

SKRIPSI

**SISTEM PENCARIAN DOKUMEN DENGAN
MENGUNAKAN ALGORITMA NAIVE BAYES
BIDANG STATISTIK DAN PERSANDIAN
DISKOMINFO PUSPEM BADUNG**



POLITEKNIK NEGERI BALI

Oleh :

Ni Luh Gita Kusumawati

NIM. 2115354022

**PROGRAM STUDI SARJANA TERAPAN
TEKNOLOGI REKAYASA PERANGKAT LUNAK
JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI BALI
2025**

ABSTRAK

Kemajuan teknologi informasi mendorong instansi pemerintah untuk mengelola data secara lebih efektif dan presisi. Di Diskominfo Puspem Badung, khususnya bagian Statistik dan Persandian, masih terdapat kendala karena kegiatan pengarsipan dan pencarian dokumen dilakukan secara manual, yang menurunkan produktivitas kerja. Penelitian ini difokuskan pada pengembangan sistem pencarian dan klasifikasi arsip digital berbasis web dengan algoritma Naive Bayes untuk mempercepat sekaligus menata proses pengelolaan dokumen. Metode pengembangan yang dipakai adalah Agile dengan dukungan HTML, JavaScript, dan MySQL. Sistem mampu memfasilitasi unggah dokumen PDF, mengelompokkan secara otomatis berdasar isi, serta menyajikan hasil pencarian melalui judul maupun kategori. Algoritma Naive Bayes dimanfaatkan untuk memperkirakan kategori dokumen berdasarkan kata-kata yang muncul di dalam teks. Pengujian dilakukan dengan pendekatan black box serta evaluasi kinerja sistem.

Hasil implementasi memperlihatkan bahwa sistem bekerja cukup baik dalam klasifikasi dengan nilai probabilitas tinggi pada dokumen relevan. Keunggulan sistem ada pada kemudahan penggunaan serta efektivitas pengelompokan, meskipun masih terbatas pada format file dan pengolahan teks lanjutan. Secara umum, penelitian ini memberikan kontribusi nyata dalam mendukung digitalisasi arsip pemerintah dan berpotensi dikembangkan di masa depan.

Kata kunci: pencarian dokumen, klasifikasi, naive bayes, arsip digital, sistem berbasis web

ABSTRACT

The advancement of information technology requires government institutions to manage data more efficiently and accurately. One of the challenges faced by Diskominfo Puspem Badung, particularly in the Statistics and Encryption Division, is the manual process of archiving and retrieving documents, which hampers work efficiency. This study aims to develop a web-based document search and classification system using the Naive Bayes algorithm to support faster and more structured digital archive management.

The system was developed using the Agile method, with HTML, JavaScript, and MySQL as the programming technologies. It allows users to upload PDF documents, automatically classify them based on their content, and display search results by title or category. The Naive Bayes algorithm is used to calculate the probability of a document belonging to a specific category based on the words it contains. Testing was conducted using black box methods and system performance parameter evaluation.

The implementation results show that the system can classify documents effectively, indicated by high probability values for relevant documents. The system offers advantages in terms of ease of use and document grouping efficiency, although it still has limitations in file format support and advanced text processing. Overall, the system contributes positively to the digitization of government archives and has potential for further development in the future.

Keywords: document retrieval, classification, naive bayes, digital archiving, web-based system

Daftar Isi

LEMBAR PERSETUJUAN UJIAN SKRIPSI	2
LEMBAR PENGESAHAN SKRIPSI	3
HALAMAN PERNYATAAN KEASLIAN KARYA SKRIPSI	4
ABSTRAK	5
KATA PENGANTAR	7
Daftar Isi	9
Daftar Gambar	10
Daftar Tabel	11
DAFTAR LAMPIRAN	12
BAB I PENDAHULUAN	13
1.1. Latar Belakang	13
1.2. Perumusan Masalah	14
1.3. Batasan Masalah	14
1.4. Tujuan Penelitian	15
1.5. Manfaat Penelitian	15
BAB II TINJAUAN PUSTAKA	16
2.1. Penelitian Sebelumnya	16
2.2. Landasan Teori	17
BAB III METODE PENELITIAN	22
3.1. Objek Dan Metode Penelitian	22
3.2. Analisis Kondisi Eksisting	24
3.3. Rancangan Penelitian	24
3.4. Pengujian Penelitian	30
3.5. Hasil Yang Diharapkan	31
BAB IV HASIL DAN PEMBAHASAN	32
4.1. Hasil Implementasi Sistem	32
4.2. Hasil Pengujian Sistem	38
4.3. Pembahasan Hasil Implementasi dan Pengujian	40
BAB V KESIMPULAN DAN SARAN	48
5.1. Kesimpulan	48
JADWAL KEGIATAN	49
DAFTAR PUSTAKA	50
LAMPIRAN	52
Lampiran 1. Form Bimbingan Skripsi	52
Lampiran 2. Pernyataan Telah Menyelesaikan Bimbingan Skripsi	54
Lampiran 3. Lembar Perbaikan.....	55

Daftar Gambar

Gambar 2.2.7 Simbol flowmap	18
Gambar 3.3.2 Use Case	25
Gambar 4.1.2 Tampilan Dashboard	32
Gambar 4.1.2 Form Unggah Dokumen.....	32
Gambar 4.1.2 Halaman Pencarian Dokumen.....	33
Gambar 4.1.2 Halaman Pencarian Berdasarkan kategori.....	34
Gambar 4.1.2 Laporan Hasil Klasifikasi.....	34

Daftar Tabel

Tabel 3.3.5 Dataset.....	31
Tabel 4.1.3 Tabel Dokumen.....	37
Tabel 4.1.3 Tabel Kategori.	38
Tabel 4.2.3 Tabel Hasil Klasifikasi.....	38
Tabel 4.2.1 Tabel Pengujian black -box.....	39
Tabel 4.3.2 tabel model.....	45

DAFTAR LAMPIRAN

Lampiran 1. Form Bimbingan Skripsi

Lampiran 2. Pernyataan Telah Menyelesaikan Bimbingan Skripsi

Lampiran 3. Lembar Perbaikan

BAB I

PENDAHULUAN

1.1. Latar Belakang

Perkembangan teknologi informasi pada era modern membawa dampak besar di berbagai sektor, termasuk administrasi pemerintahan. Penerapan teknologi informasi dan komunikasi memungkinkan data dikelola secara digital dengan lebih efisien. Namun, tantangan yang dihadapi instansi pemerintah, seperti Diskominfo

Puspem Badung pada bidang Statistik dan Persandian, adalah bagaimana memanfaatkan teknologi ini untuk mengelola arsip dengan cara yang cepat, akurat, dan terstruktur. Dengan semakin meningkatnya volume data, kebutuhan terhadap sistem pengarsipan yang efektif menjadi semakin penting.

Salah satu hambatan yang muncul adalah praktik pengarsipan manual yang masih digunakan. Meski terlihat sederhana, metode ini memiliki banyak kelemahan, seperti memakan waktu lama dalam pencarian, rentan kesalahan manusia, dan tidak efisien untuk pengambilan keputusan berbasis data. Seiring bertambahnya jumlah arsip, pengelolaan manual semakin sulit dilakukan. Kondisi ini menegaskan perlunya sistem digital yang mampu mempercepat pencarian, klasifikasi, dan pengelolaan dokumen. Kelemahan ini tentunya menghambat alur kerja yang lebih cepat dan dapat mengganggu proses pengambilan keputusan yang membutuhkan data yang tepat dan cepat. Solusi lain yang muncul adalah peningkatan volume arsip yang harus dikelola. Seiring dengan berkembangnya jumlah kegiatan pemerintah yang dilakukan, banyak data yang terus bertambah dan terkadang sulit untuk diakses atau dikelola secara efisien. Tanpa sistem pengarsipan yang baik, data yang tersimpan dalam bentuk fisik atau bahkan digital yang tidak terorganisir dengan baik dapat menyebabkan infeksi, kebingungan dalam pencarian data, dan akhirnya menyebabkan keputusan pengambilan yang kurang optimal[1].

Permasalahan ini menggaris bawahi pentingnya sistem monitoring surat yang mampu mempercepat proses pencarian, pengarsipan, dan pengelolaan surat. Dengan penerapan metode *Naive Bayes*, penelitian ini bertujuan untuk mengembangkan sistem berbasis web yang dapat mengklasifikasikan surat masuk berdasarkan kategori tertentu dikarenakan manusia tidak bisa memilah data yang terlalu banyak[2].

Penelitian yang dilakukan oleh Ilhamsah Kasim, Yuyun, dan Supriadi Sahibu menggunakan algoritma *Naive Bayes* untuk meningkatkan efisiensi dan akurasi klasifikasi surat. Algoritma *Naive Bayes* adalah metode probabilistik dan statistik yang sederhana namun sangat efektif untuk mengklasifikasikan data berdasarkan pengalaman masa lalu. Dengan pendekatan ini, model dapat memprediksi peluang kategori yang sesuai untuk dokumen baru berdasarkan pola yang telah dipelajari dari data sebelumnya. Penelitian ini menganalisis sebanyak 1.036 dokumen surat yang terdiri dari 80% data untuk pelatihan (*training*) dan 20% untuk pengujian (*testing*). Dalam pengujian ini, nilai presisi, perolehan, dan akurasi dihitung menggunakan matriks konfusi untuk mengukur kinerja model. Hasilnya, sistem yang dibangun mampu mengklasifikasikan surat dengan nilai *Correctly Classified Instances* sebesar 86,24% dan nilai presisi serta recall sebesar 86%[3].

Sistem pencarian dokumen dengan *algoritma Naive Bayes* ini diharapkan dapat memberikan solusi bagi berbagai permasalahan yang dihadapi Diskominfo Kabupaten Badung, mulai dari pengelolaan data yang lebih terstruktur hingga peningkatan efisiensi pencarian dokumen. Dengan penerapan teknologi ini, Diskominfo dapat mendukung aksesibilitas dokumen yang lebih baik, meningkatkan kualitas pelayanan publik, dan mewujudkan visi Kabupaten Badung dalam mendukung transformasi digital yang menyeluruh[2].

1.2. Perumusan Masalah

Berdasarkan latar belakang tersebut, permasalahan yang dirumuskan dalam penelitian ini adalah:

1. Bagaimana penerapan sistem pengelolaan arsip berbasis web dengan algoritma *Naive Bayes* untuk membantu pengelompokan dan pencarian dokumen?

1.3. Batasan Masalah

1. Sistem pencarian dikembangkan menggunakan HTML dan basis data MySQL.
2. Penelitian terbatas pada bidang Statistik dan Persandian Diskominfo Badung.
3. Arsip hanya berupa dokumen PDF, tidak mencakup multimedia atau bahasa asing.
4. Fokus penelitian pada efisiensi waktu, klasifikasi, dan pencarian dokumen.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Penelitian ini bertujuan menghasilkan penerapan algoritma *Naive Bayes* yang dapat mempercepat proses pengelompokan dan pencarian dokumen sehingga informasi dapat diakses secara lebih cepat, akurat, dan efisien.

1.5. Manfaat Penelitian

Manfaat dari penulis ini diharapkan bisa membantu mengelola surat dan dokumen dengan lebih cepat, mudah, dan akurat. Dengan sistem berbasis web, proses seperti mengelompokkan, mencari, dan menyimpan dokumen menjadi lebih praktis dan otomatis. Hal ini dapat mengurangi kesalahan manual, menghemat waktu, dan mempermudah akses dokumen kapan saja.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Penelitian ini berhasil mengembangkan sistem pencarian dan klasifikasi dokumen berbasis web menggunakan algoritma Naive Bayes. Sistem mampu mengotomatisasi proses klasifikasi berdasarkan isi teks dengan akurasi yang baik, sehingga meningkatkan efisiensi dan mengurangi kesalahan dalam pengelolaan arsip digital. Fitur pencarian berdasarkan judul dan kategori juga membantu mempercepat akses informasi. Dengan demikian, sistem ini efektif mendukung transformasi digital di lingkungan pemerintahan.

5.2. Saran

Merujuk pada hasil implementasi yang telah dilakukan, masih terdapat beberapa aspek yang perlu disempurnakan agar sistem dapat berfungsi secara lebih efektif. Agar lebih optimal, sistem perlu dikembangkan untuk mendukung format file lain seperti DOCX atau TXT. Selain itu, integrasi teknologi OCR dapat dilakukan agar sistem mampu membaca file hasil scan. Pengembangan antarmuka yang lebih interaktif dan responsif juga menjadi poin penting. Penambahan fitur pelatihan model otomatis dan ekspor hasil klasifikasi diharapkan dapat mendukung fleksibilitas operasional di masa depan.

DAFTAR PUSTAKA

- [1] A. Algiffary, "Implementasi Machine Learning dengan Algoritma Naive Bayes Terhadap Sistem Informasi Pelayanan Pemberkasan Kepegawaian di BKPSDM Kota Palembang," *Indones. J. Comput. Sci.*, vol. 12, no. 3, pp. 1272–1281, 2023, doi: 10.33022/ijcs.v12i3.3227.
- [2] M. D. Trisetiyo and J. S. Wibowo, "Klasifikasi Surat Menggunakan Metode Naïve Bayes Pada Sistem Informasi Manajemen Surat," *Pros. SENDI*, pp. 978–979, 2019.
- [3] I. Kasim, S. Sahibu, K. Kementerian Agama Propinsi Sulawesi Selatan, and S. Handayani Makassar, "Klasifikasi Surat Digital Menggunakan Algoritma Naïve Bayes," *Klasifikasi Surat Digit. Menggunakan Algoritma Naïve Bayes*, vol. 13, no. 2, pp. 57–62, 2020.
- [4] J. F. Indey, "Implementasi Algoritma Naïve Bayes dalam Sistem Pengarsipan Surat Berbasis AI di GPI Papua Klasis Mimika Papua Tengah," no. 45, 2024.
- [5] Y. Dolot, P. Tinno, D. Rompas, and V. P. Rantung, "Implementasi Text Mining Pada Aplikasi Pengarsipan Berbasis Web Menggunakan Algoritma Naïve Bayes," *J. Educ. Method Technol.*, vol. 1, no. 1, p. 28, 2023.
- [6] S. Hadi Wijoyo and A. Dwi Herlambang, "Algoritma Naïve Bayes Untuk Klasifikasi Sumber Belajar Naïve Bayes Algorithm for Text Based Learning Resources Classification in Productive Subject At Information," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 4, pp. 431–436, 2019, doi: 10.25126/jtiik.201961323.
- [7] D. Irianti, "Rancangan Aplikasi Pada Puskesmas Cerenti Dalam Persediaan Obat Menggunakan Metode Naive Bayes," *J. Perencanaan, Sains, Teknol. dan Komput.*, vol. 3, no. 2, pp. 418–427, 2020.
- [8] V. Rizqiyani, A. Mulwinda, and R. D. M. Putri, "Klasifikasi Judul Buku dengan Algoritma Naive Bayes dan Pencarian Buku pada Perpustakaan Jurusan Teknik Elektro," *J. Tek. Elektro*, vol. 9, no. 2, pp. 60–65, 2017, [Online]. Available: <https://journal.unnes.ac.id/nju/index.php/jte/article/view/11658/7332>
- [9] Purwono, "Arti dan Sejarah Istilah Dokumentasi," *Dasar-Dasar Dokumentasi*, pp. 2–32, 2014, [Online]. Available: <https://pustaka.ut.ac.id/lib/pust2241-dasar-dasar-dokumentasi-edisi-2/>
- [10] Chandra Christian and Apriade Voutama, "Implementasi Aplikasi Antrian Pencucian Mobil Berbasis Web Menggunakan Php, Javascript, Html, Css Dan Uml," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 2243–2248, 2024.
- [11] D. Irawan and Z. Novianto, "Perancangan E-Learning Pada Sman 1 Kota Lubuklinggau Menggunakan Framework Codeigniter (Ci)," *J. Digit. Teknol. Inf.*, vol. 3, no. 2, p. 53, 2020, doi: 10.32502/digital.v3i2.2690.
- [12] J. S. Pasaribu, "Penerapan Framork Yii Pada Pembangunan Sistem Ppdb Smp Bppi Baleendah Kabupaten Bandung," *J. Ilm. Teknol. Infomasi Terap.*, vol. 3, no. 2, pp. 154–163, 2017, doi: 10.33197/jitter.vol3.iss2.2017.132.
- [13] R. F. Ramadhan and R. Mukhaiyar, "Penggunaan Database Mysql dengan Interface PhpMyAdmin sebagai Pengontrolan Smarthome Berbasis Raspberry Pi," *JTEIN J. Tek. Elektro Indones.*, vol. 1, no. 2, pp. 129–134,

- 2020, doi: 10.24036/jtein.v1i2.55.
- [14] L. Setiyani, “Desain Sistem : Use Case Diagram Pendahuluan,” *Pros. Semin. Nas.Inov. Adopsi Teknol. 2021*, no. September, pp. 246–260, 2021, [Online]. Available: <https://journal.uui.ac.id/AUTOMATA/article/view/19517>
- [15] M. I. Burhan, F. Nawir, and K. N. Salam, “Pengembangan Sistem Tracer Study Menggunakan Agile Development Methods Pada Ibk Nitro,” *JURSIMA (Jurnal Sist. ...)*, vol. 10, no. November, pp. 160–170, 2022, doi: 10.47024/js.v10i3.
- [16] L. Trisnawati and D. Setiawan, “Sistem Monitoring Kegiatan Mahasiswa Menggunakan Metode Agile Development,” *JOISIE J. Inf. Syst. Informatics Eng.*, vol. 6, no. 1, pp. 49–57, 2022.