

Klasifikasi Karya Ilmiah (Tugas Akhir) Mahasiswa Menggunakan Metode *Naive Bayes Classifier* (Nbc)

Nurdin*, M. Suhendri, Yesy Afrilia, Rizal

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Malikussaleh
Jl. Batam, Bukit Indah – Lhokseumawe, Aceh.

*e-mail: nurdin@unimal.ac.id

(received: 21 Desember 2020, revised: 14 Januari 2021, accepted: 5 April 2021)

Abstrak

Tugas akhir atau skripsi merupakan hasil penelitian yang membahas suatu masalah sesuai bidang ilmu dari mahasiswa. Dengan bertambah jumlah lulusan, maka jumlah dokumen tugas akhir yang dihasilkan juga akan semakin besar. Jumlah dokumen karya ilmiah atau tugas akhir yang besar akan sulit dicari sesuai dengan topik jika tidak dikelompokkan. Jumlah dokumen yang besar akan tidak efektif jika dilakukan klasifikasi secara manual. Penelitian ini membuat aplikasi klasifikasi karya ilmiah bertujuan untuk mengklasifikasikan karya ilmiah (tugas akhir) mahasiswa dalam bidang ilmu Teknik Informatika. Aplikasi ini dibangun dengan mengimplementasikan algoritma *Naive Bayes Classifier* berdasarkan parameter latar belakang dan akan diklasifikasikan menjadi 5 kategori yaitu pengolahan citra, data *mining*, sistem pengambilan keputusan, sistem informasi geografis dan sistem pakar. Dengan tahapan penelitian yaitu pengumpulan data, *preprocessing*, perhitungan metode *Naive Bayes Classifier*, implementasi dan pengujian sistem. Penelitian ini menggunakan data sebanyak 170 data karya ilmiah, yang dibagi menjadi 150 data untuk pelatihan dan 20 data untuk pengujian. Hasil penelitian ini menggambarkan bahwa algoritma *Naive Bayes Classifier* merupakan algoritma sederhana yang mampu digunakan untuk melakukan klasifikasi karya ilmiah dengan rata-rata akurasi 86,68% serta rata-rata waktu proses yang dibutuhkan dalam setiap pengujian yaitu 5,7406 detik/pengujian.

Kata Kunci: Karya ilmiah, *Naive bayes classifier*, Klasifikasi, Pelatihan, Pengujian.

Abstract

The final project or thesis is the result of research that addresses a problem according to the student's field of science. By increasing the number of graduates, the number of final project documents produced will also be even greater. The large number of scientific papers or final project documents will be difficult to find according to the topic if they are not grouped. A large number of documents will not be effective if classification is done manually. This study makes a scientific paper classification application aimed at classifying the scientific work (final project) of students in the field of Informatics Engineering. This application was built by implementing the Naive Bayes Classifier algorithm based on background parameters and will be classified into 5 categories, namely image processing, data mining, decision making systems, geographic information systems and expert systems. With the research stages, namely data collection, preprocessing, calculation of the Naive Bayes Classifier method, implementation and system testing. This study uses 170 scientific papers, which are divided into 150 data for training and 20 data for testing. The results of this study illustrate that the Naive Bayes Classifier algorithm is a simple algorithm that can be used to classify scientific papers with an average accuracy of 86.68% and the average processing time required in each test is 5.7406 seconds / test.

Keywords: *scientific work, naive bayes classifier, classification, training, testing*

1 Pendahuluan

Dampak dari perkembangan teknologi informasi sangat signifikan bagi kehidupan sehari-hari, salah satu manfaatnya untuk membantu kegiatan yang membutuhkan tingkat ketelitian tinggi. Kegiatan yang dikerjakan oleh sebuah instansi atau perusahaan adalah kegiatan pengolahan dokumen.

<http://sistemasi.ftik.unisi.ac.id>

padaperguruan tinggi, pengarsipan yang sering dilakukan yaitu pengarsipan sebuah karya ilmiah berupa laporan kerja praktek, laporan penelitian, tugas akhir, tesis dan lain sebagainya.

Tugas akhir atau skripsi merupakan hasil penelitian yang membahas suatu masalah sesuai bidang ilmu dari mahasiswa dengan menggunakan aturan yang sudah ditetapkan serta dibimbing oleh dosen pembimbing[1]. Tugas akhir merupakan salah satu syarat yang harus dipenuhi oleh mahasiswa untuk menyelesaikan pendidikan di perguruan tinggi. Ilmu yang diperoleh mahasiswa dituangkan kedalam suatu penelitian yang nantinya akan menghasilkan keluaran berupa dokumen tugas akhir[2].

Dengan bertambah jumlah lulusan, maka jumlah dokumen tugas akhir yang dihasilkan juga akan semakin besar. Jumlah dokumen karya ilmiah atau tugas akhir yang besar akan terasa sulit dicari sesuai dengan topik yang diinginkan jika tidak dikelompokkan sesuai dengan topiknya. Kesulitan yang dialami oleh banyak para pengambil kebijakan di perguruan tinggi adalah dalam menentukan klasifikasi karya ilmiah atau tugas akhir. Karena selama ini dalam penentuan klasifikasi karya ilmiah atau tugas akhir hanya berdasarkan pada perkiraan terhadap isi konten yang diteliti oleh mahasiswa. Sehingga kesesuaian antara topik dengan isi terkadang sangat berbeda.

Melihat dan menganalisa dari permasalahan yang ada, maka diperlukan sebuah sistem yang bisa mengelompokkan tugas akhir secara otomatis. Sistem yang akan dibangun ini diharapkan dapat membantu pihak jurusan dalam mengklasifikasikan tugas akhir yang diajukan mahasiswa lebih efektif dan efisien. Dalam penelitian ini, penulis akan mengklasifikasikan tugas akhir kedalam beberapa bidang keahlian pada jurusan teknik informatika.

Text mining merupakan data yang berupa teks dengan sumber data diperoleh dari dokumen, bertujuan untuk mencari kembali kata yang bisa mewakili isi dari sebuah dokumen yang nantinya dapat dilakukan proses analisa dan mencari hubungan antar dokumen[3],[4]. *Text mining* adalah suatu proses untuk mengekstraksi pola berupa informasi dan pengetahuan yang berguna dari sejumlah besar sumber data text, seperti dokumen berupa word, pdf dan kutipan teks lainnya[5].

Text Preprocessing merupakan proses mengubah bentuk data yang tidak terstruktur menjadi data terstruktur (perubahan teks menjadi *term index*) yang bertujuan untuk mengecilkan dimensi data sehingga menjadi proses komputasi bisa menjadi lebih efisien dan lebih presisi [6]. Adapun tahapan dari *preprocessing* berdasarkan *case folding*, *tokenizing/parsing*, *filtering* dan *stemming* [7].

Naive Bayes Classifier (NBC) merupakan sebuah metode pengelompokkan berdasarkan pada teorema bayes [8]. Pada penelitian ini penulis menggunakan metode *naive bayes classifier* karena pada metode ini memiliki akurasi lebih baik dibandingkan dengan model klasifikasi *Neural Network* dan *Decision Tree*. Penelitian yang dilakukan oleh Xhemali, Hinde dan Stoned dalam artikelnya menjelaskan bahwa "*Naive Bayes Classifier* memiliki tingkat akurasi lebih baik dibanding model klasifikasi *Neural Network* dan *Decision Tree*".

Penelitian ini menggunakan metode *naive bayes classifier* untuk menentukan kategori karya ilmiah (tugas akhir) yang terdapat di Program Studi Teknik Informatika Universitas Malikussaleh. Berdasarkan permasalahan yang telah dipaparkan sebelumnya, adapun tujuan dilakukan penelitian ini adalah untuk menentukan kategori karya ilmiah berdasarkan pada bidang keahlian yaitu bidang keahlian Pengolahan Citra, Data Mining, Sistem pengambilan Keputusan, Sistem Informasi Geografis dan Sistem Pakar. Hasil akhir berupa, hasil klasifikasi karya ilmiah menggunakan metode *naive bayes classifier*.

2 Tinjauan pustaka

2.1 Penelitian Terkait

Penelitian terdahulu atau penelitian terkait ini menjadi salah satu acuan penulis dalam melakukan penelitian sehingga penulis dapat memperkaya teori yang digunakan dalam mengkaji penelitian yang dilakukan dan sebagai referensi dalam melakukan penelitian. Berikut ini penelitian terkait yang berhubungan dengan penelitian penulis.

Tabel 1. Penelitian Terkait

Nama Peneliti	Analisis Hasil Penelitian dan Perbedaan dengan Penelitian Penulis
Radikdo, dan Aidina	Hasil dari penelitiannya yaitu klasifikasi abstrak skripsi menggunakan naive bayes sesuai bidang kajian, didapatkan dari hasil pengujian 9 siklus menghasilkan pengetahuan bahwa siklus ke 1 merupakan siklus terbaik

<http://sistemasi.ftik.unisi.ac.id>

Ristyawan (2018)	dengan akurasi 82,76% [15]. Perbedaan: Penelitian yang penulis lakukan ini berfokus pada klasifikasi karya ilmiah menggunakan naive bayes classifier dengan latar belakang tugas akhir sebagai objeknya
Devita, dkk (2018)	Pada penelitian devita, dkk berfokus pada klasifikasi artikel bahasa Indonesia dengan membandingkan kinerja 2 metode yaitu metode KNN dan NBC, penelitiannya menggunakan data jurnal sebagai data latih dan uji serta abstrak sebagai objeknya, didapatkan hasil bahwa metode naive bayes memiliki kinerja yang lebih baik dengan akurasi 70%, sedangkan metode KNN memiliki akurasi 40% [16]. Sedangkan dalam penelitian yang penulis lakukan berfokus pada klasifikasi tugas akhir mahasiswa yang hanya menggunakan metode naive bayes sebagai proses klasifikasinya serta latar belakang sebagai objek dari penelitian.
Pramudita, dkk (2018)	Hasil dari penelitiannya yaitu klasifikasi berita olahraga menggunakan naive bayes classifier dengan enhanced confix stripping stemmer, didapat hasil dengan tingkat keakuratan 77% serta metode naive bayes classifier mampu digunakan dalam klasifikasi berita olahraga dengan baik [6]. Perbedaan: Penelitian yang penulis lakukan berfokus pada klasifikasi tugas akhir mahasiswa Teknik Informatika Unimal menggunakan naive bayes classifier dan tidak menggunakan tambahan stemmer.
Latif, Syukriyanto, dkk (2018)	Hasil dari penelitiannya yaitu klasifikasi abstrak isiyang diunduh di www.computer.org. menggunakan naive bayes, dari total data 120 dengan 75% data latih dan 25% data uji. Didapat hasil tingkat akurasi dari 3 kategori DM, ITS, dan MM adalah 100%, 100% dan 86% dan nilai kecepatan pembelajaran antara 0,1-0,5 [17]. Perbedaan: Penelitian yang penulis lakukan berfokus pada klasifikasi karya ilmiah menggunakan metode naive bayes.
Aung, aye aye myint, dan Su wai hlaing (2018)	Penelitian yang dilakukan oleh Aung yaitu mengenai tentang teks mining dalam situs berita menggunakan metode naive bayes classifier yang dibangun dengan bahasa pemrograman python didapat hasil yang cukup baik dari 5 kategori yaitu bisnis, hiburan, politik, olahraga dan teknologi dengan total data 1967 yang dibagi menjadi 2/3 data sebagai data training dan 1/3 data sebagai data testing dan didapatkan akurasi sebesar 91% [18]. Sedangkan penelitian yang penulis lakukan yaitu mengenai tentang klasifikasi karya ilmiah menggunakan metode naive bayes yang dibangun menggunakan bahasa pemrograman php dan MySQL.
Buzic, Dalibor dan Jasminka Dobsa (2018)	Penelitian yang dilakukan Buzic, Dalibor, dan Jasminka Dobsa yaitu klasifikasi lirik lagu dengan menggunakan metode naive bayes yang dibagi menjadi dua kategori yaitu Nirvana dan Metallica. Pengukuran evaluasi model menunjukkan hasil yang sangat baik yaitu, presisi 0,93, recall 0,95 dan f-measure 0,94. Maka dengan hasil tersebut metode naive bayes dapat dilakukan dalam pengklasifikasian lirik [19]. Sedangkan penulis meneliti tentang klasifikasi karya ilmiah yang dibagi menjadi 5 kategori dengan menggunakan metode naive bayes, hanya saja persamaan antara penelitian penulis dengan penelitian Buzic, dkk yaitu sama-sama melakukan proses klasifikasi teks.
B, Dwi Pramita B, dkk (2018)	Penelitian yang penulis lakukan yaitu klasifikasi karya ilmiah (tugas akhir) menggunakan algoritma naive bayes classifier berdasarkan bidangnya dengan menganalisis latar belakang. Sedangkan dalam penelitian B, Dwi Pramita B, dkk yaitu klasifikasi artikel akademik berdasarkan bidangnya dengan menganalisis judul dan abstraknya menggunakan metode naive bayes classifier, dimana didapat dalam menggunakan 292 artikel sebagai data latih dan 100 artikel sebagai data uji. Penelitian tersebut diuji dengan menerapkan nilai threshold 5 dari 1 hingga 2,5 dengan setiap threshold dijalankan 5 kali [20].

Penelitian yang peneliti lakukan hanya berfokus pada klasifikasi karya ilmiah dengan objek penelitian latar belakang dari karya ilmiah dengan menggunakan algoritma *Naive Bayes Classifier*. Penelitian ini membuat sistem aplikasi dengan menggunakan bahasa pemrograman PHP dan *MySQL*. Dari beberapa penelitian terdahulu yang terkait telah dijelaskan dan dilakukan analisis di atas belum ada yang melakukan penelitian ini. Selain penelitian ini, ada beberapa penelitian yang telah dilakukan oleh penulis yang berhubungan dengan pembuatan sistem aplikasi, diantaranya aplikasi sistem penterjemahan ayat pada surah yasin [9], sistem pendeteksian dokumen plagiarisme [10], sistem pendeteksi kemiripan judul skripsi menggunakan algoritma winnowing [11], sistem pendeteksian pola lafadz Allah dan Muhammad pada citra Al-Quran [12], aplikasi quick count pilkada berbasis android [13], dan sistem pengecekan toko *online* pada shoope menggunakan algoritma *BFS* [14].

3 Metode Penelitian

NBC adalah metode pengelompokan atau klasifikasi yang berdasarkan pada teorema Bayes. Metode klasifikasi ini menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris bernama Thomas Bayes, yaitu suatu metode untuk memprediksikan peluang di masa depan berdasarkan pada pengalaman di masa lalu, sehingga metode ini lebih dikenal dengan Teorema Bayes [9]. Metode NBC sering disebut dengan *Bayesian Classification* yang merupakan suatu metode klasifikasi statistik yang berdasarkan pada teorema bayes yang digunakan untuk melakukan prediksi probabilitas keanggotaan pada suatu kelas [21].

NBC dalam penelitian ini digunakan untuk klasifikasi dokumen teks. Dalam proses pelatihan dokumen harus ditentukan kategorinya (data latih). Untuk menghitung setiap kata pada dokumen pelatihan digunakan persamaan 1, dan untuk menghitung probabilitas kategori dokumen menggunakan persamaan 2 [16].

$$p(w_i|c_j) = \frac{1+n_j}{n+|kosakata|} \quad (1)$$

dimana:

- $p(w_i|c_j)$: probabilitas kata dari setiap kategori
- n_j : frekuensi kata dari setiap kategori
- n : jumlah seluruh kata pada dokumen
- $|kosakata|$: jumlah total kata di semua data pelatihan

$$p(c_j) = \frac{n(doc_j)}{n(sampel)} \quad (2)$$

dimana:

- $p(c_j)$: probabilitas dokumen kategori (prior)
- $n(doc_j)$: jumlah seluruh dokumen
- $n(sampel)$: jumlah seluruh dokumen pelatihan

Setelah proses pelatihan selesai, maka dilakukan proses klasifikasi. Pada proses ini dokumen yang dipakai belum diketahui apa kategorinya, sehingga membutuhkan metode *naive bayes* untuk mencari kata pada data uji yang sesuai dengan data latih $p(w_i|c_j)$. Kemudian menghitung probabilitas setiap dokumen $p(c_j)$, maka untuk setiap proses kategori dokumen dihitung dengan persamaan 3.

$$p(c_j) \prod_i p(w_i|c_j) \quad (3)$$

Untuk mencari nilai $p(w_i|c_j)$ dilakukan dengan mengalikan nilai probabilitas kemunculan kata yang sama saat data dilatih dengan nilai probabilitas dokumen sesuai kategorinya $p(c_j)$. Setelah diperoleh hasil perkalian padasetiap kategori dokumen, proses berikutnya membandingkan dan mencari nilai probabilitas terbesar C_{MAP} yang akan digunakan untuk mengklasifikasi data uji, perhitungan tersebut dapat dilihat pada persamaan 4.

$$C_{MAP} = \operatorname{argmax}_{c_j \in C} p(c_j) \prod_i p(w_i|c_j) \quad (4)$$

Tahapan yang dilakukan dalam Klasifikasi karya ilmiah (tugas akhir) mahasiswa menggunakan metode *naive bayes classifier (NBC)* penelitian ini sebagai berikut:

a. Pengumpulan Data

Pengumpulan data pada penelitian ini diambil berdasarkan kategori yang sudah ditetapkan dalam penelitian ini. Pengumpulan data dilakukan dengan cara mengambil data karya ilmiah (tugas akhir) berupa softcopy di dalam perpustakaan Teknik Informatika Universitas Malikussaleh. Data yang didapatkan dari karya ilmiah tersebut berupa latar belakang yang akan dijadikan sebagai objek dalam penelitian ini.

b. Perancangan Sistem

Perancangan sistem dilakukan dengan beberapa tahap, yaitu perancangan skema sistem (flowchart sistem) dan perancangan antarmuka untuk menghasilkan *wireframe*.

c. Implementasi Sistem

Tahapan yang dilakukan dalam implementasi sistem meliputi implementasi dari perancangan sistem dengan menjalankan antarmuka dari aplikasi.

d. Pengujian Sistem

Tahap ini melakukan pengujian sistem yang telah dilakukan pada tahap implementasi sistem. Tahapan ini untuk mengetahui keberhasilan penerapan algoritma naive bayes untuk menyelesaikan permasalahan klasifikasi dokumen tugas akhir.

4 Hasil dan Pembahasan

4.1 Hasil Pengumpulan Data

Mahasiswa pada umumnya akan melihat kepada judul, abstrak, daftar isi dan latar belakang ketika membaca laporan karya ilmiah atau tugas akhir. Sejauh ini belum ada penelitian yang melibatkan latar belakang untuk proses klasifikasi karya ilmiah.

Penelitian ini menggunakan data karya ilmiah berupa skripsi (tugas akhir) dimana latar belakang dari karya ilmiah tersebut akan dijadikan sebagai objek dalam penelitian tersebut. Data tersebut diambil di perpustakaan Teknik Informatika Universitas Malikussaleh. Data yang terkumpul berjumlah 170 data, terdiri dari 150 data sebagai data latih dan 20 data sebagai data uji. Untuk lebih detailnya dapat dilihat pada tabel 2.

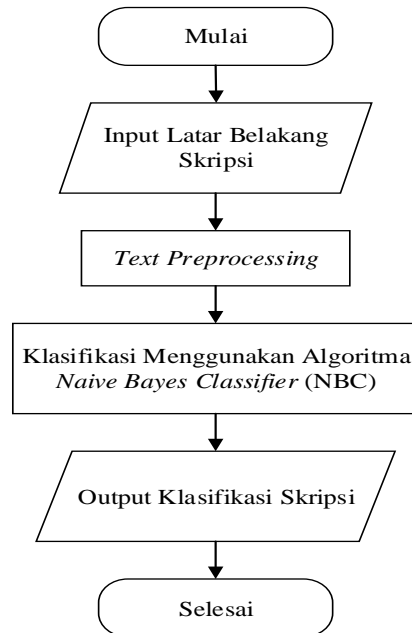
Tabel 2. Hasil Pengumpulan Data

Kategori	Jumlah Data Latih	Jumlah Data Uji
Pengolahan Citra	30 Dokumen	4 Dokumen
Data Mining	30 Dokumen	4 Dokumen
Sistem Pengambilan Keputusan	30 Dokumen	4 Dokumen
Sistem Informasi Geografis (SIG)	30 Dokumen	4 Dokumen
Sistem Pakar	30 Dokumen	4 Dokumen

4.2 Hasil Perancangan Sistem

4.2.1 Skema Sistem

Skema sistem yang dibangun dalam penelitian ini dapat dilihat pada Gambar 1.



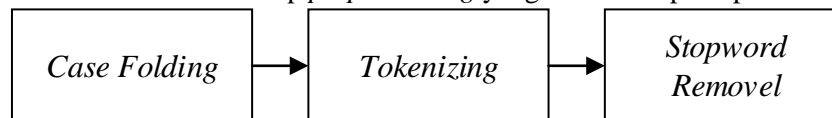
Gambar 1. Skema Sistem

Berikut ini penjelasan dari gambar skema sistem:

- Input Latar Belakang*
Merupakan proses memasukkan data dari latar belakang karya ilmiah
- Text Preprocessing*
Proses untuk dilakukan langkah-langkah *text preprocessing*
- Klasifikasi NBC*
Proses pengklasifikasian karya ilmiah yang dilakukan setelah dokumen melalui tahap *text preprocessing*. Dengan menggunakan *dataset* yang sudah disediakan, maka dilakukan proses pengklasifikasian dengan menggunakan metode NBC dan akan dicocokkan dengan data *training* yang sudah dilakukan *training* sebelumnya dengan metode NBC.
- Output Klasifikasi Karya Ilmiah*
Proses untuk menampilkan hasil klasifikasi yang sebelumnya telah melalui proses pengimputan latar belakang, *text preprocessing* dan proses klasifikasi menggunakan metode NBC.

4.2.2 Tahap *Text Preprocessing*

Tahap *preprocessing* adalah tahapan awal yang merupakan salah satu langkah penting dalam sebuah klasifikasi teks. Berikut ini tahap *preprocessing* yang dilakukan pada penelitian ini.



Gambar 2. Tahap *Text Preprocessing*

- Case Folding*
Tahapan ini merupakan tahap dimana dilakukan proses perubahan pada setiap huruf dalam dokumen sehingga menjadi huruf kecil dan membuang karakter selain a-z.
- Tokenizing*
Tahapan ini dilakukan untuk memotong kalimat setiap kata yang menyusunnya. Proses ini menguraikan deskripsi berupa kalimat-kalimat menjadi kata-kata.
- Stopword Removal*
Tahapan ini merupakan tahap dimana penghapusan kata yang dianggap tidak penting dalam suatu dokumen.

4.3 Implementasi Sistem

4.3.1 Text Preprocessing

Text preprocessing yaitu tahap pembersihan untuk menghilangkan noise pada latar belakang agar menjadi bersih yang siap untuk diolah menjadi klasifikasi karya ilmiah. berikut merupakan hasil *text preprocessing* dari data uji seperti pada tabel 3.

Tabel 3. Text Preprocessing

No	Latar Belakang
1.	tugas peneliti kartographi mengekstraksi fitur data disajikan terperinci memvisualisasikan peta sederhana dianalisa teknologi komputer proses melibatkan otomatis algoritma efisien menyederhanakan mengidentifikasi permasalahan kompleks garis linier menghilangkan pengecilan ukuran file penyimpan titik vektor graphis pendekatan diajukan merepresentasikan openshaw douglas peucker metode direct indirect penelitian dimodelkan sistem pemilihan mudah diprogram menghasilkan struktur hirarki
...
20.	mobile alat komunikasi memudahkan kehidupannya sehari-hari seiring perkembangannya mengakibatkan mobilitas akses cepat dihadirkan perangkat dibawa menikmati kecanggihan perkembangan menunjang aktifitasnya jarak terdekat wisata wisatawan berkunjung halnya pencarian lhokseumawe ditetapkan statusnya undang-undang batas-batas berbatasan selat malaka kuta makmur dewantara syamtalira bayu sumatera banda medan inemiliki eksotis menakjubkan sensasi berlibur menyenangkan ketentuan uu no perjalanan mengunjungi tujuan rekreasi pengembangan pribadi mempelajari keunikan daya tarik dikunjungi jangka dipisahkan kesibukan padat aktifitas beragam berusaha terlepas kesulitan terpendek shortest part problem ditempuh tempuh pendek banyaknya smartphone kecanggihannya primer mulyanto fitur layanan internet dilengkapi gps positioning system satelit navigasi penentuan koordinat latitude longitude longitude geografis permukaan garis lintang disebelah ekuator diukur derajat khatulistiwa kutub bujuradalah meridian bujur greenwich international line algoritma sort pengurutan pengguna bucket radix sorting prinsip mengurutkan data-data satuan wadah ember perbandingan menampilkan kompleksitas tercepat penelitian sebelumnya horsmalahti membahas sorr dibandingkan memori membutuhkan kapasitas dibanding audy uraian dibutuhkan aplikasi berbasis android menemukan penulis berkcinginan tugas berjudul

4.3.2 Perhitungan Manual Metode *Naive Bayes Classifier* (NBC)

Pada perhitungan manual ini, penulis menggunakan data latar belakang yang berguna sebagai data pengujian. Contoh perhitungan manualnya:

Tugas peneliti kartographi mengekstraksi fitur data disajikan terperinci memvisualisasikan peta sederhana dianalisa teknologi komputer proses melibatkan otomatis algoritma efisien menyederhanakan mengidentifikasi permasalahan kompleks garis linier menghilangkan pengecilan ukuran file penyimpan titik vektor grafis pendekatan diajukan merepresentasikan openshaw douglas peucker metode direct indirect penelitian dimodelkan sistem pemilihan mudah diprogram menghasilkan struktur hirarki.

Langkah-langkah perhitungan manual metode *Naive Bayes Classifier* (NBC), adalah sebagai berikut:

1. Menghitung probabilitas dokumen kategori: $P(c_j)$

Tabel 4. Hitung $P(c_j)$

Kategori	$P(c_j)$
Pengolahan Citra	0,2
Data Mining	0,2
Sistem Informasi Geografis	0,2
Sistem Pakar	0,2
Sistem Pengambilan Keputusan	0,2

2. Menghitung nilai frekuensi kata dari setiap kategori: $P(n_i)$

Tabel 5. Hitung $P(n_i)$

Kata	$P(n_i)$				
	Pengolahan Citra	Data Mining	SIG	Sistem Pakar	SPK
Tugas	17	12	6	15	7
...
Hirarki	0	1	2	0	1

3. Menghitung probabilitas kata dari setiap kategori: $P(w_i|c_j)$

Tabel 6. Hitung $P(w_i|c_j)$

Kata	$P(w_i c_j)$				
	Pengolahan Citra	Data Mining	SIG	Sistem Pakar	SPK
Tugas	0.0006244896	0.0006244896	0.0007599867	0.0003762581	0.0003320683
...
Hirarki	0.0000482998	0.0000960753	0.0000474992	0.0000940645	0.0001423150

4. Nilai posterior: $P(c_j) * P(w_i|c_j)$

Tabel 7. Hitung $P(c_j) * P(w_i|c_j)$

Kategori	$P(c_j) * P(w_i c_j)$ atau Posterior
Pengolahan Citra	5.0712716885157E-199
Data Mining	2.3303855837471E-200
Sistem Informasi Geografis	1.3444228726216E-202
Sistem Pakar	3.0427056719605E-201
Sistem Pengambilan Keputusan	6.971689614762E-204

Karena nilai posterior pengolahan citra lebih tinggi dibandingkan dengan nilai posterior kategori yang lainnya, maka hasil klasifikasi dari latar belakang karya ilmiah tersebut adalah pengolahan citra.

4.3.3 Hasil Output

Berdasarkan perancangan sistem yang telah dibuat, maka hasil *output* dari sistem klasifikasi karya ilmiah dengan menerapkan algoritma *Naive Bayes Classifier* dapat dilihat pada Gambar 3.



Gambar 3. Halaman Hasil Pengujian

Untuk menampilkan data klasifikasi karya ilmiah yang telah dilakukan sebelumnya, seluruh data klasifikasi akan ditampilkan pada gambar 4.

No	Judul Karya Ilmiah	Latar Belakang	Kategori	Aksi
1	PENYEDERHANAAN TITIK-TITIK GARIS POLYLINE DENGAN MENGGUNAKAN ALGORITMA DOUGLAS PEUCKER	Tugas penting para peneliti kartografi adalah mengekstraksi berbagai fitur dari data data yang disajikan terperinci dan sederhananya memvisualisasikan data-data tersebut ke dalam bentuk peta sederhana yang dapat dibaca dan diarsip. Teknologi-komputer sangat terlibat dalam proses-proses yang melibatkan kartografi otomatis, sehingga algoritma yang efisien sangat diperlukan oleh komputer untuk mengekstraksi dan menyederhanakan data-data peta awal yang masih terperinci tersebut. Para peneliti kartografi telah mengidentifikasi permasalahan yang kompleks dalam penyederhanaan garis sebagai tugas penting dalam memvisualisasikan berbagai fitur tema. Penyederhanaan digunakan karena berbagai hal, seperti menghilangkan detail yang tidak perlu dan pengecilan ukuran file penyimpanan titik vektor grafis. Berbagai Algoritma atau pendekatan telah diajukan untuk merepresentasikan penyederhanaan garis diantaranya algoritma Li-Openshaws, algoritma Douglas-Peucker, metode direct dan indirect, dan berbagai algoritma lainnya. Di dalam penelitian ini, algoritma Douglas-Peucker dimodelkan untuk sistem penyederhanaan garis. Alasan pemilihan metode Douglas-Peucker adalah (1) mudah untuk diprogram, dan (2) dalam penyederhanaannya algoritma Douglas-Peucker akan menghasilkan suatu struktur hirarki yang baik.	Pengolahan Citra	
2	IDENTIFIKASI KOMPARAN KARYA ILMIAH DENGAN	Karya ilmiah yaitu karya tulis yang telah diakui dalam bidang ilmu pengetahuan, teknologi atau seni. Pada dasarnya karya ilmiah merupakan penemuan logis dari	Data Mining	

Gambar 4. Halaman Hasil Klasifikasi

4.4 Pengujian Sistem

Berdasarkan pengujian klasifikasi karya ilmiah dari 20 data uji latar belakang karya ilmiah, didapat hasil klasifikasi karya ilmiah yang tertera seperti pada Tabel 8.

Tabel 8. Jumlah Hasil Klasifikasi

Kategori	Hasil Klasifikasi
Pengolahan Citra	5
Data Mining	4
Sistem Pengambilan Keputusan	5
Sistem Informasi Geografis (SIG)	3
Sistem Pakar	3

Berdasarkan hasil pengujian 20 data latar belakang yang tertera pada tabel 8, penulis mendapatkan data bahwasannya aplikasi dapat melakukan pengklasifikasian karya ilmiah dengan baik. Aplikasi ini berbasis web yang dijalankan pada Laptop HP V110, dengan spesifikasi: AMD A9-9420 Radeon R5, 3.00 Ghz, VGA AMD Radeon (TM) R5 Graphics, RAM 8 GB dan HDD 500 GB. Waktu yang diperlukan sistem untuk melakukannya setiap data latar belakang dengan menerapkan algoritma *naive bayes classifier* dapat dilihat pada tabel 9.

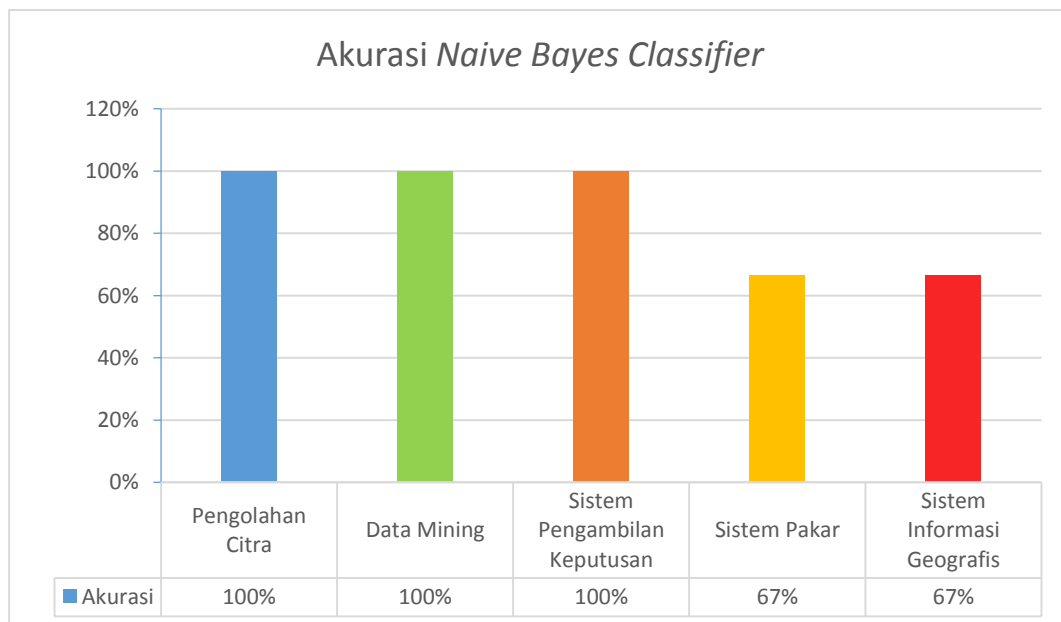
Tabel 9. Waktu Proses Sistem

No	Latar Belakang	Waktu
1.	tugas peneliti kartografi mengekstraksi fitur data disajikan terperinci memvisualisasikan peta sederhana dianalisa teknologi komputer proses melibatkan otomatis algoritma efisien menyederhanakan mengidentifikasi permasalahan kompleks garis linier menghilangkan pengecilan ukuran file penyimpanan titik vektor grafis pendekatan diajukan merepresentasikan openshaw douglas peucker metode direct indirect penelitian dimodelkan sistem pemilihan mudah diprogram menghasilkan struktur hirarki	3,814
...
20.	mobile alat komunikasi memudahkan kehidupannya sehari-hari seiring perkembangannya mengakibatkan mobilitas akses cepat dihadirkan perangkat dibawa menikmati kecanggihan perkembangan menunjang aktifitasnya jarak terdekat wisata	10,527

<http://sistemasi.ftik.unisi.ac.id>

wisatawan berkunjung halnya pencarian lhokseumawe ditetapkan statusnya undang-undang batas-batas berbatasan selat malaka kuta makmur dewantara syamtalira bayu sumatera banda medan inemiliki eksotis menakjubkan sensasi berlibur menyenangkan ketentuan uu no perjalanan mengunjungi tujuan rekreasi pengembangan pribadi mempelajari keunikan daya tarik dikunjungi jangka dipisahkan kesibukan padat aktifitas beragam berusaha terlepas kesulitan terpendek shortest part problem ditempuh tempuh pendek banyaknya smartphone kecanggihannya primer mulyanto fitur layanan internet dilengkapi gps positioning system satelit navigasi penentuan koordinat latitude longitude longitude geografis permukaan garis lintang disebelah ekuator diukur derajat khatulistiwa kutub bujuradalah meridian bujur greenwich international line algoritma sort pengurutan pengguna bucket radix sorting prinsip mengurutkan data-data satuan wadah ember perbandingan menampilkan kompleksitas tercepat penelitian sebelumnya horsmalahiti membahas sorr dibandingkan memori membutuhkan kapasitas dibanding audy uraian dibutuhkan aplikasi berbasis android menemukan penulis berkcinginan tugas berjudul

Berdasarkan tabel di atas, penulis mendapatkan data bahwa dari 20 data pengujian data latar belakang diatas didapatkan rata-rata waktu proses pengujian dengan menerapkan algoritma *Naive Bayes Classifier* adalah 5,7406 detik. Pengujian akurasi dilakukan dengan beberapa pengujian data uji dengan menggunakan data training dengan jumlah seluruh data training 150 data. Hasil pengujian sistem dapat dilihat pada grafik dibawah ini.



Gambar 5. Grafik Akurasi *Naive Bayes Classifier*

Dari pengujian 5 kategori yang penulis lakukan seperti pada gambar 5 didapat rata-rata keakuratan yang cukup baik yaitu sebesar 86,68% dari total data uji sebanyak 20 data latar belakang karya ilmiah mahasiswa.

5 Kesimpulan

Berdasarkan pembahasan yang telah dilakukan maka dapat diambil kesimpulan yaitu algoritma *Naive Bayes Classifier* mampu melakukan proses klasifikasi data karya ilmiah atau tugas akhir secara otomatis dan proses klasifikasi akan semakin akurat jika data latih yang digunakan dalam *training* dengan jumlah yang lebih banyak. Algoritma *Naive Bayes Classifier* memiliki kinerja yang baik dalam klasifikasi karya ilmiah (tugas akhir) ditunjukkan dengan hasil pengujian 20 data karya ilmiah berdasarkan parameter latar belakang menghasilkan 18 data diklasifikasikan dengan benar dan 2 data lainnya terdeteksi salah, dan 20 data pengujian tersebut didapat rata-rata waktu proses pengujian

<http://sistemasi.ftik.unisi.ac.id>

dengan menerapkan algoritma *Naive Bayes Classifier* yaitu 5,7406 detik/pengujian. Dan tingkat akurasi dari pengujian tersebut yang diklasifikasikan kedalam 5 kelas didapatkan nilai rata-rata akurasi yang cukup baik yaitu 86,68%.

Referensi

- [1] A. Salam, V. P. Wicaksana, and K. Hastuti, "Sistem Rekomendasi Penentuan Dosen Pembimbing Tugas Akhir dengan Menggunakan Algoritma Rabin-Karp," *Techno.COM*, vol. 14, no. 3, pp. 225–233, 2015.
- [2] Yusra, D. Olivita, and Y. Vitriani, "Perbandingan Klasifikasi Tugas Akhir Mahasiswa Jurusan Teknik Informatika Menggunakan Metode Naïve Bayes Classifier dan K-Nearest Neighbor," *J. Sains, Teknol. dan Ind.*, vol. 14, no. 1, pp. 79–85, 2016.
- [3] Aprilliana, N. Ransi, and J. Nangi, "Implementasi Text Mining Klasifikasi Skripsi Menggunakan Metode Naive Bayes Classifier," *semantik*, vol. 3, no. 2, pp. 187–194, 2017.
- [4] V. Gupta and G. S. Lehal, "A Survey of Text Mining Techniques and Applications - Volume 1, No. 1, August 2009 - JETWI," *J. Emerg. Technol. Web Intell.*, vol. 1, no. 1, pp. 60–76, 2009.
- [5] H. Nindito, "Teori Text Mining dan Web Mining," *sis.binus.ac.id*, 2016. <https://sis.binus.ac.id/2016/12/15/teori-text-mining-dan-web-mining/> (accessed Jan. 23, 2021).
- [6] Y. D. Pramudita, S. S. Putro, and N. Makhmud, "Klasifikasi Berita Olahraga Menggunakan Metode Naive Bayes Dengan Enhanced Confix Stripping Stemmer," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 3, pp. 269–276, 2018, doi: 10.25126/jtiik.201853810.
- [7] C. Triawati, "Metode Pembobotan Statistical Concept Based untuk Klastering dan Kategorisasi Dokumen Berbahasa Indonesia," *Institut Teknologi Telkom, Bandung*. 2009.
- [8] M. S. Mustafa, M. R. Ramadhan, and A. P. Thenata, "Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Creat. Inf. Technol. J.*, vol. 4, no. 2, p. 151, 2018, doi: 10.24076/citec.2017v4i2.106.
- [9] Nurdin and N. B. Puteri, "Perbandingan Metode Transformasi Wavelet Dan Fourier Dalam Penterjemahan Ayat Pada Surah Yasin," *Sist. J. Sist. Inf.*, vol. 9, no. 1, p. 16, 2020, doi: 10.32520/stmsi.v9i1.563.
- [10] Nurdin, Rizal, and Rizwan, "Pendeteksian Dokumen Plagiarisme Dengan Menggunakan Metode Weight Tree," *J. Telemat.*, vol. 1, no. 1, pp. 31–45, 2019.
- [11] Nurdin and A. Munthoha, "Sistem Pendeteksi Kemiripan Judul Skripsi Menggunakan Algoritma Winnowing," *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 2, no. 1, pp. 90–97, 2017.
- [12] Nurdin, D. Hamdhana, and M. J. Setiawan, "Sistem Pendeteksi Pola Lafadz Allah Dan Muhammad Pada Citra Al-Qur'an Menggunakan Metode Peirce," *e-Journal Techsi Tek. Inf.*, vol. 9, no. 2, pp. 78–90, 2017.
- [13] Nurdin, D. Hamdhana, and M. Iqbal, "Aplikasi Quick Count Pilkada Dengan Menggunakan Metode Random Sampling Berbasis Android," *e-Journal Techsi Tek. Inf.*, vol. 10, no. 1, pp. 141–154, 2018.
- [14] Nurdin, M. Hutomi, M. Qamal, and Bustami, "Sistem Pengecekan Toko Online Asli atau Dropship pada Shopee Menggunakan Algoritma Breadth First Search," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 6, pp. 1117–1123, 2020, doi: 10.29207/resti.v4i6.2514.
- [15] A. C. Pradikdo and A. Ristyawan, "Model Klasifikasi Abstrak Skripsi Menggunakan Text Mining Untuk Pengkategorian Skripsi Sesuai Bidang Kajian," *J. SIMETRIS*, vol. 9, no. 2, pp. 1091–1098, 2018.
- [16] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, "Perbandingan Kinerja Metode Naive Bayes Dan K-Nearest Neighbor Untuk Klasifikasi Artikel Berbahasa Indonesia," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, pp. 427–434, 2018, doi: 10.25126/jtiik.201854773.
- [17] S. Latif, U. Suwardoyo, and E. A. Wihelmus Sanadi, "Content Abstract Classification Using Naive Bayes," *J. Phys. Conf. Ser.*, vol. 979, no. 1, pp. 1–8, 2018, doi: 10.1088/1742-6596/979/1/012036.
- [18] A. A. M. Aung and S. W. Hlaing, "Text Mining from News Website using Machine Learning with Naïve Bayes Algorithm," *Int. J. Sci. Eng. Technol. Res.*, vol. 7, no. 10, pp. 707–711,

- 2018.
- [19] D. Buzic and J. Dobsa, "Lyrics Classification using Naive Bayes," *2018 41st Int. Conv. Inf. Commun. Technol. Electron. Microelectron. MIPRO 2018 - Proc.*, no. June, pp. 1011–1015, 2018, doi: 10.23919/MIPRO.2018.8400185.
 - [20] D. P. B. Bestari, R. Saptono, and R. Anggrainingsih, "Academic Articles Classification Using Naive Bayes Classifier (Nbc) Method," *J. Ilm. Teknol. dan Inf.*, vol. 7, no. 2, pp. 74–81, 2018.
 - [21] H. Muhamad, C. A. Prasajo, N. A. Sugianto, L. Surtiningsih, and I. Cholissodin, "Optimasi Naive Bayes Classifier Dengan Menggunakan Particle Swarm Optimization Pada Data Iris," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 3, pp. 180–184, 2017.