

Optimalisasi Pencarian Dokumen Akademik Menggunakan Naïve Bayes Classifier : Studi Kasus pada *Repository Digital*

M. Aidil Putra¹, Meri Azmi^{2,*}, Yuhefizar³

^{1,2,3} Jurusan Teknologi Informasi, Politeknik Negeri Padang, Padang, 25164, Indonesia
* meriazmi@pnp.ac.id
* penulis korespondensi

ARTICLE INFO	ABSTRACT
Article history Received 14/11/2024 Revised 09/12/2024 Accepted 03/01/2025	<i>The advancement of information technology offers opportunities to improve efficiency and accessibility across various fields, including education. This study aims to develop a web-based digital repository system utilizing the Naïve Bayes Classifier (NBC) algorithm to classify academic documents based on course categories. By applying the Waterfall method in software development, the system is designed to address the needs of Information Technology students at Politeknik Negeri Padang, who often face challenges in finding academic references. The system employs the Naïve Bayes algorithm to categorize documents by analyzing the probability of relevant keywords in document titles. Testing was conducted on 200 documents, achieving an accuracy rate of 90%. A keyword-based search feature enables users to quickly and efficiently locate relevant documents. The results demonstrate that the implementation of NBC is effective for text-based document classification, enhancing accessibility and efficiency in managing academic documents. This system has the potential to be widely integrated into the academic ecosystem as an innovative digital solution</i> Keywords : Naïve Bayes Classifier, Digital Repository, Academic System, Document Search, Information Technology.
	ABSTRAK Perkembangan teknologi informasi memberikan peluang untuk meningkatkan efisiensi dan aksesibilitas dalam berbagai bidang, termasuk pendidikan. Penelitian ini bertujuan untuk mengembangkan sistem repository digital berbasis website yang memanfaatkan algoritma Naïve Bayes Classifier untuk mengklasifikasikan dokumen akademik berdasarkan kategori mata kuliah. Dengan menerapkan metode Waterfall dalam pengembangan perangkat lunak, sistem ini dirancang untuk memenuhi kebutuhan mahasiswa Jurusan Teknologi Informasi di Politeknik Negeri Padang yang sering mengalami kendala dalam mencari referensi akademik. Sistem ini menggunakan algoritma Naïve Bayes untuk mengelompokkan dokumen berdasarkan probabilitas kata kunci yang relevan dalam judul dokumen. Pengujian dilakukan pada 200 dokumen, dengan hasil akurasi mencapai 90%. Fitur pencarian berbasis kata kunci memungkinkan pengguna menemukan dokumen yang relevan secara cepat dan efisien. Hasil penelitian menunjukkan bahwa implementasi NBC efektif dalam klasifikasi dokumen berbasis teks, sehingga meningkatkan aksesibilitas dan efisiensi pengelolaan dokumen akademik. Sistem ini berpotensi untuk diintegrasikan lebih luas dalam ekosistem akademik sebagai solusi digital yang inovatif. Kata kunci : Naïve Bayes Classifier, Repository Digital, Sistem Akademik, Pencarian Dokumen, Teknologi Informasi. This is an open access article under the CC-BY-SA license. 

1. Pendahuluan

Teknologi informasi memainkan peran penting dalam mendukung berbagai aspek kehidupan manusia, termasuk dalam bidang pendidikan. Perpustakaan fisik di institusi pendidikan sering kali menjadi sumber utama referensi bagi mahasiswa. Namun, berbagai kendala seperti keterbatasan waktu, lokasi, serta situasi pandemi COVID-19 yang membatasi akses ke fasilitas fisik telah menimbulkan masalah bagi mahasiswa, khususnya mahasiswa baru Jurusan Teknologi Informasi Politeknik Negeri Padang. Ketergantungan pada dokumen hardcopy dan keterbatasan dokumen yang dapat diakses melalui senior menyebabkan kebutuhan akan sistem *repository digital* semakin mendesak.

Sistem *repository* berbasis website menjadi solusi inovatif yang tidak hanya memungkinkan pengelolaan dokumen secara digital, tetapi juga mendukung pencarian dokumen secara cepat dan efisien. Dengan memanfaatkan algoritma *Naïve Bayes Classifier* (NBC), dokumen dapat dikelompokkan berdasarkan mata kuliah secara akurat, sehingga proses pencarian menjadi lebih terstruktur dan efisien. Penelitian ini berfokus pada pengembangan sistem *repository* yang mampu memberikan kemudahan akses bagi mahasiswa untuk menemukan referensi akademik secara aman dan cepat, tanpa keterbatasan waktu dan tempat.

Naïve Bayes Classifier adalah model pembelajaran mesin berbasis probabilitas yang digunakan untuk tugas klasifikasi, dengan menerapkan Teorema Bayes dan asumsi independensi yang kuat (naïf) antar fitur[1]. Metode ini sangat efektif untuk dataset berukuran besar dan banyak digunakan di berbagai bidang, seperti kesehatan, analisis sentimen, dan data pendidikan. *Classifier* ini dikenal karena kesederhanaannya, efisiensinya, serta kemampuannya menangani masalah klasifikasi biner maupun multikelas[2][3].

Beberapa studi penelitian terdahulu sudah banyak mengangkat penggunaan *Naïve Bayes Classifier* ini. Diantaranya di bidang kesehatan, *Naïve Bayes Classifier* telah diterapkan untuk mengklasifikasikan stunting pada balita, mencapai nilai presisi, recall, dan akurasi masing-masing 85%, 91%, dan 80%, yang menunjukkan efektivitasnya dalam identifikasi dini dan pencegahan stunting[4]. Metode ini juga digunakan untuk mengklasifikasikan status gizi pada lansia, dengan akurasi 91%, yang membantu dalam deteksi dini dan penanganan malnutrisi[5].

Metode ini juga bisa diterapkan dalam melakukan analisis sentimen. Seperti pada penelitian yang bertujuan untuk mengklasifikasikan sentimen dalam judul berita, dengan mencapai akurasi sebesar 91,04%, yang menyoroti kegunaannya dalam analisis media. Nilai presisi dan recall sering kali tinggi, yang menunjukkan efektivitas model dalam mengidentifikasi *instance* yang relevan dengan benar. Sebagai contoh, dalam analisis sentimen terhadap judul berita, nilai presisi adalah 89% dan recall adalah 94% [6]. Metode ini juga telah digunakan untuk menganalisis sentimen pengguna YouTube terhadap desain mata uang baru, dengan akurasi 80%[7]. Di Instagram, metode ini mengklasifikasikan sentimen terhadap akun Presiden Joko Widodo dengan akurasi 84,1%[8]. Metode ini juga diterapkan pada data Twitter untuk menganalisis sentimen tentang *rebranding* Twitter menjadi 'X', dengan mencapai akurasi sebesar 83%[9], selain itu juga banyak penelitian analisis sentiment terkait penggunaan media sosial [10][11].

Bidang pendidikan juga sering menjadi objek penelitian terkait dengan penggunaan metode *Naïve Bayes Classifier* ini, diantaranya untuk klasifikasi penentuan kelulusan tepat waktu, dimana metode ini telah digunakan untuk memprediksi kelulusan tepat waktu mahasiswa universitas, dengan akurasi klasifikasi sebesar 70% menggunakan *K-Fold Cross Validation* [12]. Selain itu juga digunakan untuk penentuan penerimaan beasiswa [13][14], penentuan loyalitas siswa [15] dan masih banyak lagi.

Penelitian ini bertujuan untuk menjawab beberapa pertanyaan utama, yaitu bagaimana membangun sistem *repository* berbasis website untuk mendukung pencarian dokumen akademik secara digital, bagaimana algoritma *Naïve Bayes Classifier* diterapkan untuk klasifikasi dokumen

berbasis teks, dan bagaimana sistem ini dapat memberikan aksesibilitas serta efisiensi yang lebih baik dibandingkan metode manual. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi praktis berupa sistem yang relevan dengan kebutuhan mahasiswa, tetapi juga mendukung transformasi digital dalam lingkungan akademik yang lebih luas

2. Metodologi Penelitian

Penelitian ini menggunakan pendekatan berbasis pengembangan perangkat lunak dengan beberapa tahapan utama yang dilakukan secara sistematis berdasarkan metode *Waterfall*. Berikut adalah penjelasan rinci untuk setiap tahapan :

1. Identifikasi masalah

Pada tahap ini, dilakukan analisis terhadap permasalahan utama yang dihadapi mahasiswa dalam mengakses referensi akademik. Observasi dilakukan untuk memahami kebutuhan sistem dan hambatan yang ada, seperti keterbatasan akses ke perpustakaan fisik serta ketiadaan sistem digital yang efektif untuk menyimpan dan mencari dokumen.

2. Studi literatur

Studi literatur dilakukan untuk mendukung perancangan dan pengembangan sistem. Penelusuran terhadap jurnal, buku, dan artikel ilmiah terkait repository digital serta algoritma *Naïve Bayes Classifier* memberikan landasan teori yang kuat. Informasi yang dikumpulkan mencakup prinsip kerja algoritma NBC, metode pengelolaan data digital, serta studi kasus implementasi repository di institusi lain.

3. Analisis kebutuhan

Analisis kebutuhan dilakukan untuk menentukan fitur-fitur yang harus dimiliki oleh sistem. Fitur utama yang dirancang meliputi unggah, pencarian dan klasifikasi dokumen serta fitur manajemen pengguna.

4. Perancangan sistem

Pada tahap ini, dilakukan perancangan arsitektur sistem, antarmuka pengguna, dan basis data. Use case diagram digunakan untuk menggambarkan interaksi antara pengguna dan sistem, sementara perancangan basis data dilakukan untuk memastikan data dapat dikelola dengan baik. Antarmuka dirancang dengan mempertimbangkan kemudahan penggunaan oleh mahasiswa.

5. Implementasi Sistem

Sistem Implementasi dilakukan dengan membangun sistem menggunakan bahasa pemrograman PHP dan database MySQL. Framework CodeIgniter 3 digunakan untuk mempercepat pengembangan. Proses implementasi mencakup:

- Pengkodean fitur unggah dan pencarian dokumen.
- Integrasi algoritma *Naïve Bayes Classifier* untuk klasifikasi dokumen berdasarkan judul.
- Pengujian fungsionalitas antarmuka pengguna.

6. Pengujian Sistem

Pengujian dilakukan untuk memastikan sistem berfungsi sesuai dengan kebutuhan. Tahapan pengujian meliputi:

- Pengujian Fungsional: Memastikan setiap fitur sistem berjalan dengan baik.
- Pengujian Algoritma *Naïve Bayes* : Algoritma ini memanfaatkan prinsip probabilitas untuk memprediksi kategori dokumen berdasarkan data pelatihan. Dataset yang digunakan terdiri dari 200 dokumen, dengan 180 dokumen untuk data training dan 20 dokumen untuk data testing. Proses klasifikasi dimulai dengan *preprocessing* teks seperti tokenisasi, penghapusan kata-kata umum (*stopwords*),

dan stemming. Setelah itu, probabilitas setiap kata dalam dokumen dihitung untuk menentukan kategori dokumen berdasarkan mata kuliah. Rumus yang digunakan dalam algoritma Naïve Bayes adalah sebagai berikut :

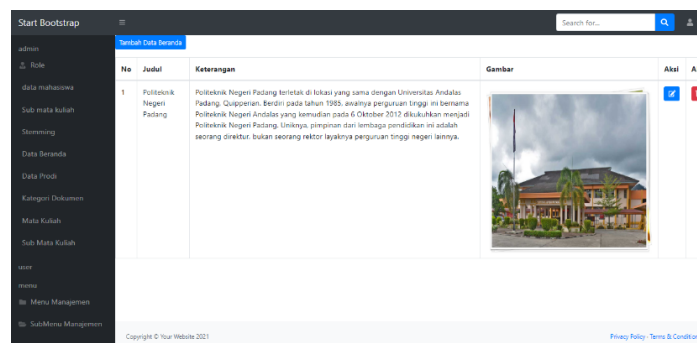
$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

dimana $P(H|X)$ adalah nilai probabilitas kelas H berdasarkan data X (*posterior probability*). $P(X|H)$ adalah nilai probabilitas data X diberikan kelas H (*likelihood*). $P(H)$ merupakan probabilitas awal kelas H (*prior probability*) dan $P(X)$ adalah probabilitas awal data X (*evidence*).

3. Hasil dan Pembahasan

3.1. Implementasi Sistem

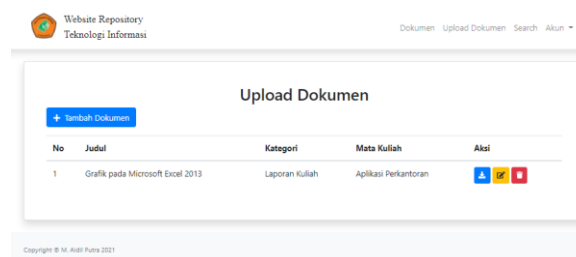
Sistem dirancang sebagai aplikasi berbasis web dengan antarmuka yang ramah pengguna. Antarmuka dirancang menggunakan framework CSS seperti Bootstrap untuk memastikan aksesibilitas dan responsivitas pada berbagai perangkat. Sistem dikembangkan dengan 2 level hak akses yakni administrator sistem dan user mahasiswa. Gambar 1 berikut ini memperlihatkan halaman utama dari administrator sistem yang terdiri dari beberapa menu yakni : data mahasiswa, sub mata kuliah, stemming, data beranda, data prodi, kategori dokumen, mata kuliah, sub mata kuliah, user, menu, menu manajemen, sub menu manajemen.



Gambar 1. Halaman Utama Administrator Sistem

3.2. Pengelolaan Dokumen

Sistem dirancang dengan dilengkapi dengan fitur unduh dan unggah dokumen. Fitur unggah dokumen seperti pada Gambar 2 berikut memungkinkan pengguna untuk menyimpan file dokumen akademik dalam format digital, seperti PDF. File yang diunggah disimpan dalam struktur direktori yang terorganisasi di server, dengan metadata disimpan dalam basis data.



Gambar 2. Menu Unggah Dokumen

3.3. Klasifikasi Dokumen dengan Naïve Bayes Classifier

Klasifikasi dilakukan berdasarkan judul dokumen untuk menentukan kategori mata kuliah yang sesuai. Langkah-langkah implementasi algoritma ini meliputi:

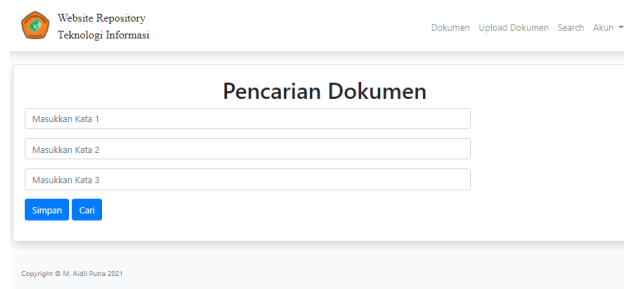
- a. Preprocessing Data
 - *Tokenisasi* : Memisahkan teks menjadi kata-kata individu.
 - *Stopword Removal* : Menghapus kata-kata umum yang tidak relevan, seperti "dari", "dan", "atau".
 - *Stemming* : Mengubah kata menjadi bentuk dasar. Setelah preprocessing, algoritma menghitung probabilitas setiap kata untuk menentukan mata kuliah yang paling mungkin sesuai dengan judul dokumen. Contoh kata "komputerisasi" menjadi "komputer".
- b. Perhitungan Probabilitas
Setelah preprocessing, algoritma menghitung probabilitas setiap kata dalam dokumen untuk menentukan kategori mata kuliah.
- c. Klasifikasi
Berdasarkan hasil probabilitas, algoritma memilih kategori dengan nilai probabilitas tertinggi sebagai hasil klasifikasi dokumen. Contoh: Dokumen berjudul "Dasar Jaringan Komputer" menghasilkan probabilitas tertinggi untuk kategori "Jaringan Komputer".
- d. Hasil Klasifikasi
Dari pengujian menggunakan 200 dokumen (180 data training dan 20 data testing), algoritma berhasil mengklasifikasikan 18 dari 20 dokumen dengan benar, memberikan tingkat akurasi sebesar 90%, seperti terlihat pada Tabel 1 berikut ini :

Tabel 1. Hasil Pengujian

No	Judul Dokumen	Kategori Asli	Hasil Klasifikasi	Status
1	Dasar Jaringan Komputer	Jaringan Komputer	Jaringan Komputer	Benar
2	Algoritma dan Pemrograman	Algoritma Pemrograman	Algoritma Pemrograman	Benar
3	Analisis Sistem Basis Data	Basis Data	Basis Data	Benar
...	...	...	...	...
20	Aplikasi Perkantoran Modern	Aplikasi Perkantoran	Aplikasi Perkantoran	Benar

3.4. Pencarian Dokumen

Sistem dilengkapi dengan fitur pencarian berbasis kata kunci yang terintegrasi dengan hasil klasifikasi seperti terlihat pada Gambar 3. Pengguna dapat mencari dokumen dengan memasukkan kata kunci tertentu, dan sistem akan menampilkan daftar dokumen yang relevan berdasarkan kategori mata kuliah yang dihasilkan oleh algoritma Naïve Bayes.



Gambar 3. Menu Pencarian Dokumen

4. Kesimpulan

Penelitian ini berhasil mengembangkan sistem repository berbasis website yang dirancang untuk meningkatkan aksesibilitas dan efisiensi dalam pengelolaan dokumen akademik. Sistem ini memanfaatkan algoritma Naïve Bayes Classifier untuk mengklasifikasikan dokumen berdasarkan kategori mata kuliah dengan tingkat akurasi mencapai 90%. Implementasi metode *Waterfall* memastikan setiap tahap pengembangan dilakukan secara terstruktur, mulai dari analisis kebutuhan hingga pengujian.

Namun, penelitian ini memiliki beberapa batasan. Dataset yang digunakan masih terbatas pada 200 dokumen, sehingga hasil klasifikasi mungkin kurang representatif jika diterapkan pada skala data yang jauh lebih besar. Selain itu, fokus penelitian hanya pada klasifikasi berbasis judul dokumen tanpa mempertimbangkan isi dokumen secara keseluruhan. Hal ini dapat membatasi tingkat akurasi jika dokumen memiliki judul yang kurang informatif.

Untuk penelitian lanjutan, disarankan untuk memperluas dataset yang digunakan agar hasil klasifikasi lebih representatif. Penelitian juga dapat mengintegrasikan analisis isi dokumen (*content-based analysis*) untuk meningkatkan akurasi klasifikasi. Selain itu, pengembangan sistem dapat diperluas dengan menambahkan fitur rekomendasi dokumen atau integrasi dengan sistem akademik yang lebih luas guna mendukung kebutuhan pengguna secara menyeluruh.

References

- [1] A. Wibawa, A. Kurniawan, D. Murti, R. P. Adiperkasa, S. Putra, S. Kurniawan, and Y. Nugraha, "Naïve Bayes Classifier for Journal Quartile Classification," *International Journal of Recent Contributions from Engineering, Science & IT (iJES)*, vol. 7, pp. 91–91, Jun. 2019, doi: 10.3991/ijes.v7i2.10659.
- [2] R. Achmad and A. S. Girsang, "Implementation of Naive Bayes Classifier Algorithm in Classification of Civil Servants," *Journal of Physics: Conference Series*, vol. 1485, no. 1, p. 012018, Mar. 2020, doi: 10.1088/1742-6596/1485/1/012018.
- [3] R. Adrian, M. A. J. S. Perdana, Asroni, and S. Riyadi, "Applying the Naive Bayes Algorithm to Predict the Student Final Grade," *Emerging Information Science and Technology*, vol. 1, no. 2, pp. 49–57, 2020, doi: 10.18196/eist.127.
- [4] T. Hardiani, R. Noviyanti, and P. Putri, "Implementasi metode Naïve Bayes classifier untuk klasifikasi stunting pada balita," *Digital Transformation Technology*, vol. 4, no. 1, pp. 621–627, 2024, doi: 10.47709/digitech.v4i1.4481.
- [5] K. Kartarina, A. Azzahrah, H. Rismayati, B. F. Rahmiati, F. Fatimatussahra, and A. H. Husaini, "Klasifikasi gizi lansia menggunakan metode Naïve Bayes classifier," *JTIM : Jurnal Teknologi Informasi dan Multimedia*, 2024, doi: 10.35746/jtim.v6i2.502.

- [6] P. A. Astuti, A. R. Wibowo, E. K. Kartikadarma, E. R. Subhiyakto, N. A. S. Winarsih, and M. S. R. Rohman, "Penerapan metode Naïve Bayes classifier untuk klasifikasi sentimen pada judul berita," *LogicLink*, vol. 1, no. 1, pp. 1–12, 2024, doi: 10.28918/logiclink.v1i1.7684.
- [7] A. M. Putra, I. Muhammad, and L. Jundillah, "Analisis sentimen pengguna YouTube terhadap uang baru tahun emisi 2022 menggunakan metode Naïve Bayes classifier," *Adopsi Teknologi dan Sistem Informasi*, vol. 3, no. 1, pp. 17–27, 2024, doi: 10.30872/atasi.v3i1.1177.
- [8] B. Della, A. Ulya, and H. A. Azizah, "Analisis sentimen pada media sosial Instagram terhadap akun Presiden Joko Widodo menggunakan metode Naïve Bayes classifier," *Deleted Journal*, vol. 2, no. 2, pp. 67–83, 2024, doi: 10.54066/jptis.v2i2.1895.
- [9] F. M. Azzahra, S. N. Sulistiyowati, and A. Voutama, "Analisis sentimen terhadap respon perubahan nama Twitter menjadi 'X' menggunakan metode Naïve Bayes classifier," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 5, pp. 10987–10994, 2024, doi: 10.36040/jati.v8i5.10720.
- [10] M. A. Djamaludin, A. Triayudi, and E. Mardiani, "Analisis sentimen tweet KRI Nanggala 402 di Twitter menggunakan metode Naïve Bayes classifier," *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, 2022.
- [11] D. Jozu, Y. Yusra, M. Fikry, S. Agustian, and O. Lolo, "Klasifikasi sentimen terhadap topik pindah ibu kota negara pada Twitter menggunakan metode Naïve Bayes classifier," *Jurnal Sistem Komputer dan Informatika (JSON)*, 2024, doi: 10.30865/json.v5i3.7475.
- [12] R. M. Tomu, M. S. Noya, and N. L. Van Delsen, "Klasifikasi ketepatan waktu lulus mahasiswa jurusan matematika universitas pattimura menggunakan metode Naïve Bayes classifier," *Proximal*, vol. 7, no. 2, pp. 864–872, 2024, doi: 10.30605/proximal.v7i2.4153.
- [13] N. Suryani and E. Priyanti, "Optimasi Naïve Bayes dan Algoritma Genetika untuk Prediksi Penerimaan Beasiswa Pendidikan pada SMP Utama," *Jurnal Teknik Komputer AMIK BSI*, vol. V, no. 2, pp. 189–196, Aug. 2019, P-ISSN 2442-2436, E-ISSN 2550-0120.
- [14] A. Pebdika, R. Herdiana, and D. Solihudin, "Klasifikasi menggunakan metode Naïve Bayes untuk menentukan calon penerima PIP," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, pp. 452–458, Feb. 2023, ISSN 2598-828X.
- [15] I. Purnamasari, "Analisa klasifikasi loyalitas siswa lembaga pendidikan tari dengan metode Naïve Bayes," *Jurnal Ilmiah KOMPUTASI*, vol. 19, no. 1, pp. 59–68, Mar. 2020, P-ISSN 1412-9434, E-ISSN 2549-7227, doi: 10.32409/jikstik.19.1.2731.