0.8860, 0.8746, 0.8581, dan 0.8662. *Fold 1* merupakan *fold* dengan nilai *Cross-Validation* tertinggi yaitu dengan skor 0.8974 atau 89.74%. Secara keseluruhan, model ini memiliki tingkat akurasi atau nilai rata-rata sebesar 0.8765 atau 87.65%.

- *Confusion Matrix*

Tahap evaluasi performa model *Naïve Bayes* selanjutnya adalah *Confusion Matrix*, yang bertujuan untuk menganalisis lebih dalam mengenai kinerja model klasifikasi pada masing-masing label atau kelas, dengan menyajikan *Classification Report* untuk mengetahui nilai akurasi, *precision*, *recall*, *F1-score*, dan *support* dalam setiap label atau kelas yang diprediksi oleh model. Karena data yang digunakan untuk klasifikasi terdiri dari banyak kelas (*multiclass*), maka evaluasi model hanya ditampilkan dalam bentuk akurasi, *precision*, *recall*, dan *F1-score* secara keseluruhan (*macro average*) yang disajikan pada tabel 6 sebagai berikut.

TABEL 6
(HASIL *CLASSIFICATION REPORT*)

Report	Value
Accurate	0.88
Precision	0.86
Recall	0.85
F1-Score	0.85

Berdasarkan tabel 6. Hasil evaluasi *Confusion Matrix* menunjukkan bahwa model *Naïve Bayes* memiliki kinerja yang cukup baik, dengan nilai akurasi sebesar 88%, *precision* sebesar 86%, *recall* sebesar 85%, dan *F1-score* sebesar 85%. Hal ini dapat disimpulkan bahwa model mampu mengklasifikasikan data dengan cukup optimal antara ketepatan prediksi dan kemampuan menangkap label yang relevan.

F. Deployment

Tahap *deployment* merupakan tahap implementasi model yang hasilnya dapat ditampilkan dalam bentuk visualisasi maupun antarmuka pengguna (*user interface*) berbasis *website* atau aplikasi. Dalam tahap penelitian ini, dihasilkan berupa sistem rekomendasi berbasis *website* yang telah terintegrasi dengan *framework Flask*. Berikut merupakan hasil tampilan awal sistem rekomendasi yang disajikan pada gambar 6.



GAMBAR 6
(TAMPILAN AWAL SISTEM REKOMENDASI)

Lalu, berikut ini merupakan tampilan rekomendasi yang akan menampilkan beberapa profil dosen peneliti di ruang lingkup Telkom University Surabaya berdasarkan *input* kata kunci yang dicari dengan disajikan pada gambar 7.



GAMBAR 7
(TAMPILAN HASIL REKOMENDASI)

V. KESIMPULAN

Penelitian ini memanfaatkan model *Latent Dirichlet Allocation (LDA)* untuk melakukan pemodelan topik terhadap data abstrak setiap dosen, sehingga dihasilkan beberapa topik beserta kata kunci yang relevan. Hasil pemodelan topik berupa kata kunci tersebut digunakan sebagai fitur dalam klasifikasi menggunakan model *Naïve Bayes*. Sehingga model *Naïve Bayes* yang dibangun mampu mencapai akurasi sebesar 88%, *precision* 86%, *recall* 85%, dan *F1-score* 85%. Hasil ini menunjukkan bahwa model mampu melakukan prediksi serta memberikan rekomendasi dosen peneliti yang relevan berdasarkan kata kunci yang dicari oleh pengguna.

REFERENSI

- [1] C. M. D. Butar Butar *et al.*, "Analisis Produktivitas Publikasi Ilmiah Mengenai Program Pertukaran

- Mahasiswa Merdeka (Pmm),” *Communnity Development Journal*, vol. 4, no. 2, pp. 4921–4924, 2023.
- [2] D. K. Kemdikbudristek, “Analisis Total Publikasi Nasional 5 Tahun Terakhir,” PTNBH. Accessed: Dec. 13, 2024. [Online]. Available: <https://sinta.kemdikbud.go.id/ptnbhanalytics/v2/indicators?indicator=14>
- [3] D. Diktiristek, “Mahasiswa PMDSU Didorong Produktif Hasilkan Publikasi Berkualitas,” 2022. [Online]. Available: <https://dikti.kemdikbud.go.id/kabar-dikti/kabar/mahasiswa-pmdsu-didorong-produktif-hasilkan-publikasi-berkualitas/>
- [4] H. Setiawan, “Implementasi Algoritma Fuzzy Topsis Pada Sistem Rekomendasi Beasiswa,” *Jurnal Sistem Telekomunikasi Elektronika Sistem Kontrol Power Sistem & Komputer*, vol. 1, no. 2, pp. 2776–5822, 2021.
- [5] M. Alkaff, H. Khatimi, and A. Eriadi, “Sistem Rekomendasi Buku pada Perpustakaan Daerah Provinsi Kalimantan Selatan Menggunakan Metode Content-Based Filtering,” *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 20, no. 1, pp. 193–202, Sep. 2020, doi: 10.30812/matrik.v20i1.617.
- [6] F. R. Hariri and L. W. Rochim, “Sistem Rekomendasi Produk Aplikasi Marketplace Berdasarkan Karakteristik Pembeli Menggunakan Metode User Based Collaborative Filtering,” *Teknika*, vol. 11, no. 3, pp. 208–217, Nov. 2022, doi: 10.34148/teknika.v11i3.538.
- [7] Z. Fayyaz, M. Ebrahimian, D. Nawara, A. Ibrahim, and R. Kashef, “Recommendation systems: Algorithms, challenges, metrics, and business opportunities,” *Applied Sciences (Switzerland)*, vol. 10, no. 21, pp. 1–20, Nov. 2020, doi: 10.3390/app10217748.
- [8] D. Rosmala and R. N. Cahyadi, “Analisis Sentimen Web Novel Menggunakan Metode Latent Dirichlet Allocation (LDA) (Study Kasus Komentar Novel Harry Potter),” *Jurnal Riset Sistem Informasi dan Teknik Informatika*, vol. 2, no. 2, 2024, doi: 10.61132/mercurius.v2i2.74.
- [9] H. Oktafiandi, “Implementasi LDA untuk Pengelompokan Topik Twitter Bertagat #Mypertamina,” *JURNAL EKONOMI DAN TEKNIK INFORMATIKA*, vol. 11, no. 1, pp. 2579–3322, 2023.
- [10] N. L. P. M. Putu, Ahmad Zuli Amrullah, and Ismarmiaty, “Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, pp. 123–131, Feb. 2021, doi: 10.29207/resti.v5i1.2587.
- [11] K. Anwar, “Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone Menggunakan Naive Bayes,” *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 2, no. 4, pp. 148–155, 2022, [Online]. Available: <https://djournals.com/klik>
- [12] R. Rizqi Robbi Arisandi, B. Warsito, and A. Rachman Hakim, “Aplikasi Naive Bayes Classifier (Nbc) Pada Klasifikasi Status Gizi Balita Stunting Dengan Pengujian K-Fold Cross Validation,” *JURNAL GAUSSIAN*, vol. 11, no. 1, pp. 130–139, 2022, [Online]. Available: <https://ejournal3.undip.ac.id/index.php/gaussian/>
- [13] N. Nurainun, E. Haerani, F. Syafria, and L. Oktavia, “Penerapan Algoritma Naive Bayes Classifier Dalam Klasifikasi Status Gizi Balita dengan Pengujian K-Fold Cross Validation,” *Journal of Computer System and Informatics (JoSYC)*, vol. 4, no. 3, pp. 578–586, May 2023, doi: 10.47065/josyc.v4i3.3414.
- [14] D. Zakeshia Tiara Kannitha and P. Kartikasari, “Pemodelan Topik Pada Keluhan Pelanggan Menggunakan Algoritma Latent Dirichlet Allocation Dalam Media Sosial Twitter,” *JURNAL GAUSSIAN*, vol. 11, no. 2, pp. 266–277, 2022, [Online]. Available: <https://ejournal3.undip.ac.id/index.php/gaussian/>
- [15] V. Alpiana, A. Salam, F. Alzami, I. Rizqa, and D. Aqmal, “Analisis Topic-Modelling Menggunakan Latent Dirichlet Allocation (LDA) Pada Ulasan Sosial Media Youtube,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 1, p. 332, Jan. 2024, doi: 10.30865/mib.v8i1.7127.
- [16] M. R. Fahlevvi and A. SN, “Topic Modeling on Online News.Portal Using Latent Dirichlet Allocation (LDA),” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 16, no. 4, p. 335, Oct. 2022, doi: 10.22146/ijccs.74383.
- [17] H. F. Putro, R. T. Vlandari, and W. L. Y. Saptomo, “Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan,” *Jurnal Teknologi Informasi dan Komunikasi (TIKOMSiN)*, vol. 8, no. 2, Oct. 2020, doi: 10.30646/tikomsin.v8i2.500.
- [18] S. H. Ramadhani and M. I. Wahyudin, “Analisis Sentimen Terhadap Vaksinasi Astra Zeneca pada Twitter Menggunakan Metode Naive Bayes dan K-NN,” *Jurnal Teknologi Informasi dan Komunikasi*, vol. 6, no. 4, p. 2022, 2022, doi: 10.35870/jti.
- [19] N. Paramitha, E. Junianto, and S. Susanti, “Penerapan Teorema Bayes Untuk Diagnosis Penyakit Pada Ibu Hamil Berbasis Android,” 2019. [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji/article/view/4>
- [20] D. F. Ningtyas and N. Setiyawati, “Implementasi Flask Framework pada Pembangunan Aplikasi Purchasing Approval Request,” *Jurnal Janitra Informatika dan Sistem Informasi*, vol. 1, no. 1, pp. 19–34, Apr. 2021, doi: 10.25008/janitra.v1i1.120.