

# PENERAPAN ALGORITMA *NAÏVE BAYES* UNTUK KLASIFIKASI RETENSI ARSIP

Budi Harijanto<sup>1</sup>, Yuri Ariyanto<sup>2</sup>, Luthfia Miftahurroifa<sup>3</sup>

<sup>1,2</sup>Jurusan Teknologi Informasi, Program Studi Teknik Informatika, <sup>3</sup>Politeknik Negeri Malang

<sup>1</sup>budi.hijet@gmail.com, <sup>2</sup>yuri.bjn@gmail.com, <sup>3</sup>miftaluthfia@gmail.com

## Abstrak

Pengelolaan arsip pada Bagian Persuratan dan Pengarsipan Kantor Pusat Politeknik Negeri Malang ada kendala dalam penentuan jangka waktu penyimpanan arsip atau retensi arsip, sehingga berkas arsip yang belum di tentukan retensi arsipnya menumpuk. Dalam penelitian ini dibuat sistem pengarsipan untuk mengklasifikasikan retensi arsip surat. Sistem ini menggunakan data arsip yang berstatus dinilai kembali yang akan diklasifikasikan ke dalam arsip yang dapat dimusnahkan atau arsip permanen. Untuk mengklasifikasikan retensi arsip surat, peneliti menerapkan Algoritma *Naïve Bayes*. Perhitungan dalam penerapan Algoritma *Naïve Bayes* untuk mendapatkan hasil probabilitas sebagai rekomendasi dalam pengklasifikasian retensi arsip memiliki 5 (lima) kriteria yaitu klasifikasi arsip, nilai guna, proses arsip, waktu aktif dan waktu inaktif.

Dalam penelitian ini, dilakukan beberapa uji coba penerapan Algoritma *Naïve Bayes* ke dalam sistem pengarsipan dengan membandingkan hasil *output* sistem dan hasil pakar dari 1008 data *training* dan 100 data *testing*. Hasil uji coba menunjukkan tingkat akurasi mencapai 95%. Banyaknya data dapat mempengaruhi hasil rekomendasi klasifikasi retensi arsip.

**Kata kunci:** retensi arsip, algoritma, *Naïve Bayes*, klasifikasi

## 1. Pendahuluan

Perkembangan teknologi informasi telah berpengaruh dalam pengelolaan data dan penyampaian informasi bagi setiap instansi sehingga mempermudah data maupun informasi yang diolah menggunakan cara manual menjadi berbasis komputerisasi. Dengan adanya teknologi informasi yang digunakan dalam dunia persuratan, setiap saat pencatatan sebuah surat didokumentasikan sebagai barang bukti dan disimpan menjadi arsip.

Dalam pengelolaan tentang kearsipan pada Bagian Persuratan dan Pengarsipan Kantor Pusat di Politeknik Negeri Malang, terdapat pengelolaan retensi arsip, retensi arsip berfungsi agar tidak terjadi penumpukan berkas arsip yang sudah tidak terpakai atau tingkat kepentingan yang telah lampau. Retensi arsip merupakan jangka waktu penyimpanan arsip, retensi arsip ditentukan atas dasar nilai kegunaan setiap berkas arsip. Retensi arsip terbagi menjadi tiga macam yaitu arsip musnah, permanen dan dinilai kembali. Dalam pengelolaan retensi arsip masih memiliki kendala yaitu kurang optimal dalam pengklasifikasiannya.

Dalam penelitian ini akan dibuat sistem dengan menerapkan Algoritma *Naïve Bayes Classifier* untuk klasifikasi retensi arsip. Data yang digunakan untuk diklasifikasikan yaitu data arsip berjenis arsip dinilai kembali. Arsip dinilai kembali merupakan arsip yang belum ditentukan rekomendasinya yaitu arsip musnah atau

permanen sehingga diperlukan pengkajian ulang, oleh karena itu perlu adanya metode untuk menentukan klasifikasi arsip dinilai kembali yang akhirnya menjadi arsip musnah atau permanen.

Dalam menerapkan perhitungan Algoritma *Naïve Bayes Classifier* untuk probabilitas tiap kelas dalam pengklasifikasian retensi arsip memiliki lima kriteria yaitu klasifikasi arsip, nilai guna, proses arsip, waktu aktif dan waktu inaktif. Dengan menggunakan Algoritma *Naïve Bayes Classifier* untuk klasifikasi retensi arsip, dapat memprediksi probabilitas sebuah class, sehingga dapat menghasilkan rekomendasi berdasarkan data pembelajaran atau data training.

## 2. Kajian Pustaka

### 2.1 Retensi Arsip

Sebagai suatu endapan informasi, maka pertumbuhan akumulasi arsip akan sejalan dengan perkembangan pelaksanaan fungsi-fungsi unit kerja dalam struktur organisasi instansi. Artinya arsip akan berkembang terus dalam jumlah dan jenisnya sehingga diperlukan fasilitas penyimpanan yang terus berkembang. Sementara itu sebagai suatu proses yang bersifat fungsional, nilai guna arsip akan berubah dan menyusut sejalan dengan tahapan proses tersebut. Dengan demikian esensi pemahaman terhadap nilai guna arsip adalah kemampuan untuk melakukan penyusutan secara sistematis agar hanya arsip

yang yang bernilai guna sajarah yang harus disimpan.

Berikut ini merupakan macam retensi arsip secara umum :

- a. Arsip permanen adalah arsip yang memiliki nilai guna sekunder atau permanen, penyimpanannya di serahkan ke Arsip Nasional Republik Indonesia (ANRI) sebagai bukti pertanggungjawaban.
- b. Arsip musnah adalah jenis arsip yang dapat dimusnahkan karena jangka waktu penyimpanan telah habis dan tidak memiliki nilai guna lagi.
- c. Arsip dinilai kembali adalah arsip yang belum dapat ditentukan rekomendasi akhirnya apakah dimusnahkan atau dipermanenkan, sehingga perlu penilaian dan pengkajian kembali. Asichin, (2012).

**2.2 Algoritma Naïve Bayes Classifier**

Bayes merupakan teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan teorema bayes dengan asumsi independensi (ketidaktergantungan) yang kuat (naif). Dengan kata lain, dalam Naïve Bayes, model yang digunakan adalah “model fitur independen”.

Prediksi Bayes didasarkan pada Teorema Bayes dengan formula umum sebagai berikut :

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)} \tag{1}$$

Keterangan :

**P(H|E)** : Probabilitas akhir bersyarat (*conditional probability*) suatu hipotesis H terjadi jika diberikan bukti (*evidence*) E terjadi.

**P(E|H)** : Probabilitas sebuah bukti E terjadi akan mempengaruhi hipotesis H.

**P(H)** : Probabilitas awal (*priori*) hipotesis H terjadi tanpa memandang bukti apapun.

**P(E)** : Probabilitas awal (*priori*) bukti E terjadi tanpa memandang hipotesis/bukti yang lain.

Ide dasar dari aturan Bayes adalah bahwa hasil dari hipotesis atau peristiwa (H) dapat diperkirakan berdasarkan pada beberapa bukti (E) yang diamati. Ada beberapa hal penting dari aturan Bayes tersebut, yaitu :

- a. Sebuah probabilitas awal/priori H atau P(H) adalah probabilitas dari suatu hipotesis sebelum bukti diamati.
- b. Sebuah probabilitas akhir H atau (P(H|E)) adalah probabilitas dari suatu hipotesis setelah bukti diamati.

Kaitan antara Naïve Bayes dengan klasifikasi, kolerasi hipotesis, dan bukti dengan klasifikasi adalah bahwa hipotesis dalam Teorema Bayes merupakan label kelas yang menjadi target

pemetaan dalam klasifikasi, sedangkan bukti merupakan fitur – fitur yang menjadi masukan dalam model klasifikasi. Jika X adalah vektor masukan yang berisi fitur dan Y adalah label kelas, Naïve Bayes dituliskan dengan P(Y|X). Notasi tersebut berarti probabilitas label kelas Y didapatkan setelah fitur – fitur X diamati. Notasi ini disebut juga probabilitas akhir (*posterior probability*) untuk Y, sedangkan P(Y) disebut disebut probabilitas awal (*prior probability*) Y.

Selama proses pelatihan harus dilakukan pembelajaran probabilitas akhir P(Y|X) pada model untuk setiap kombinasi X dan Y berdasarkan informasi yang didapat dari data latih. Dengan membangun model tersebut, suatu data uji X’ dapat diklasifikasikan dengan mencari nilai Y’ dengan memaksimalkan nilai (P(Y’|X’)) yang didapat.

Formulasi Naïve Bayes untuk klasifikasi adalah

$$P(Y | X) = \frac{P(Y) \prod_{i=1}^q P(X_i | Y)}{P(X)} \tag{2}$$

P(Y|X) adalah probabilitas data dengan vector X pada kelas Y. P(Y) adalah probabilitas awal kelas Y.  $\prod_{i=1}^q P(X_i | Y)$  adalah probabilitas independen kelas Y dari semua fitur dalam vektor X. Nilai p(X) selalu tetap sehingga dalam perhitungan prediksi nantinya kita tinggal menghitung bagian  $P(Y) \prod_{i=1}^q P(X_i | Y)$  dengan memilih yang terbesar sebagai kelas yang dipilih sebagai hasil prediksi. Sementara probabilitas independen  $\prod_{i=1}^q P(X_i | Y)$  tersebut merupakan pengaruh semua fitur dari data terhadap setiap kelas Y. yang dinotasikan dengan :

$$P(X|Y = y) = \prod_{i=1}^q P(X_i | Y = y) \tag{3}$$

Setiap set fitur X =[X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>, ..., X<sub>q</sub>] terdiri atas q atribut (q dimensi). Prasetyo, (2012)

Dalam banyak bidang ilmu informasi, untuk menemukan hasil prediksi dari suatu data adalah hal yang penting. Penemuan awal dari hasil prediksi biasanya dilakukan dengan *Training set* sementara *Test set* dan *Validation Set* digunakan untuk mengevaluasi apakah hasil ditemukan. *Training set* adalah suatu set data yang digunakan untuk menemukan data yang berpotensi prediktif. *Test set* adalah suatu set data yang digunakan untuk menilai kekuatan dan utilitas dari hubungan prediktif. Berikut ini merupakan tahapan dalam perhitungan Algoritma Naïve Bayes Classifier :

- a. *Learning* (Pembelajaran)
  - 1) Bentuk klasifikasi pada setiap *data training*.
  - 2) Hitung probabilitas pada setiap kategori P(i).
  - 3) Tentukan frekuensi setiap kata pada setiap kategori P(x|i).

b. *Classify* (Pengklasifikasian)

- 1) Hitung  $P(x|i) = \prod P(x|E)$  untuk setiap kategori.
- 2) Tentukan kategori dengan nilai  $P(x|i) = \prod P(x|E)$  maksimal.

3. **Metodologi Penelitian**

Dalam penelitian ini akan dibuat sistem dengan menerapkan Algoritma *Naïve Bayes Classifier* untuk klasifikasi retensi arsip. Data yang digunakan untuk diklasifikasikan yaitu data arsip berjenis arsip dinilai kembali. Arsip dinilai kembali merupakan arsip yang belum ditentukan rekomendasinya yaitu arsip musnah atau permanen sehingga diperlukan pengkajian ulang, oleh karena itu perlu adanya metode untuk menentukan klasifikasi arsip dinilai kembali yang akhirnya menjadi arsip musnah atau permanen. Data-data yang dibutuhkan dalam penelitian ini antara lain :

Tabel 1. Data Yang Dibutuhkan

Data	Keterangan
Klasifikasi arsip	Klasifikasi arsip surat
Nilai guna	Nilai guna arsip surat
Proses arsip	Proses untuk kajian arsip
Waktu aktif	Waktu aktif arsip surat
Waktu inaktif	Waktu inaktif arsip surat

Pada aplikasi ini dibutuhkan sebuah basis data yang digunakan untuk menampung data, baik bahan penilaian maupun hasil penilaian. Basis data tersebut berisi klasifikasi arsip, nilai guna, proses arsip, waktu aktif dan waktu inaktif pada arsip bagian Persuratan Dan Pengarsipan Kantor Pusat Politeknik Negeri Malang. *Output* aplikasi ini adalah hasil klasifikasi retensi arsip untuk menentukan jadwal retensi arsip.

Metode pengumpulan data dalam pengerjaan penelitian ini melalui dua metode, yaitu sebagai berikut :

- a. Observasi yaitu pengamatan secara langsung prosedur dan alur yang berjalan saat ini. Serta pengamatan permasalahan yang berhubungan dengan pelayanan dan pengolahan data dalam pengarsipan pada Persuratan dan Pengarsipan Kantor Pusat dan UPT Kearsipan Politeknik Negeri Malang.
- b. Wawancara (*interview*) merupakan suatu data yang dilakukan dengan cara tanya jawab secara langsung dengan pihak – pihak Persuratan dan Pengarsipan yang terkait dengan pengelolaan arsip. Dengan adanya wawancara maka didapatkan data yang dapat dijadikan kriteria – kriteria yang digunakan dalam perhitungan Algoritma *Naïve Bayes Classifier*.
- c. Studi literatur merupakan pengumpulan data dengan cara mempelajari dan memahami hal – hal yang berkaitan dengan pengerjaan sistem dan sumber-sumber studi literatur yang digunakan sebagai studi pustaka dalam

laporan. Hal – hal yang diperlukan dalam pengerjaan sistem meliputi retensi arsip dan klasifikasi menggunakan Algoritma *Naïve Bayes Classifier*.

Tahap pengumpulan data dilakukan dengan cara wawancara yang dilakukan wawancara pada tanggal 13 dan 15 Maret 2017 di UPT Kearsipan Politeknik Negeri Malang mengenai layanan – layanan yang akan dibuat. Layanan – layanan yang akan dibuat dalam aplikasi yang memiliki layanan registrasi pegawai, pengarsipan surat, pengolahan jadwal retensi arsip atau klasifikasi retensi arsip dan peminjaman arsip. Untuk dapat dibuat layanan – layanan tersebut maka dibutuhkan data – data untuk mendukung pembuatannya yaitu data arsip untuk pengolahan jadwal retensi arsip atau klasifikasi retensi arsip.

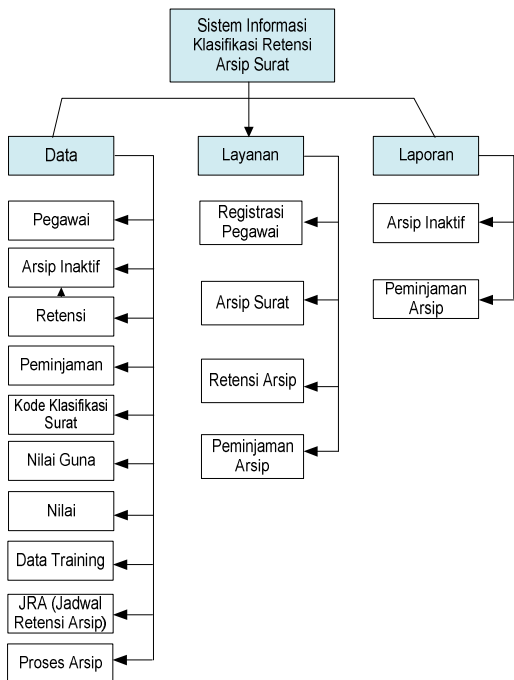
Layanan registrasi pegawai merupakan layanan yang dapat dijalankan oleh admin. Apabila ada pegawai yang belum terdaftar, pegawai tersebut harus memberikan data diri ke admin lalu admin akan menambahkan data pegawai tersebut ke sistem. Data yang digunakan dalam layanan registrasi pegawai yaitu data pegawai.

Layanan arsip inaktif adalah layanan penyimpanan data arsip. Dari layanan pengarsipan ada laporan yang akan dibuat dari layanan pengarsipan surat yaitu laporan arsip.

Pengelolaan retensi arsip merupakan layanan yang digunakan untuk menentukan jadwal retensi arsip yang terdapat pada layanan arsip inaktif. Setelah mengisi data arsip inaktif, maka data akan terklasifikasi kedalam jenis retensi arsip musnah atau permanen. Pada layanan ini akan diterapkan Algoritma *Naïve Bayes Classifier* untuk mengklasifikasikan arsip yang berjenis dinilai kembali kedalam arsip musnah atau permanen sebagai rekomendasi dari hasil perhitungan. Data – data yang dibutuhkan sebagai pendukung layanan retensi arsip yaitu data arsip inaktif, data *training* yang digunakan sebagai acuan untuk klasifikasi retensi arsip, data klasifikasi arsip, data nilai guna dan data proses arsip.

Peminjaman arsip merupakan layanan yang digunakan untuk menyimpan data peminjaman arsip. Data – data yang dibutuhkan sebagai pendukung layanan peminjaman arsip yaitu data arsip. Peminjaman arsip dilengkapi pencarian data – data arsip yang diperlukan.

Penjelasan di atas dapat digambarkan dalam *Work Breakdown Structure* WBS sebagai berikut:



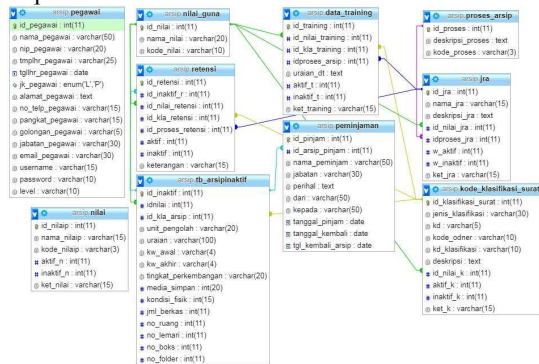
Gambar 1. Work Breakdown Structure (WBS)

4. Hasil dan Pembahasan

Bab ini membahas tentang implementasi sistem pada aplikasi yang menerapkan Algoritma *Naïve Bayes* untuk klasifikasi retensi arsip surat yang didasarkan pada hasil analisis dan perancangan yang telah dibuat. Pembahasan implementasi ini terdiri implementasi *database*, sistem dan metode.

4.1 Implementasi Database

Berikut ini merupakan *database* pada implementasi sistem :

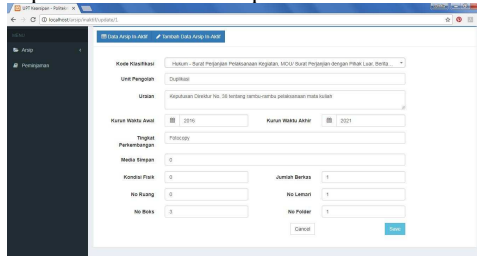


Gambar 2. Implementasi Database

4.2 Implementasi Proses

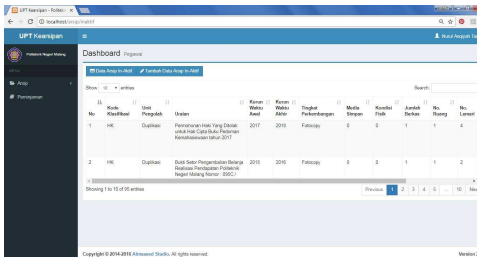
Algoritma *Naïve Bayes Classifier* yang digunakan untuk memberikan hasil klasifikasi retensi arsip. Proses perhitungan menggunakan Algoritma *Naïve Bayes Classifier* dilakukan setelah *user* pegawai mengisi data arsip inaktif lalu mengisi data peninjauan kembali arsip yang

meliputi nilai guna arsip, klasifikasi arsip, proses arsip, waktu aktif dan waktu inaktif. Berikut ini tampilan tambah data arsip inaktif.



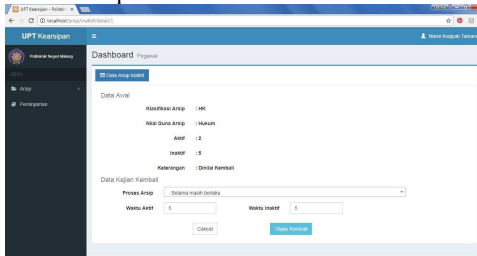
Gambar 3. Tambah Data Arsip Inaktif

Berikut ini tampilan *list view* dari data arsip inaktif.



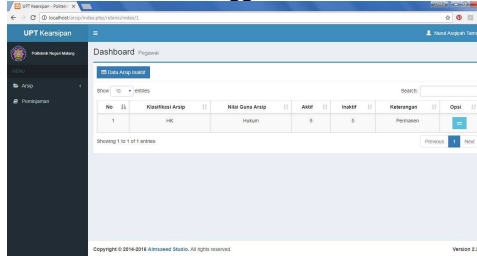
Gambar 4. List View Data Arsip Inaktif

Setelah menambah data arsip inaktif, pada halaman lihat data list arsip terdapat tombol untuk tinjauan kembali arsip. Berikut halaman tinjauan kembali arsip inaktif.



Gambar 5. Tinjauan Kembali Data Arsip Inaktif

Berikut ini merupakan hasil perhitungan dari peninjauan kembali arsip inaktif dengan hasil probabilitas akhir tertinggi.



Gambar 6. Hasil Klasifikasi Data Arsip Inaktif

4.3 Pengujian Hasil Algoritma *Naïve Bayes Classifier*

Pengujian hasil sistem dengan *microsoft excel* akan di uji dengan menghitung nilai probabilitas dan menghitung hasil untuk

mengklasifikasikan retensi arsip. Untuk data training menggunakan 152 data retensi arsip.

Perhitungan probabilitas untuk mengklasifikasikan retensi arsip kedalam tiga kelas yaitu musnah dan permanen yang memiliki 5 fitur meliputi kode klasifikasi arsip, nilai guna, proses arsip, waktu aktif dan waktu inaktif. Untuk menghitung probabilitas akan di uji dengan menggunakan nilai sebagai berikut yang ditunjukkan pada Tabel 2.

Tabel 2. Tabel Data *Testing*

Deskripsi Arsip	Proses Arsip	Klasifikasi Arsip
Daftar Permintaan Gaji dan Tunjangan PNS Dosen Jurusan Administrasi Niaga Politeknik Negeri Malang (991731-2911-30) Bagian Bulan Maret 2015	Setelah tindak lanjut hasil pemeriksaan telah selesai	Keuangan – KUI
Nilai Guna	Waktu	
	Aktif	Inaktif
Keuangan	2	3

Berikut ini merupakan Tabel 3 yaitu nilai probabilitas tiap fitur klasifikasi permanen hasil manual.

Tabel 3. Tabel Nilai Probabilitas Tiap Fitur Klasifikasi Permanen Hasil *Excel*

Fitur Klasifikasi	Excel
Kode klasifikasi arsip	0,5
Nilai guna	0,5
Proses Arsip	0,5
Waktu Aktif	0,333
Waktu Inaktif	0,5

Berikut ini merupakan Gambar 6 yaitu nilai probabilitas tiap fitur klasifikasi permanen hasil sistem.

Permanen	
Nilai Guna	0.5
Klasifikasi Arsip	0.5
Proses Arsip	0.5
Waktu Aktif	0.333
Waktu Inaktif	0.5

Gambar 7. Nilai Probabilitas Tiap Fitur Klasifikasi Permanen Hasil Sistem

Berikut ini merupakan Tabel 4 yaitu nilai probabilitas akhir klasifikasi permanen hasil manual.

Tabel 4. Tabel Nilai Probabilitas Akhir Klasifikasi Permanen Hasil *Excel*

Permanen
0,01041667

Berikut ini merupakan Gambar 7 yaitu nilai probabilitas akhir klasifikasi permanen hasil sistem.

Permanen
0.01040625

Gambar 8. Nilai Probabilitas Akhir Klasifikasi Permanen Hasil Sistem

Tabel 5. Tabel Nilai Probabilitas Tiap Fitur Klasifikasi Musnah Hasil *Excel*

Fitur Klasifikasi	Excel
Kode klasifikasi arsip	0,5
Nilai guna	0,5
Proses Arsip	0,5
Waktu Aktif	0,667
Waktu Inaktif	0,5

Musnah	
Nilai Guna	0.5
Klasifikasi Arsip	0.5
Proses Arsip	0.5
Waktu Aktif	0.667
Waktu Inaktif	0.5

Gambar 9. Nilai Probabilitas Tiap Fitur Klasifikasi Musnah Hasil Manual

Berikut ini merupakan Tabel 6 yaitu nilai probabilitas akhir klasifikasi permanen hasil manual.

Tabel 6. Tabel Nilai Probabilitas Akhir Klasifikasi Musnah Hasil *Excel*

Musnah
0,02051875

Berikut ini merupakan Gambar 10 yaitu nilai probabilitas akhir klasifikasi permanen hasil sistem.

Musnah
0.02084375

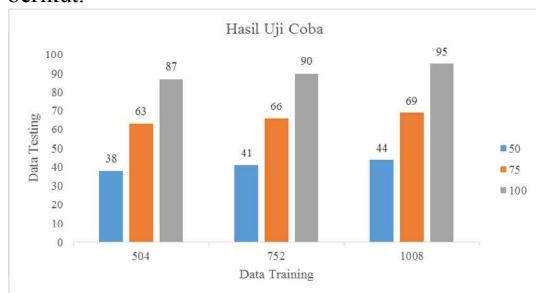
Gambar 10. Nilai Probabilitas Akhir Klasifikasi Musnah Hasil Sistem

Perhitungan hasil probabilitas akhir yaitu hasil perkalian dari nilai setiap fitur retensi yang dikali dengan nilai setiap kelas retensi, selanjutnya hasil probabilitas tertinggi merupakan hasil klasifikasi. Jadi, dapat disimpulkan bahwa data X dari data *testing* memiliki kelas="Musnah" karena  $P(Musnah | X)$  memiliki nilai maksimum pada perhitungan probabilitas akhir.

Pengujian hasil sistem dengan menerapkan Algoritma *Naïve Bayes Classifier* terhadap hasil klasifikasi retensi oleh pakar dengan data *training* dan data *testing* yang berbeda. Dibawah ini menunjukkan rumus perhitungan akurasi:

$$\text{Akurasi (\%)} = \frac{\text{Banyak data benar}}{\text{Jumlah data keseluruhan}} \times 100\% \quad (1)$$

Hasil akurasi dari 50, 75 dan 100 data yang diuji dengan menggunakan data *training* sebanyak 504, 752 dan 1008 ditunjukkan pada grafik berikut.



Gambar 11. Grafik Uji Coba

Dari 1008 data training dan 100 data testing, terdapat 95 data yang sesuai dan 5 data yang tidak sesuai dari hasil perhitungan sistem dengan pakar. Dari hasil analisis tersebut, maka dapat dilakukan perhitungan akurasi sebagai berikut.

$$\text{Akurasi (\%)} = \frac{95}{100} \times 100\% = 95\%$$

Hasil akurasi yang ditampilkan pada gambar grafik menunjukkan bahwa hasil data *testing* yang memiliki akurasi paling tinggi dengan akurasi mencapai 95% yaitu pengujian yang menggunakan data *training* sebanyak 1008. Hal ini dikarenakan semakin banyak data *training* yang digunakan maka semakin baik pula hasil yang didapat.

## 5. Kesimpulan

Berdasarkan hasil dari pengujian sistem klasifikasi retensi arsip, dapat di ambil beberapa kesimpulan sebagai berikut :

- Sistem pengarsipan yang dibuat menggunakan Algoritma *Naïve Bayes Classifier* untuk klasifikasi retensi arsip ini berpengaruh terhadap hasil rekomendasi data *testing*. Apabila data *training* tersebut semakin banyak maka rekomendasi retensi akan semakin tepat dan akurat.
- Tingkat keakuratan hasil perhitungan sistem menggunakan Algoritma *Naïve Bayes Classifier* dengan menggunakan 1008 data *training* dari pada pengujian yang dilakukan memiliki hasil keakuratan mencapai 95% bila dibandingkan dengan hasil pakar.

## Daftar Pustaka:

Asichin, M, (2012), *Peraturan Kepala Arsip Nasional Republik Indonesia No. 37 Tahun 2012 tentang Jadwal Retensi Arsip Fasilitas Non Keuangan dan Non Kepegawaian*. Jakarta: Arsip Nasional Republik Indonesia.

Mustofa, Mufid, (2016), *Pengembangan Sistem Pendukung Keputusan Penjurusan Bagi Siswa Baru Menggunakan Metode Naive Bayes (Studi Kasus MAN 1 Malang*. SKRIPSI: Politeknik Negeri Malang.

Pascapraharastyan, Rizki Alfiasca, dkk, (2014), *Rancang Bangun Sistem Informasi Manajemen Arsip Berbasis Web Pada Rumah Sakit Bedah Surabaya*. JSIKA Vol 3, No 1 (2014)/ISSN 2338-137X : STMIK AMIKOM Surabaya.

Prasetyo, Eko, (2012), *DATA MINING – Konsep dan Aplikasi Menggunakan MATLAB*, Ed 1. Yogyakarta : ANDI OFFSET, pp. 59-72.