
Penerapan Metode Naïve Bayes Untuk Pengelolaan Status Retensi Arsip

Divia Rizqa Maharani¹, Leny Tritanto Ningrum^{2*}, Lis Utari³

^{1,2}Sistem Informasi/Universitas Binaniaga Indonesia.

¹Email: divarizqa167@gmail.com

²Email: lenytrinie@unbin.ac.id

³Teknik Informatika/Universitas Binaniaga Indonesia.

Email: lisutari@stikombinaniaga.ac.id

*)Corresponding Author

ABSTRACT

Archival management is an important aspect of information management in various institutions, one of which is determining the retention status of archives. Determining the retention status of archives is still not accurate and effective, so there is a possibility of human error or inaccurate errors when determining the retention status of archives. One method that is appropriate for the problem is the classification method. The classification method used is Naive Bayes which will help archive operators. The purpose of this research is to classify the retention status of archives and create an application prototype. This approach allows archive managers to make more informed and efficient decisions about how to store or destroy documents, so they can reduce the risk of errors and optimize the use of storage space. The variables used are archive type, index type, active retention time, inactive retention time and class predictor, which is the archive retention status. Prototype application to determine archive retention status using the Naive Bayes Algorithm method. The test results using confusion matrix showed an accuracy of 96%, the accuracy of the user test questionnaire results was 100%, and the accuracy of the system expert test questionnaire results was 92.86%.

Keywords: Accuracy, Archives Management, Classification, Naïve Bayes, Prototype.

ABSTRAK

Pengelolaan kearsipan adalah aspek penting dari manajemen informasi di berbagai lembaga, salah satu pengelolaan kearsipan adalah menentukan status retensi arsip. Penentuan status retensi arsip masih belum akurat dan belum efektif, sehingga ada kemungkinan kesalahan manusia atau kesalahan yang kurang akurat saat menentukan status retensi arsip. Salah satu metode yang tepat untuk permasalahan adalah metode klasifikasi. Metode klasifikasi yang digunakan adalah Naive Bayes yang akan membantu operator kearsipan. Tujuan penelitian ini untuk mengklasifikasikan status retensi arsip dan membuat prototype aplikasi. Pendekatan ini memungkinkan pengelola kearsipan untuk membuat keputusan yang lebih tepat dan efisien tentang bagaimana menyimpan atau memusnahkan dokumen, sehingga mereka dapat mengurangi risiko kesalahan dan mengoptimalkan penggunaan ruang penyimpanan. Variabel yang digunakan yaitu jenis arsip, jenis indeks, waktu retensi aktif, waktu retensi in-aktif dan prediktor kelas, yaitu status retensi arsip. Prototype aplikasi untuk menentukan status retensi arsip dengan menggunakan metode Algoritma Naive Bayes. Hasil uji menggunakan confusion matrix menunjukkan akurasi sebesar 96%, akurasi dari hasil kuesioner uji pengguna sebesar 100%, dan akurasi hasil kuesioner uji ahli sistem sebesar 92,86%.

Keywords: Akurasi, Klasifikasi, Naïve Bayes, Pengelolaan Kearsipan, Prototype.

A. PENDAHULUAN

1. Latar Belakang

Dalam Lembaga pemerintahan terdapat bidang yang terkait dengan pengelolaan kearsipan. Setiap kegiatan yang dilakukan dalam suatu organisasi mengarah pada terciptanya arsip, sehingga menambah jumlah arsip. Mengingat banyaknya jumlah arsip yang dimiliki suatu organisasi, maka hal tersebut harus dicapai melalui pengelolaan arsip yang memadai dan akurat agar arsip yang diciptakan dapat memberikan nilai manfaat, yaitu informasi yang

menjadi pedoman pengambilan keputusan dalam kegiatan organisasi dan lembaga harus dikelola. Proses pengelolaan kearsipan dapat ditingkatkan dengan menggunakan teknologi seperti sistem informasi pengelolaan kearsipan yang berbasis teknologi.

Arsip merupakan informasi yang disimpan dalam berbagai media untuk pelaksanaan kehidupan bermasyarakat, berbangsa, dan bernegara. Dalam organisasi atau lembaga, baik pemerintahan maupun swasta arsip sangat penting, karena arsip tersebut berisi keterangan atau informasi yang dapat membantu organisasi menjalankan operasinya atau membuat keputusan yang akurat tentang masalah yang sedang dihadapi, sehingga arsip tersebut dapat nilai guna, proses arsip, waktu aktif, dan waktu inaktif untuk mendapatkan rekomendasi klasifikasi status retensi arsip. Nilai guna arsip menentukan masa penyimpanan arsip.

Pengelolaan kearsipan ini dapat menerapkan Jadwal Retensi Arsip (JRA) di tempat kerja dapat membantu pengelola kearsipan terutama dalam menentukan status retensi arsip yang bertujuan untuk menjaga keberadaan arsip dan memastikan bahwa rencana tersebut dilaksanakan dengan akurat yang akan meningkatkan pengelolaan arsip agar mencegah penumpukan dokumen atau arsip yang sudah tidak terpakai atau tidak relevan. Jadwal Retensi Arsip (JRA) adalah daftar yang mencantumkan jenis arsip, jangka waktu penyimpanan atau retensi, dan keterangan yang berisi saran tentang bagaimana jenis arsip harus dimusnahkan, dinilai kembali, atau dipermanenkan. Dalam pengelolaan status retensi arsip mengatasi kendala dalam penentuan status retensi arsip yang belum akurat dan belum efektif.

Arsip yang tidak dikelola sering diabaikan oleh pengelola karena dianggap tidak berguna atau tidak penting karena berbagai alasan, seperti sumber daya manusia dan kualifikasi akademis yang kurang, pemahaman yang kurang tentang pentingnya kearsipan dan standar operasional prosedur (SOP) yang dikeluarkan oleh pihak yang berwenang, dan peningkatan volume. Rencana penyimpanan arsip adalah aturan yang ditetapkan oleh pihak yang berwenang untuk menentukan lokasi arsip, biaya yang diperlukan, dan tingkat kehati-hatian yang diperlukan untuk menentukan apakah arsip layak untuk disimpan secara permanen.

Dalam penelitian ini metode naïve bayes memudahkan pengelolaan dan menerapkan sistem pengelolaan kearsipan dengan menerapkan metode naïve bayes untuk penentuan status retensi arsip. Data yang digunakan berasal dari jenis arsip yang dinilai kembali, yang merupakan arsip yang belum ditentukan rekomendasinya, yaitu arsip musnah atau permanen.

2. Permasalahan

Dalam pengelolaan arsip dan ketersediaan arsiparis atau petugas arsip yang memiliki pemahaman tentang arsip sangat penting dalam menerapkan Jadwal Retensi Arsip (JRA) di tempat kerja dapat membantu pengelola kearsipan terutama dalam menentukan status retensi arsip, namun beberapa institusi mungkin tidak memiliki Jadwal Retensi Arsip (JRA) secara keseluruhan, sumber daya manusia, terutama di bidang kearsipan, biaya pengelolaan file meningkat seiring dengan jumlah file dan masalah pemusnahan arsip, termasuk prosedur pemusnahan arsip dan persetujuan untuk rencana retensi arsip. Proses pengelolaan arsip yang belum efektif serta membutuhkan waktu yang lama dan perlu ditangani secara profesional dan sungguh-sungguh, sesuai dengan peraturan perundang-undangan yang berlaku untuk meningkatkan efisiensi dan efektivitas pengelolaan arsip. Rencana retensi arsip hanya berlaku untuk arsip yang dibuat setelah keputusan dikeluarkan. Penentuan status retensi arsip adalah proses dalam manajemen arsip, yang bertujuan untuk menentukan apakah suatu arsip harus disimpan untuk jangka waktu tertentu atau dimusnahkan. Penentuan status retensi arsip yang belum akurat dan belum efektif, sehingga retensi arsip dilakukan untuk mencegah penumpukan dokumen atau arsip yang sudah tidak terpakai maupun tidak relevan, tentu saja hal tersebut membutuhkan pengelola atau operator kearsipan yang terlatih di bidang tersebut agar tidak terjadinya kesalahan yang kurang akurat pada saat penentuan status retensi arsip. Metode Naive Bayes memudahkan pengelolaan kearsipan dengan mengklasifikasikan status retensi arsip. Proses pengelolaan kearsipan dapat ditingkatkan dengan menggunakan teknologi

seperti sistem informasi pengelolaan kearsipan yang berbasis teknologi. Berikut data pendukung dari permasalahan tersebut:

Tabel 1 Data Pendukung Permasalahan

No	Kode Klasifikasi	Jenis Arsip	Deskripsi Arsip	Kurun Waktu	Indeks / Kata Kunci	Jumlah Item	Waktu Retensi Aktif	Waktu Retensi Inaktif	Retensi/ Jangka Simpan	Status Retensi Arsip
1	KU.03.02	KEUANGAN	Laporan Keuangan Periode T.A 2021 (Unaudited)	2022	Laporan Keuangan	1 Item	2	8	2032	Permanen
2	KP.14.02	KEPEGAWAIAN	Pemohonan Pensiun	2023	Pemohonan	21 Item	1	2	2026	Musnah
3	KS.00.01	KERJA SAMA	Tanggapan Terhadap Permohonan Tindak Lanjut Komitmen Kerja Sama Pendidikan RI-Papua Nugini	2023	Tindak Lanjut dan Evaluasi	2 Item	2	3	2028	Musnah
4	KS.00.01	KERJA SAMA	Pemantauan dan Evaluasi Program Kerja Sama Kemendikbudristek dengan KRIVET	2023	Tindak Lanjut dan Evaluasi	4 Item	2	3	2028	Permanen
5	KS.00.01	KERJA SAMA	Pemantauan dan Evaluasi Program NIHONGO Partners Gelombang 18	2023	Tindak Lanjut dan Evaluasi	3 Item	2	3	2028	Permanen
6	TU.01.01	KETATAUSAHAAN	Penilaian Arsip Usul Musnah 2021	2021	Usul	8 Item	2	3	2026	Musnah
7	TU.01.04	KETATAUSAHAAN	Penyusutan/Pemindahan Arsip Inaktif Tahun 2021	2021	Pemindahan/ Penyusutan Arsip dan Penataan	3 Item	2	2	2025	Musnah
8	KP.16	KEPEGAWAIAN	Surat Pernyataan a.n. Eva Komalasari, S.T., M.M.	2023	Surat Pernyataan	2 Item	1	9	2033	Musnah
9	TU.01.04	KETATAUSAHAAN	Pelaksanaan Pemindahan/ Penyusutan Arsip Inaktif Tahun 2023	2023	Pemindahan/ Penyusutan Arsip dan Penataan	4 Item	2	2	2027	Musnah
10	TU.01.01	KETATAUSAHAAN	Penilaian Arsip Usul Musnah 2020	2020	Usul	1 Item	2	3	2025	Musnah
11	KP.11.00	KEPEGAWAIAN	Pengiriman Data Pegawai	2020	Pendataan	1 Item	1	4	2025	Musnah
12	KP.04.08	KEPEGAWAIAN	Usul Perceraian PNS	2020	Usul	1 Item	2	3	2025	Musnah
13	KU.00.00	KEUANGAN	Usul Pejabat Perbendaharaan	2020	Usul	1 Item	2	3	2025	Musnah
14	PR.05.04	PERENCANAAN DAN PROGRAM	Evaluasi Kinerja Pegawai 2020	2020	Tindak Lanjut dan Evaluasi	1 Item	2	3	2025	Permanen
15	TU.01.02	KETATAUSAHAAN	Pendataan dan Pengendalian Arsip 2020	2020	Pendataan	1 Item	2	3	2025	Musnah
16	PR.00.02	PERENCANAAN DAN PROGRAM	Penyusunan Rencana Strategis Setditjen Pendidikan Vokasi	2020	Renstra	12 Item	2	3	2025	Permanen
17	KP.11.03	KEPEGAWAIAN	0007/D1/KP/2021 Hal Usul Mutasi Jabatan Pegawai dan Penerbitan Surat Keputusan	2021	Usul	2 Item	1	2	2024	Musnah

No	Kode Klasifikasi	Jenis Arsip	Deskripsi Arsip	Kurun Waktu	Indeks / Kata Kunci	Jumlah Item	Waktu Retensi Aktif	Waktu Retensi Inaktif	Retensi/ Jangka Simpan	Status Retensi Arsip
			Tunjangan Kinerja di lingkungan BBPPMV BMTI							
18	KU.03.02	KEUANGAN	Kegiatan Penyusunan Laporan Keuangan Triwulan III, Tanggal 20 s.d. 22 Oktober 2022	2022	Laporan	2 Item	2	5	2029	Musnah
19	TU.01.04	KETATAUSAHAAN	Pemusnahan Arsip Eks Arsip Ditjen PAUD dan Dikmas	2020	Pemusnahan Arsip	10 Item	2	3	2025	Musnah
20	PR.07.03	PERENCANAAN DAN PROGRAM	Kegiatan Penyusunan Program Kerja, Tanggal 16 s.d. 18 Oktober 2022	2022	Program Kerja	2 Item	2	3	2027	Musnah
21	TU.00.01	KETATAUSAHAAN	Agenda Surat Masuk Tahun 2020	2021	Surat	1 Item	2	3	2026	Musnah
22	TU.01.02	KETATAUSAHAAN	Kegiatan Penilaian Arsip Usul Musnah, Tanggal 6 s.d. 7 Juni 2022	2022	Usul	16 Item	2	3	2027	Musnah
23	KS.00.01	KERJASAMA	Kegiatan Rapat Focus Group Discussion (FGD) Praktis Mengajar, Tanggal 8 s.d. 10 Maret 2023	2023	Kegiatan FGD	2 Item	2	3	2028	Permanen
24	KS.06.02	KERJASAMA	Rekomendasi izin belajar kursus Bahasa Indonesia bagi WNA pada LKP "IMLAC", a.n. Kim Kyungsook	2022	Rekomendasi	7 Item	2	4	2028	Musnah
25	KS.00.01	KERJASAMA	Penyampaian Informasi Aus4ASEAN Short Course Programme	2023	Informasi	8 Item	2	3	2028	Permanen
...	...	...	...	...	...	...	...	...	...	...
223	PR.00.02	PERENCANAAN DAN PROGRAM	Kegiatan Penyusunan Renstra Setditjen Pendidikan Vokasi Tahun 2020-2021 pada hari Minggu s.d. Selasa, 6 s.d. 8 September 2020 di Swiss-Belhotel Bogor	2021	Renstra 2021-2022	2 Item	2	3	2026	Musnah

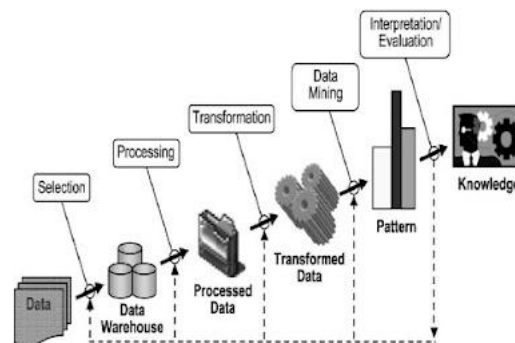
Pada tabel 1 diatas terdapat permasalahan yang terjadi pada data ke 3, dan data ke 223, dimana jenis arsip tersebut masih ada yang belum akurat dalam menentukan status retensi arsip karena petugas arsip seringkali bergantung pada pengetahuan mereka sendiri daripada mengikuti peraturan pedoman Jadwal Retensi Arsip (JRA) yang dapat menyebabkan kejanggalan dalam proses penentuan status arsip dan petugas arsip tidak selalu memiliki pemahaman tentang arsip yang memadai. Dalam memproses penentuan status retensi arsip adanya kesalahan dalam penilaian arsip yang disebabkan oleh komunikasi yang belum efektif karena monitoring dan evaluasi yang tidak dilakukan secara teratur dapat menyebabkan kesalahan dalam pelaksanaan Jadwal Retensi Arsip (JRA).

3. Tinjauan Pustaka

a. Data Mining

Menurut [1] Data mining merupakan proses mengekstraksi pola atau pengetahuan yang berguna dan menarik (non-trivial, implisit, dan sebelumnya belum diketahui kegunaannya) dari kumpulan data yang sangat besar. Menurut [2] Istilah data mining dan Knowledge Discovery In Databases (KDD) sering kali digunakan untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar, Tahapan knowledge discovery in databases (KDD) terdiri dari 6 tahapan [3] sebagai berikut:

- (1) Seleksi atau Pemilihan Data (*Data Selection*) merupakan proses yang membuat dan memilih himpunan data yang berkonsentrasi pada subset data sampel atau variabel di mana proses penemuan akan dilakukan;
- (2) Pembersihan Data (*Data Cleaning atau Pre-processing*) mencakup pembersihan data, memeriksa data yang dinilai tidak konsisten, dan mencari dan memperbaiki kesalahan data, seperti kesalahan cetak (typography);
- (3) Transformasi Data (*Data Transformation*) merupakan proses mencari berbagai fitur dalam proses data mining dengan mengubah data yang telah dipilih ke dalam model analitis dan mengubah model menjadi sesuai dengan analisis yang diharapkan dan diperlukan oleh algoritma data mining;
- (4) Pembentukan Model (*Data Mining*) merupakan proses menemukan pola dan informasi khusus dalam data yang telah dipilih sebelumnya dengan menggunakan teknik atau pendekatan tertentu, seperti klasifikasi, klustering, regresi, atau asosiasi;
- (5) Penginterpretasian dan Visualisasi (*Interpretation and Evaluation*) merupakan Pola informasi yang dihasilkan dari proses ini harus ditampilkan sehingga berbagai pihak yang berkepentingan dapat memahaminya;
- (6) Pengetahuan (*Knowledge*) merupakan langkah terakhir dari proses data mining, pola yang telah ditemukan untuk mendapatkan tindakan dipresentasikan melalui tahap evaluasi dalam sistem informasi dan perilaku organisasi perusahaan.



Gambar 1 Tahapan dalam knowledge discovery in databases (KDD) [2]

b. Crisp-dm

Menurut [4], terdapat siklus hidup CRISP-DM yang terdiri dari enam fase dan penjelasannya sebagai berikut;

- (1) *Business Understanding Phase* (Fase Pemahaman Bisnis) merupakan tahapan dilakukan dengan menentukan tujuan proyek dan kebutuhan dalam lingkup unit penelitian atau bisnis secara keseluruhan, serta menciptakan strategi awal untuk mencapai tujuan;
- (2) *Data Understanding Phase* (Fase Pemahaman Data) merupakan tahapan dilakukan pengumpulan data yang kemudian dianalisis untuk mendapatkan informasi tentang pola dan kualitas data;
- (3) *Data Preparation Phase* (Fase Persiapan Data) merupakan tahapan menganalisis dan menyiapkan data dengan menyesuaikan kasus pada beberapa variabel yang diperlukan untuk tahapan pemodelan;

- (4) *Modelling Phase* (Fase Pemodelan) merupakan tahapan yang dilakukan pemilihan dan pengaplikasian teknik yang menyesuaikan pemodelan dengan kasus serta permasalahan;
- (5) *Evaluation Phase* (Fase Pengujian) merupakan tahapan di mana model yang dirancang dievaluasi untuk mengidentifikasi masalah penting yang terkait dengan bisnis atau penelitian dan untuk mengevaluasi apakah model tersebut telah mencapai tujuan bisnisnya atau tidak.
- (6) *Deployment Phase* (Fase Penyebaran) merupakan tahapan di mana model data mining digunakan dengan melibatkan penerapan proses data mining pada perusahaan atau pembuatan laporan tentang hasilnya.

c. Confusion Matrix

Confusion matrix adalah tool yang digunakan untuk evaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah, confusion matrix dapat digunakan untuk menghitung berbagai performance metrics untuk mengukur kinerja model yang telah dikerjakan/dibuat. [5]

Tabel 2 Confusion Matrix [6]

Clasification	Predicted Class	
	Class = Yes	Class = No
Class = Yes	a (true positive-TP)	b (false p[ositive-FP)
Class = No	c (false negative-FN)	d (true negative-TN)

Beberapa performance metrics populer yang umum dan sering digunakan, accuracy, precision, dan recall. Accuracy merupakan rasio prediksi benar (positif dan negatif) dengan keseluruhan data. Dengan kata lain prediksi dengan nilai aktual. Nilai accuracy dapat diperoleh dengan persamaan:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{A + D}{A + D + B + C}$$

Keterangan:

A = Jika hasil prediksi positif dan data sebenarnya positif

B = Jika hasil prediksi negatif dan data sebenarnya positif

C = Jika hasil prediksi positif dan data sebenarnya negatif

D = Jika hasil prediksi negatif dan data sebenarnya negatif

4. Hipotesis

Metode Naive Bayes merupakan metode pengklasifikasian yang didasarkan pada probabilitas sederhana untuk digunakan ketika variabel penjelas saling bebas (independen). Beberapa penelitian menunjukkan bahwa hasil penelitian dengan metode Naive Bayes yang diperoleh [7] dari penelitian sebelumnya tentang Klasifikasi Surat Digital Menggunakan Algoritma Naïve Bayes mencapai nilai precision sebesar 86%, Recall 86 % dan Akurasi sebesar 76%. Pengelolaan Status Retensi Arsip dapat mengklasifikasikan dokumen yang paling akurat untuk dipermanenkan atau dimusnahkan, serta mengurangi kesalahan dalam penentuan status arsip, dan mencegah kehilangan informasi penting. Berdasarkan pemahaman sebelumnya, maka hipotesis dari penelitian & pengembangan ini dapat ditetapkan bahwa penerapan metode Naïve Bayes diduga dapat menentukan status retensi arsip.

B. METODE

Pendekatan yang efektif dalam penentuan retensi arsip dapat mengoptimalkan ruang penyimpanan, memastikan kepatuhan terhadap regulasi, dan meningkatkan efisiensi operasional. Variabel yang digunakan dalam penelitian ini ditentukan berdasarkan pada tujuan penelitian dalam menentukan status retensi arsip. Variabel yang digunakan meliputi jenis arsip, jenis indeks, waktu retensi aktif, dan waktu retensi in-aktif.

1. Algoritma Naïve Bayes

Algoritma Naive Bayes adalah algoritma sederhana karena setiap atributnya unik, sehingga setiap atribut dapat ikut serta dalam hasil akhir[8], Algoritma Naive Bayes mengasumsikan bahwa karakteristik tertentu dari suatu kelas tidak berhubungan dengan kelas lainnya dan dirancang untuk dapat memprediksi peluang masa depan berdasarkan informasi masa lalu ini dikenal sebagai teorema Bayes [9], adapun langkah-langkah dalam metode naïve bayes yaitu:

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

Keterangan:

X : Data yang memiliki class yang belum diketahui;

H : Hipotesis data merupakan kelas tertentu;

P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X, yang dikenal sebagai probabilitas posterior;

P(H) : Probabilitas hipotesis H, yang dikenal sebagai probabilitas prior;

P(X) : Probabilitas X;

P(X|H) : Probabilitas X berdasarkan kondisi hipotesis H.

2. Teknik Pengumpulan Data

a. Pengumpulan Dokumen

Pengumpulan dokumen untuk menemukan masalah dan kebutuhan data yang akan digunakan selama proses pengembangan aplikasi. Adapun dokumen yang dikumpulkan adalah data arsip selama 1 (satu) tahun, yang terdiri dari data arsip aktif dan in-aktif yang akan digunakan sebagai data training atau dasar dalam membuat confusion matriks.

b. Wawancara

Pada hari Rabu 15 November 2023, wawancara dilakukan dengan operator kearsipan. untuk mengetahui proses penentuan status retensi arsip. Adapun hasil wawancara menunjukkan bahwa bagian tersebut dalam penentuan status retensi arsip masih ada yang belum akurat dan belum efektif dalam penentuan status retensi arsip yang memerlukan waktu lama.

3. Teknik Analisis Data

Menurut Arikunto [10], Data kuantitatif yang berwujud angka-angka hasil perhitungan atau pengukuran dapat diproses dengan cara dijumlah, dibandingkan dengan jumlah yang diharapkan dan diperoleh persentase. Rumus persentase kelayakan:

$$\text{Persentase kelayakan (\%)} = \frac{\text{Skor yang didapatkan}}{\text{Skor yang maksimal}} \times 100\%$$

Berikut kategori kelayakan menurut Arikunto [10] yang menentukan nilai kelayakan produk yang dikembangkan.

Tabel 3 Kategori Kelayakan

Presentasi Pencapaian	Kategori Kelayakan
<21%	Sangat Tidak Layak
21% - 40%	Tidak Layak
41% - 60%	Cukup Layak
61% - 80%	Layak
81% - 100%	Sangat Layak

C. HASIL DAN PEMBAHASAN

1. HASIL

a. Tahap Pemahaman Bisnis (*Business Understanding*)

Penentuan status retensi arsip yang belum akurat dan belum efektif, sehingga retensi arsip dilakukan untuk mencegah penumpukan dokumen atau arsip yang sudah tidak terpakai maupun tidak relevan, tentu saja hal tersebut membutuhkan pengelola atau operator kearsipan yang terlatih di bidang tersebut agar tidak terjadinya kesalahan yang kurang akurat pada saat penentuan status retensi arsip. Penelitian ini bertujuan untuk memudahkan pengelolaan kearsipan dalam mengklasifikasikan status retensi arsip.

b. Tahap Pemahaman Data (*Data Understanding*)

Pada Fase Pemahaman Data menggunakan data arsip aktif dan in-aktif tahun 2023, total data Dari data arsip yang telah dikumpulkan, maka tahap berikutnya adalah melakukan pengolahan data. Pada proses pengolahan data mining yang dilakukan pada penelitian ini yaitu menggunakan data preparation dan data transformation.

2. PEMBAHASAN

a. Tahap Persiapan Data (*Data Preparation*)

Pada data asli dilakukan perubahan data informasi diantaranya yaitu kolom indek/kata kunci ditransformasikan menjadi kolom jenis indeks. Pada tabel 4 menunjukkan data yang telah melakukan proses persiapan data dan trasformasi data sebagai berikut:

Tabel 4 Persiapan Data

No	Jenis Arsip	Jenis Indeks	Waktu Retensi Aktif	Waktu Retensi Inaktif	Status Retensi Arsip
1	KEUANGAN	Laporan Keuangan	2	8	Permanen
2	KEPEGAWAIAN	Permohonan	1	2	Musnah
3	KERJA SAMA	Tindak Lanjut dan Evaluasi	2	3	Musnah
4	KERJA SAMA	Tindak Lanjut dan Evaluasi	2	3	Permanen
5	KERJA SAMA	Tindak Lanjut dan Evaluasi	2	3	Permanen
6	KETATAUSAHAAN	Usul	2	3	Musnah
7	KETATAUSAHAAN	Pemindahan/ Penyusutan Arsip dan Penataan	2	2	Musnah
8	KEPEGAWAIAN	Surat Pernyataan	1	9	Musnah
9	KETATAUSAHAAN	Pemindahan/ Penyusutan Arsip dan Penataan	2	2	Musnah
10	KETATAUSAHAAN	Usul	2	3	Musnah
...	...	...	...	...	...
214	KETATAUSAHAAN	Pemindahan/ Penyusutan Arsip dan Penataan	2	3	Musnah
215	KETATAUSAHAAN	Pemindahan/ Penyusutan Arsip dan Penataan	2	3	Musnah
216	KETATAUSAHAAN	Pemindahan/ Penyusutan Arsip dan Penataan	2	3	Musnah
217	KETATAUSAHAAN	Pemindahan/ Penyusutan Arsip dan Penataan	2	3	Musnah

No	Jenis Arsip	Jenis Indeks	Waktu Retensi Aktif	Waktu Retensi Inaktif	Status Retensi Arsip
218	KETATAUSAHAAN	Pemindahan/ Penyusutan Arsip dan Penataan	2	3	Musnah
219	KEUANGAN	Laporan	2	5	Permanen
220	KERJA SAMA	Rekomendasi	2	4	Musnah
221	KERJA SAMA	Rekomendasi	2	4	Musnah
222	KEPEGAWAIAN	Surat Pernyataan	1	9	Musnah
223	PERENCANAAN DAN PROGRAM	Renstra	2	3	Musnah

Setelah melakukan persiapan data, ada beberapa atribut yang tidak digunakan dalam proses pengolahan data dan proses ini mencakup pemilihan variabel yang sudah disesuaikan untuk diterapkan pada bagian modeling. variabel yang digunakan yaitu jenis arsip, jenis indeks, waktu retensi aktif, waktu retensi in-aktif dan prediktor kelas, yaitu status retensi arsip.

b. Tahap Pemodelan (*Modeling*)

Metode penyelesaian masalah penentuan status retensi arsip menggunakan metode Naive Bayes untuk menentukan status retensi arsip dimasa mendatang. Metode ini menghasilkan probabilitas yang akan digunakan untuk setiap kelas berdasarkan data yang akan diuji atau data testing. Data yang akan dipilih untuk menentukan status retensi arsip yaitu jenis arsip, jenis indeks, waktu retensi aktif, waktu retensi in-aktif, dan status retensi arsip. Data rekapitulasi Arsip Aktif dan In-aktif yang digunakan di atas berjumlah 223 data, untuk keperluan analisis metode naïve bayes diperlukan pembagian data menjadi data latih dan data uji. Berdasarkan hasil evaluasi komposisi pembagian data untuk data latih dan data uji dengan perbandingan 90% dan 10% mendapatkan hasil akurasi paling baik dibandingkan komposisi pembagian data lainnya. 10% dari data berjumlah 23 data, tetapi ditambahkan menjadi 25 data untuk mewakili semua kategori yang akan digunakan dalam pengujian data baru dengan algoritma Naive Bayes, sedangkan 90% sebanyak 198 data sebagai data latih. Data uji yang terdapat pada tabel 5 sebagai berikut:

Tabel 1 Data Uji

No	Jenis Arsip	Jenis Indeks	Waktu Retensi Aktif	Waktu Retensi Inaktif	Status Retensi Arsip
1	KEUANGAN	Laporan Keuangan	2	8	Permanen
2	KEPEGAWAIAN	Permohonan	1	2	Musnah
3	KERJA SAMA	Tindak Lanjut dan Evaluasi	2	3	Musnah
4	KERJA SAMA	Tindak Lanjut dan Evaluasi	2	3	Permanen
5	KERJA SAMA	Tindak Lanjut dan Evaluasi	2	3	Permanen
6	KETATAUSAHAAN	Usul	2	3	Musnah
7	KETATAUSAHAAN	Pemindahan/ Penyusutan Arsip dan Penataan	2	2	Musnah
8	KEPEGAWAIAN	Surat Pernyataan	1	9	Musnah
9	KETATAUSAHAAN	Pemindahan/ Penyusutan Arsip dan Penataan	2	2	Musnah
10	KETATAUSAHAAN	Usul	2	3	Musnah
11	KEPEGAWAIAN	Pendataan	1	4	Musnah
12	KEPEGAWAIAN	Usul	2	3	Musnah

No	Jenis Arsip	Jenis Indeks	Waktu Retensi Aktif	Waktu Retensi Inaktif	Status Retensi Arsip
13	KEUANGAN	Usul	2	3	Musnah
14	PERENCANAAN DAN PROGRAM	Tindak Lanjut dan Evaluasi	2	3	Permanen
15	KETATAUSAHAAN	Pendataan	2	3	Musnah
16	PERENCANAAN DAN PROGRAM	Renstra	2	3	Permanen
17	KEPEGAWAIAN	Usul	1	2	Musnah
18	KEUANGAN	Laporan	2	5	Musnah
19	KETATAUSAHAAN	Pemusnahan Arsip	2	3	Musnah
20	PERENCANAAN DAN PROGRAM	Program Kerja	2	3	Musnah
21	KETATAUSAHAAN	Surat	2	3	Musnah
22	KETATAUSAHAAN	Usul	2	3	Musnah
23	KERJA SAMA	Kegiatan FGD	2	3	Permanen
24	KERJA SAMA	Rekomendasi	2	4	Musnah
25	KERJA SAMA	Informasi	2	3	Permanen

Adapun langkah-langkah dari Algoritma Naïve Bayes sebagai berikut;

(1) menghitung jumlah kelas / label "P(H)";

Tabel 2 Jumlah Kemunculan Setiap Kelas

P(Permanen)	P(Musnah)
7/25	18/25
0,28	0,72

(2) menghitung jumlah kelas untuk setiap atribut;

a) Tahap 1: Tabel frekuensi jenis arsip (X1)

Tabel 7 frekuensi jenis arsip (X1)

Jenis Arsip	Class		Total
	Permanen	Musnah	
Keuangan	1	3	4
Kepegawaian	0	5	5
Kerja Sama	4	2	6
Ketatausahaan	0	7	7
Perencanaan dan Program	2	1	3
Total	7	18	25

b) Tahap 2: Tabel frekuensi jenis indeks (X2)

Tabel 8 frekuensi jenis indeks (X2)

Jenis Indeks	Class		Total
	Permanen	Musnah	
Laporan Keuangan	1	0	1
Permohonan	0	1	1
Tindak Lanjut dan Evaluasi	3	1	4
Usul	0	6	6
Pemindahan/ Penyusutan Arsip dan Penataan	0	2	2
Surat Pernyataan	0	1	1

Jenis Indeks	Class		Total
	Permanen	Musnah	
Pendataan	0	2	2
Renstra	1	0	1
Laporan	0	1	1
Pemusnahan Arsip	0	1	1
Program Kerja	0	1	1
Surat	0	1	1
Kegiatan FGD	1	0	1
Rekomendasi	0	1	1
Informasi	1	0	1
Total	7	18	25

c) Tahap 3: Tabel frekuensi waktu retensi aktif (X3)

Tabel 9 frekuensi waktu retensi aktif (X3)

Waktu Retensi Aktif	Class		Total
	Permanen	Musnah	
1	0	3	3
2	7	15	22
Total	7	18	25

d) Tahap 4: Tabel frekuensi waktu retensi in-aktif (X4)

Tabel 10 frekuensi jenis arsip (X1)

Waktu Retensi In-Aktif	Class		Total
	Permanen	Musnah	
2	0	4	4
3	6	10	16
4	0	2	2
5	0	1	1
8	1	0	1
9	0	1	1
Total	7	18	25

(3) Mencari predictor prior probability =P(X);

a) Tahap 1: predictor prior probability jenis arsip (X1)

Tabel 11 predictor prior probability jenis arsip (X1)

P(Kuangan)	4/25	0,16
P(Kepegawaian)	5/25	0,2
P(Kerja Sama)	6/25	0,24
P(Ketatausahaan)	7/25	0,28
P(Perencanaan dan Program)	3/25	0,12

b) Tahap 2: predictor prior probability jenis index (X2)

Tabel 12 predictor prior probability jenis index (X2)

P(Laporan Keuangan)	1/25	0,04
P(Permohonan)	1/25	0,04
P(Tindak Lanjut dan Evaluasi)	4/25	0,16
P(Usul)	6/25	0,24
P(Pemindahan/Penyusutan Arsip dan Penataan)	2/25	0,08
P(Surat Pernyataan)	1/25	0,04
P(Pendataan)	2/25	0,08
P(Renstra)	1/25	0,04
P(Laporan)	1/25	0,04

P(Pemusnahan Arsip)	1/25	0,04
P(Program Kerja)	1/25	0,04
P(Surat)	1/25	0,04
P(Kegiatan FGD)	1/25	0,04
P(Rekomendasi)	1/25	0,04
P(Informasi)	1/25	0,04

c) Tahap 3: predictor prior probability waktu retensi aktif (X3)

Tabel 13 predictor prior probability waktu retensi aktif (X3)

P(1)	3/25	0,12
P(2)	22/25	0,88

d) Tahap 4: predictor prior probability waktu retensi in-aktif (X4)

Tabel 14 predictor prior probability waktu retensi in-aktif (X4)

P(2)	4/25	0,16
P(3)	16/25	0,64
P(4)	2/25	0,08
P(5)	1/25	0,04
P(8)	1/25	0,04
P(9)	1/25	0,04

(4) mencari nilai probabilitas pada kelas permanen dan musnah;

a) Data ke-1

$$\frac{P(\text{Keuangan, Laporan Keuangan, 2,8}) \times P(\text{Permanen})}{P(\text{Keuangan, Laporan Keuangan, 2,8})} = \frac{\left(\left(\frac{1}{7}\right) \times \left(\frac{1}{7}\right) \times \left(\frac{2}{7}\right) \times \left(\frac{1}{7}\right)\right) \times 0,28}{(0,16 \times 0,04 \times 0,88 \times 0,04)} = 1,04$$

$$\frac{P(\text{Keuangan, Laporan Keuangan, 2,8}) \times P(\text{Musnah})}{P(\text{Keuangan, Laporan Keuangan, 2,8})} = \frac{\left(\left(\frac{3}{18}\right) \times \left(\frac{0}{18}\right) \times \left(\frac{15}{18}\right) \times \left(\frac{0}{18}\right)\right) \times 0,72}{(0,16 \times 0,04 \times 0,88 \times 0,04)} = 0,00$$

Keputusan prediksi : Pada nilai probabilitas dari kelas “Permanen” memiliki nilai paling besar yaitu 1,04, maka Keputusan akhir dalam penentuan status retensi arsip adalah “Permanen”.

b) Data ke-2

$$\frac{P(\text{Kepegawaian, Permohonan, 1,2}) \times P(\text{Permanen})}{P(\text{Kepegawaian, Permohonan, 1,2})} = \frac{\left(\left(\frac{0}{7}\right) \times \left(\frac{0}{7}\right) \times \left(\frac{0}{7}\right) \times \left(\frac{0}{7}\right)\right) \times 0,28}{(0,2 \times 0,04 \times 0,12 \times 0,16)} = 0,00$$

$$\frac{P(\text{Kepegawaian, Permohonan, 1,2}) \times P(\text{Musnah})}{P(\text{Kepegawaian, Permohonan, 1,2})} = \frac{\left(\left(\frac{5}{18}\right) \times \left(\frac{1}{18}\right) \times \left(\frac{3}{18}\right) \times \left(\frac{4}{18}\right)\right) \times 0,72}{(0,2 \times 0,04 \times 0,12 \times 0,16)} = 2,68$$

Keputusan prediksi: Pada nilai probabilitas dari kelas “Musnah” memiliki nilai paling besar yaitu 2,68, maka Keputusan akhir dalam penentuan status retensi arsip adalah “Musnah”.

c) Data ke-3

$$\frac{P(\text{Kepegawaian, Permohonan, 1,2}) \times P(\text{Permanen})}{P(\text{Kepegawaian, Permohonan, 1,2})} = \frac{\left(\left(\frac{0}{7}\right) \times \left(\frac{0}{7}\right) \times \left(\frac{0}{7}\right) \times \left(\frac{0}{7}\right)\right) \times 0,28}{(0,2 \times 0,04 \times 0,12 \times 0,16)} = 0,00$$

$$\frac{P(\text{Kepegawaian, Permohonan, 1,2}) \times P(\text{Musnah})}{P(\text{Kepegawaian, Permohonan, 1,2})} = \frac{\left(\left(\frac{5}{18}\right) \times \left(\frac{1}{18}\right) \times \left(\frac{3}{18}\right) \times \left(\frac{4}{18}\right)\right) \times 0,72}{(0,2 \times 0,04 \times 0,12 \times 0,16)} = 2,68$$

Keputusan prediksi: Pada nilai probabilitas dari kelas “Musnah” memiliki nilai paling besar yaitu 2,68, maka Keputusan akhir dalam penentuan status retensi arsip adalah “Musnah”.

d) Data ke-4

$$\frac{P(\text{Kepengawaaian, Permohonan, 1,2}) \times P(\text{Permanen})}{P(\text{Kepengawaaian, Permohonan, 1,2})} = \frac{\left(\left(\frac{0}{7}\right) \times \left(\frac{0}{7}\right) \times \left(\frac{0}{7}\right) \times \left(\frac{0}{7}\right)\right) \times 0,28}{(0,2 \times 0,04 \times 0,12 \times 0,16)} = 0,00$$

$$\frac{P(\text{Kepengawaaian, Permohonan, 1,2}) \times P(\text{Musnah})}{P(\text{Keuangan, Permohonan, 1,2})} = \frac{\left(\left(\frac{5}{18}\right) \times \left(\frac{1}{18}\right) \times \left(\frac{3}{18}\right) \times \left(\frac{4}{18}\right)\right) \times 0,72}{(0,2 \times 0,04 \times 0,12 \times 0,16)} = 2,68$$

Keputusan prediksi: Pada nilai probabilitas dari kelas “Musnah” memiliki nilai paling besar yaitu 2,68, maka Keputusan akhir dalam penentuan status retensi arsip adalah “Musnah”.

e) Data ke-5

$$\frac{P(\text{Kerja Sama, Tindak Lanjut dan Evaluasi, 2,3}) \times P(\text{Permanen})}{P(\text{Kerja Sama, Tindak Lanjut dan Evaluasi, 2,3})} = \frac{\left(\left(\frac{5}{9}\right) \times \left(\frac{4}{9}\right) \times \left(\frac{9}{9}\right) \times \left(\frac{9}{9}\right)\right) \times 0,36}{(0,24 \times 0,2 \times 0,84 \times 0,64)} = 3,06$$

$$\frac{P(\text{Kerja Sama, Tindak Lanjut dan Evaluasi, 2,3}) \times P(\text{Musnah})}{P(\text{Kerja Sama, Tindak Lanjut dan Evaluasi, 2,3})} = \frac{\left(\left(\frac{1}{16}\right) \times \left(\frac{0}{16}\right) \times \left(\frac{12}{16}\right) \times \left(\frac{8}{16}\right)\right) \times 0,64}{(0,24 \times 0,2 \times 0,84 \times 0,64)} = 0,00$$

Keputusan prediksi: Pada nilai probabilitas dari kelas “Permanen” memiliki nilai paling besar yaitu 3,06, maka Keputusan akhir dalam penentuan status retensi arsip adalah “Permanen”.

Berikut perbandingan status retensi arsip dari data asli dengan data setelah perhitungan pada tabel 15:

Tabel 15 Perbandingan Data Asli dengan Data setelah Perhitungan

No	Jenis Arsip	Jenis Indeks	Status Retensi Arsip	Prediksi Status Retensi Arsip
1	KEUANGAN	Laporan Keuangan	Permanen	Permanen
2	KEPEGAWAIAN	Permohonan	Musnah	Musnah
3	KERJA SAMA	Tindak Lanjut dan Evaluasi	Musnah	Permanen
4	KERJA SAMA	Tindak Lanjut dan Evaluasi	Permanen	Permanen
5	KERJA SAMA	Tindak Lanjut dan Evaluasi	Permanen	Permanen
6	KETATAUSAHAAN	Usul	Musnah	Musnah
7	KETATAUSAHAAN	Pemindahan/Penyusutan Arsip dan Penataan	Musnah	Musnah
8	KEPEGAWAIAN	Surat Pernyataan	Musnah	Musnah
9	KETATAUSAHAAN	Pemindahan/Penyusutan Arsip dan Penataan	Musnah	Musnah
10	KETATAUSAHAAN	Usul	Musnah	Musnah
11	KEPEGAWAIAN	Pendataan	Musnah	Musnah
12	KEPEGAWAIAN	Usul	Musnah	Musnah
13	KEUANGAN	Usul	Musnah	Musnah
14	PERENCANAAN DAN PROGRAM	Tindak Lanjut dan Evaluasi	Permanen	Permanen
15	KETATAUSAHAAN	Pendataan	Musnah	Musnah

No	Jenis Arsip	Jenis Indeks	Status Retensi Arsip	Prediksi Status Retensi Arsip
16	PERENCANAAN DAN PROGRAM	Renstra	Permanen	Permanen
17	KEPEGAWAIAN	Usul	Musnah	Musnah
18	KEUANGAN	Laporan	Musnah	Musnah
19	KETATAUSAHAAN	Pemusnahan Arsip	Musnah	Musnah
20	PERENCANAAN DAN PROGRAM	Program Kerja	Musnah	Musnah
21	KETATAUSAHAAN	Surat	Musnah	Musnah
22	KETATAUSAHAAN	Usul	Musnah	Musnah
23	KERJA SAMA	Kegiatan FGD	Permanen	Permanen
24	KERJA SAMA	Rekomendasi	Musnah	Musnah
25	KERJA SAMA	Informasi	Permanen	Permanen

c. Tahap Pengujian (*Evaluation*)

Pada tahap ini mengukur keakuratan hasil yang dicapai dengan menggunakan confusion matrix. Kemudian, berdasarkan variabel yang telah ditetapkan, Pada langkah ini pengukuran keakuratan dilakukan antara hasil yang dicapai menggunakan confusion matrix sebagai tahapan evaluasi. Ini hasil confusion matriks yang dilakukan dengan membandingkan data nyata dengan data prediksi yang ditunjukkan pada tabel 16 di bawah ini:

Tabel 16 Confussion Matrix

Klasifikasi	Prediksi	
	Permanen	Musnah
Permanen	7	0
Musnah	1	17

Berdasarkan tabel 16, terlihat bahwa data nyata dengan keputusan penentuan status retensi arsip yang "Permanen" setelah dilakukan prediksi dan keputusan penentuan status retensi arsip yang "Permanen" tetap "Permanen" sebanyak 7 data, kemudian data nyata dengan keputusan penentuan status retensi arsip "Permanen" setelah dilakukan prediksi dan keputusan penentuan status retensi arsip nya berubah menjadi "Musnah" tidak ada atau 0 data, Untuk data nyata dengan keputusan penentuan status retensi arsip "Musnah" setelah dilakukan prediksi dan keputusan penentuan status retensi arsip "Permanen" sebanyak 1 data, kemudian data nyata dengan keputusan penentuan status retensi arsip "Musnah" setelah dilakukan prediksi dan keputusan penentuan status retensi arsip nya tetap "Musnah" sebanyak 17 data. Selanjutnya, perhitungan akurasi dilakukan dengan cara berikut:

$$Akurasi = \frac{7 + 17}{7 + 17 + 0 + 1} \times 100\% = 96\%$$

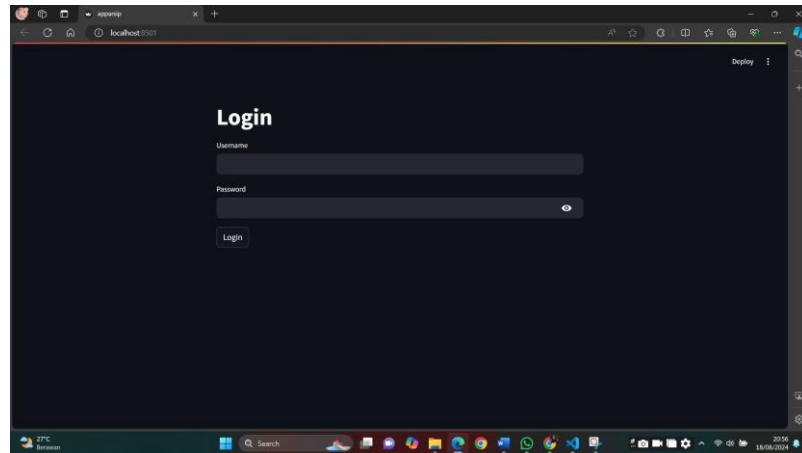
Berdasarkan perhitungan akurasi, maka hasil akurasi yang didapatkan sebesar 96% dan dapat dinyatakan pemodelan dengan naïve bayes untuk penentuan status retensi arsip termasuk ke dalam kategori sangat baik.

d. Tahap Penyebaran (*Deployment*)

Pada tahap penyebaran dibuat prototype aplikasi untuk memvisualisasikan hasil modeling dengan Naïve Bayes pada penentuan status retensi arsip yang terdiri dari beberapa tampilan sebagai berikut:

1. Halaman login

Halaman ini merupakan halaman awal operator untuk dapat mengakses laman utama



Gambar 2. Halaman Login

2. Halaman utama

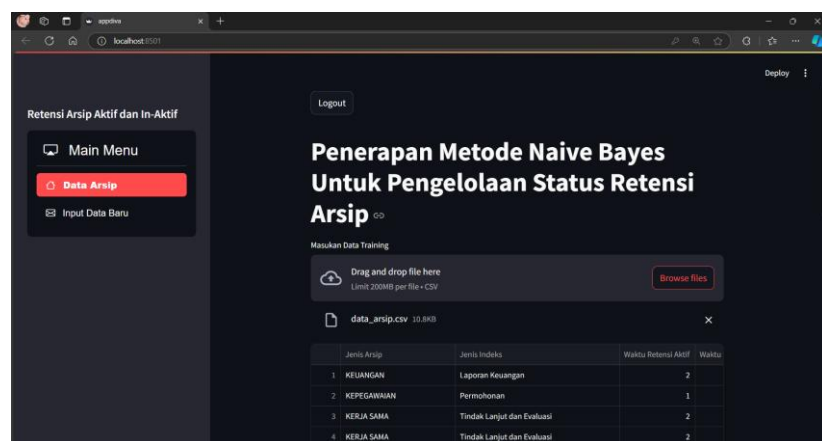
Halaman ini merupakan menu utama operator kearsipan yang menampilkan beberapa fitur diantaranya halaman data arsip dan halaman input data baru



Gambar 3. Halaman Menu Utama

3. Halaman data arsip

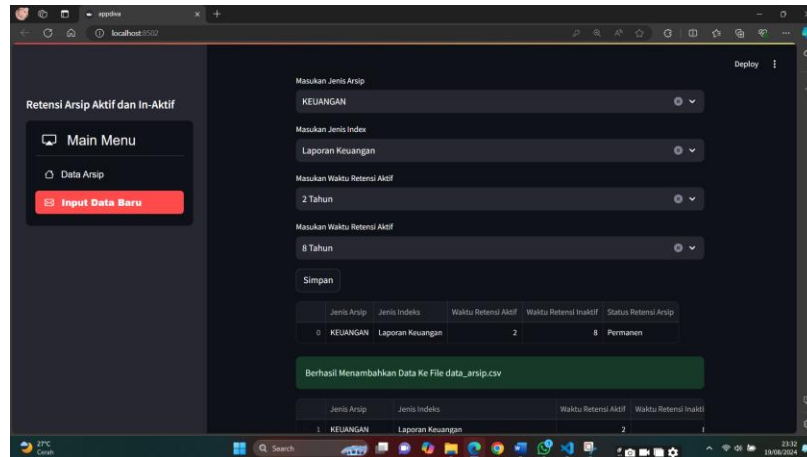
Halaman ini menampilkan data arsip yang sudah tersimpan dalam database kearsipan



Gambar 4. Tampilan data arsip

4. Halaman penentuan status retensi arsip

Halaman ini menampilkan tahap input data arsip baru dan sudah dapat ditentukan status retensi arsipnya.



Gambar 5. Tampilan penentuan status retensi arsip

D. KESIMPULAN

Berdasarkan Hasil Penelitian yang dilakukan, Kesimpulan yang dapat diuraikan antara lain:

1. Penerapan metode Naïve Bayes dapat memperoleh status retensi arsip dengan uji akurasi menggunakan confission matrix dan Metode Naive Bayes menjadi lebih efektif dalam menentukan status retensi arsip.
2. Hasil uji menunjukkan akurasi perhitungan sebesar 96% menyatakan bahwa prediksi model naïve bayes sangat baik
3. Saran untuk pengembangan penelitian selanjutnya, prototype dapat dikembangkan dalam bentuk sistem informasi yang lebih lengkap dengan dikombinasikan dengan metode penelitian lain

E. DAFTAR PUSTAKA

- [1] U. Gellysa, *Penerapan Data Mining DI BERBAGAI BIDANG*, 1st ed. Bali: SONPEDIA, 2023.
- [2] Amna, *Data mining*, 1st ed., vol. 2, no. January 2013. Padang: PT GLOBAL EKSEKUTIF TEKNOLOGI, 2023.
- [3] Y. S. Hidayat, A. G. Herdipriyansah, I. A. Agustin, J. Jaenudin, and D. Desrizal, "Pemodelan Data Mining Menggunakan Neural Network Untuk Seleksi Mahasiswa Penerima Beasiswa Bantuan," *TeknoIS: Jurnal Ilmiah Teknologi Informasi dan Sains*, vol. 14, no. 1, pp. 9–20, Jan. 2024, doi: 10.36350/jbs.v14i1.222.
- [4] J. W. Iskandar and Y. Nataliani, "Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 6, pp. 1120–1126, 2021, doi: 10.29207/resti.v5i6.3588.
- [5] Alfandi Safira and F. N. Hasan, "Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier," *ZONAsi: Jurnal Sistem Informasi*, vol. 5, no. 1, pp. 59–70, 2023, doi: 10.31849/zn.v5i1.12856.
- [6] F. Gorunescu, *Data Mining*, vol. 12. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. doi: 10.1007/978-3-642-19721-5.

- [7] I. Kasim, Yuyun, and S. Sahibu, “KLASIFIKASI SURAT DIGITAL MENGGUNAKAN ALGORITMA NAÏVE BAYES,” *JURNAL IT Media Informasi IT STMIK Handayani*, vol. 13, no. 2, pp. 57–62, Aug. 2020.
- [8] A. Triawan and D. Lintang Melinda, “Penerapan Metode Naïve Bayes Untuk Rekomendasi Topik Tugas Akhir Berdasarkan Daftar Hasil Studi Mahasiswa di Perguruan Tinggi,” *Teknois : Jurnal Ilmiah Teknologi Informasi dan Sains*, vol. 10, no. 2, pp. 58–70, Nov. 2020, doi: 10.36350/jbs.v10i2.91.
- [9] F. Septianingrum and A. S. Y. Irawan, “Metode Seleksi Fitur Untuk Klasifikasi Sentimen Menggunakan Algoritma Naive Bayes: Sebuah Literature Review,” *Jurnal Media Informatika Budidarma*, vol. 5, no. 3, p. 799, 2021, doi: 10.30865/mib.v5i3.2983.
- [10] S. Arikunto, *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: PT. Bina Aksara, 2010.