Jurnal Ilmiah Ilmu Komputer
Banthayo Lo Komputer Vol. 3, No. 1 (2024)
E-ISSN : 2827-9425
P-ISSN : 2828-4666

PENERAPAN ALGORITMA *NAIVE BAYES CLASSIFIER* UNTUK KLASIFIKASI JUDUL SKRIPSI BERDASARKAN KONSENTRASI

Salmin Dania[1], Rezqiwati Ishak, M.Kom[2], Hastuti Dalai M.Kom[3]
Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Ichsan, Kota Gorontalo, Indonesia
Email:salmindania275@gmail.com

Abstrak - Penerapan algoritma *Naive Bayes Classifier* untuk klasifikasi judul skripsi berdasarkan konsentrasi merupakan penelitian yang bertujuan untuk mengembangkan sistem klasifikasi judul skripsi berdasarkan konsentrasi menggunakan algoritma *Naive Bayes Classifier*. Sistem klasifikasi ini dapat digunakan untuk membantu mahasiswa dalam menentukan konsentrasi skripsi yang sesuai dengan minat dan kemampuannya. Penelitian ini menggunakan data judul skripsi dari jurusan Teknik Informatika, Fakultas Ilmu Komputer Universitas Ichsan Gorontalo dengan menggunakan atribut data berupa judul skripsi dan konsentrasi. Data tersebut dibersihkan dan *dipreprocessing* terlebih dahulu sebelum digunakan untuk pelatihan dan pengujian algoritma. Algoritma *Naive Bayes Classifier* diimplementasikan menggunakan bahasa pemrograman *Python*. Hasil penelitian menunjukkan bahwa algoritma *Naive Bayes Classifier* dapat mengklasifikasi judul skripsi dengan akurasi sebesar 80% pada proses evaluasi model menggunakan *Confusion Matrix.* Hasil ini menunjukkan bahwa algoritma *Naive Bayes Classifier* dapat menjadi alternatif yang efektif untuk klasifikasi judul skripsi berdasarkan konsentrasi.

Kata Kunci: klasifikasi, judul skripsi, konsentrasi, *Python, Confusion Matrix, Naive Bayes Classifier*

Abstract *The application of the Naive Bayes Classifier algorithm to classify thesis titles based on concentration is research that aims to develop a classification system for thesis titles based on concentration using the Naive Bayes Classifier algorithm. This classification system helps students determine the thesis concentration that suits their interests and abilities. This research uses thesis title data from the Informatics Engineering Department, Faculty of Computer Science, Universitas Ichsan Gorontal. It employs data attributes in the form of thesis title and concentration. The data are cleaned and preprocessed before being used for algorithm training and testing. The implementation of the Naive Bayes Classifier algorithm is through the Python programming language. The research results show that the Naive Bayes Classifier algorithm can classify thesis titles with an accuracy of 80% in the model evaluation process using the Confusion Matrix. The results indicate that the Naive Bayes Classifier algorithm is an effective alternative for classifying thesis titles based on concentration.*

Keywords: *classification, thesis title, concentration, Python, Confusion Matrix, Naive Bayes Classifier*

1. PENDAHULUAN

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. Pengetahuan yang didapatkan dituangkan dalam sebuah karya ilmiah yang akan menghasilkan dokumen tugas akhir yang baik dan bermanfaat[1]. Klasifikasi merupakan salah satu teknik yang paling banyak digunakan dalam *machine learning*. Klasifikasi teks adalah proses pengklasifikasian data menurut kelompok atau kelas yang telah ditentukan sebelumnya. Menurut Nicolasi klasifikasi terdiri dari dua tahap; tahap pembelajaran yang menganalisis data pelatihan dan menetapkan aturan klasifikasi untuk data tersebut; dan tahap klasifikasi yang mengklasifikasikan data uji menggunakan aturan yang dihasilkan ke dalam kelompok dimana kelompok tersebut didefinisikan berdasarkan nilai atribut data[2]. Salah satu metode pengelompokan data yang dapat

E-ISSN : 2827-9425
P-ISSN : 2828-4666

digunakan adalah metode *Naive Bayes Classifier. Naive Bayes Classifier* (NBC) adalah salah satu dari algoritma supervised document classification yang sederhana tetapi efisien. Model probabilistik dari algoritma ini didasarkan pada teori Bayes. *Naive Bayes Classifier* telah diterapkan di berbagai bidang antara lain, kedokteran, perbankan, perpustakaan, instalasi perkantoran, dan lain sebagainya[3]. Beberapa penelitian menggunakan metode *Naive Bayes Classifier* diantaranya yang dilakukan oleh Utomo Pujianto, Triyanna Widiyaningtyas, Didik Dwi Prasetya, dan Bintang Romadhon dengan judul "Penerapan algoritma *Naive Bayes Classifier* Untuk Klasifikasi Judul Skripsi dan Tugas Akhir Berdasarkan Kelompok Bidang Keahlian". Pengujian performa penerapan algoritma *Naive Bayes Classifier* menggunakan teknik *K-Fold Cross Validation*, dengan jumlah tahap pengujian sebanyak 10 kali, terhadap 1103 judul skripsi dan tugas akhir, didapatkan hasil rata-rata akurasi 94%, presisi 80%, dan recall 69%. Selanjutnya penelitian yang dilakukan oleh Nurdin, M. Suhendri, Yesy Afrilia, dan Rizal dengan judul "Klasifikasi Karya Ilmiah (Tugas Akhir) Mahasiswa Menggunakan Metode Naive Bayes Classifier (Nbc)". Hasil pengujian 20 data karya ilmiah berdasarkan parameter latar belakang menghasilkan 18 data diklasifikasikan dengan benar dan 2 data lainnya terdeteksi salah. Dan tingkat akurasi dari pengujian tersebut yang diklasifikasikan ke dalam 5 kelas didapatkan nilai rata-rata akurasi yang cukup baik yaitu 86,68%

2. TINJAUAN PUSTAKA

2.1 Skripsi

Skripsi merupakan karya tulis ilmiah yang ditulis oleh mahasiswa sebagai tugas akhir dalam rangka menyelesaikan studinya pada program sarjana (S1). Dalam penulisan skripsi mahasiswa dibimbing oleh dosen pembimbing skripsi dengan mengacu pada buku panduan penulisan karya ilmiah yang telah ditetapkan oleh masing-masing perguruan tinggi. Melalui bimbingan oleh dosen pembimbing dalam menulis skripsi, diharapkan skripsi yang ditulis mahasiswa dapat memenuhi standar penulisan karya tulis ilmiah[9].

2.2 Konsentrasi

Pemilihan konsentrasi dalam kegiatan akademik memang bukan hal yang mudah karena tergantung pada minat, bakat dan keinginan. Oleh karena itu perlu pertimbangan yang matang supaya mahasiswa tidak salah dalam memilih konsentrasi yang diinginkan. Pemilihan konsentrasi yang asal-asalan tanpa pertimbangan yang matang, menyebabkan dampak negatif pada mahasiswa, yaitu kesulitan dalam penyerapan materi-materi perkuliahan.

2.3 Text Mining

Pada *text mining* selalu melibatkan pra proses dokumen yaitu; melakukan kategorisasi teks, melakukan ekstraksi informasi, dan mengekstraksi kata. Metode ini untuk mengekstraksi informasi yang diambil dari sumber data dengan cara mengidentifikasi serta melakukan eksplorasi pola yang menarik. *Text mining* merupakan teknik yang digunakan untuk menangani permasalahan klasifikasi, *clustering*, *information extraction* dan *information retrival*. Secara umum *text mining* terdiri dari tiga langkah yaitu: teks preprocessing, operasi penggalian teks, postprocessing.

2.4 Klasifikasi

Klasifikasi adalah proses pencarian sekumpulan model atau fungsi yang menggambarkan dan membedakan kelas data dengan tujuan agar model tersebut dapat digunakan untuk memprediksi kelas dari suatu objek yang belum diketahui kelasnya. Model itu sendiri bisa berupa aturan "jika-maka", berbentuk pohon keputusan (*decision tree*), formula matematis seperti *naive bayesian* dan *support vector machine*.

2.5 *Naive Bayes Classifier*

Algoritma *Naive Bayes* adalah salah satu algoritma yang terdapat pada teknik data mining klasifikasi. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuan Inggris yaitu Thomas Bayes, *Naive Bayes* memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya, sehingga dikenal dengan *Teorema Bayes*. Teorema tersebut dikombinasikan dengan *Naive* dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi *Naive Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya (Bustami, 2013)[8].

E-ISSN : 2827-9425
P-ISSN : 2828-4666

Persamaan dari teorema bayes adalah :

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

Keterangan :
X : Data dengan class yang belum diketahui
H : Hipotesis data merupakan suatu class spesifik
P(H|X) : Probabilitas hipotesis berdasar kondisi (posteriori probability)
P(H) : Probabilitas hipotesis (*prior probability*)
P(H|X) : Probabilitas berdasarkan kondisi pada hipotesis
P(X) : Probabilitas

2.6 *Python*

Python adalah bahasa pemrograman *interpretatif* multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. *Python* diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang besar serta *komperhensif*. *Python* juga didukung oleh komunitas yang besar[13].

2.7 *Confusion Matrix*

Confusion matrix adalah tabel yang menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah. Contoh confusion matrix untuk klasifikasi biner ditunjukkan pada Tabel 2.7 berikut ini.

Tabel 2.7 *Confusion Matrix*

		Kelas Prediksi	
		1	0
Kelas Sebenarnya	1	TP	FN
	01	FP	TN

Keterangan:
TP (True Positive) = jumlah dokumen dari kelas 1 yang benar diklasifikasikan sebagai kelas 1
TN (True Negative) = jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0
FP (False Positive) = jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1
FN (False Negative) = jumlah dokumen dari kelas 1yang salah diklasifikasikan sebagai kelas 0

Rumus *confusion matrix* untuk menghitung accuracy, precision, dan recall seperti berikut.

3. METODE PENELITIAN

3.1 Jenis, Metode, Subjek, Objek, Waktu, dan Lokasi Penelitian

Dilihat dari tingkat penerapannya penelitian ini adalah penelitian terapan dan jenis informasi yang diolah pada penelitian ini adalah penelitian kuantitatif. Dilihat dari informasi data, maka penelitian ini adalah penelitian konfirmatori. Dengan demikian jenis penelitian ini adalah penelitian deskriptif. Subjek penelitian ini adalah "**Klasifikasi Judul Skripsi Menggunakan Metode *Naive Bayes Classifier"***. Penelitian ini dimulai November 2022 sampai dengan Februari 2023 yang berlokasi di Universitas Ichsan Gorontalo.

3.2 Pengumpulan Data

Adapun jenis pengumpulan data ini yaitu data primer dan data sekunder. Data primer adalah data yang dikumpulkan langsung dilapangan, sedangkan data sekunder adalah data yang dikumpulkan dari penelitian sebelumnya seperti jurnal yang membahas *data mining* serta membahas klasifikasi yang menggunakan metode *naive bayes*, baik dari internet maupun dari perpustakaan.

3.3 Pemodelan

3.3.1 Pra Pengolahan Data

Melakukan pemilihan pada data yang akan diolah nantinya, agar sesuai dengan data yang dibutuhkan. Hal ini dilakukan untuk menetapkan standar atribut yang selanjutnya akan dijadikan sebagai tolak ukur untuk mengukur kontribusi setiap atribut terhadap klasifikasi data.

E-ISSN : 2827-9425
P-ISSN : 2828-4666

3.3.2 Validasi

Tujuan validasi adalah untuk memisahkan data awal menjadi data training dan data testing. Data training adalah data yang akan diproses dengan menggunakan metode klasifikasi, sedangkan data testing adalah data yang akan digunakan dalam proses pengujian dengan menggunakan program komputer.

4. HASIL DAN PEMBAHASAN

4.1 Hasil Pengumpulan Data

Dataset atau data penelitian yang digunakan pada penelitian ini adalah merupakan data yang diambil dari jurusan Teknik Informatika, Universitas Ichsan Gorontalo terkait judul skripsi Mahasiswa sebanyak 404 data, dari tahun akademik 2020/2021 sampai dengan 2022/2023. Adapun dataset yang dikumpulkan dapat dilihat pada **Tabel 4.1** berikut ini :

Tabel 4.1 Hasil Pengumpulan Dataset

No	NIM	Nama Lengkap	Judul Skripsi	Ket
1	T3117369	Marsela Herawati Sarga	PENERAPAN METODE PREFENCE SELECTION INDEX UNTUK SISTEM PENDUKUNG KEPUTUSAN PENILAIAN KINERJA APARAT DESA	SE
2	T3117259	Isda Moha	sistem pendukung keputusan penerima bantuan rumah rehab pada desa botubilotahu menggunakan metode vikor	SE
...	...	...	...	...
404	T3117306	Reska Ngabito	Sistem pendukung keputusan penetapan penerima bantuan rumah rehab menggunakan metode fuzzy mamdani	SE

4.2 Tahapan *Naive Bayes Classifier*

1. Data Training

Tabel 4.1 Data Training

No	Judul Skripsi	Ket
1	PENERAPAN METODE PREFENCE SELECTION INDEX UNTUK SISTEM PENDUKUNG KEPUTUSAN PENILAIAN KINERJA APARAT DESA	SE
2	sistem pendukung keputusan penerima bantuan rumah rehab pada desa botubilotahu menggunakan metode vikor	SE
...	...	...
404	Sistem pendukung keputusan penetapan penerima bantuan rumah rehab menggunakan metode fuzzy mamdani	SE

E-ISSN : 2827-9425
P-ISSN : 2828-4666

2. Preprocessing Data

Tabel 4.6 Hasil Preprocessing Data

No	Judul Skripsi	Ket
1	penerapan metode prefence selection index sistem pendukung keputusan penilaian kinerja aparat desa	SE
2	sistem pendukung keputusan penerima bantuan rumah rehab desa botubilotahu metode vikor	SE
...	...	...
404	sistem pendukung keputusan penetapan penerima bantuan rumah rehab metode fuzzy mamdani	SE

3. Ekstraksi Fitur

Pada tahapan ini dilakukan pengubahan teks judul skripsi menjadi representasi angka.

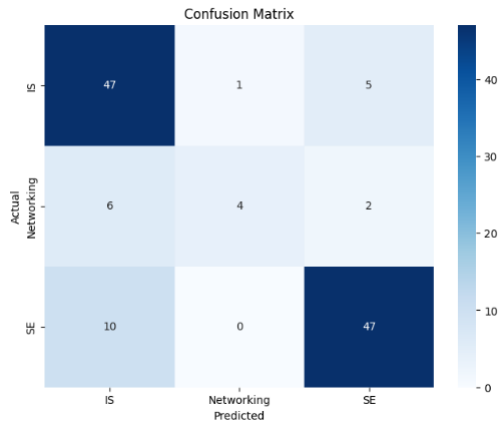
```
(0, 434)      0.38292243566428324
(0, 478)      0.11035805047823478
(0, 861)      0.38292243566428324
(0, 406)      0.43937947843996056
```

4. Pelatihan Model *Naive Bayes*

Pada tahapan ini dilakukan pembagian data menjadi *data training* dan *data testing* . Model *Naive Bayes* dilatih menggunakan data yang telah diproses, model akan memperhitungkan probabilitas munculnya kata-kata dalam setiap kelas berdasarkan *data training*.

5. Evaluasi Model

Setelah dilakukan pelatihan model *Naive Bayes* perlu dilakukan evaluasi terhadap model yang sudah dibuat dengan melihat hasil akurasi.



Tabel 5.1 hasil Akurasi

No.	Konsentrasi	Precision	Recall	F1-Score	Support
1	IS	0.75	0.89	0.81	53
2	Networking	0.80	0.33	0.47	12
3	SE	0.87	0.82	0.85	57
4	Accuracy		0.80		122

E-ISSN : 2827-9425
P-ISSN : 2828-4666

6. Klasifikasi Judul Skripsi

Setelah model dilatih dan dievaluasi dengan baik, maka dapat digunakan untuk mengklasifikasikan teks yang baru atau belum terlihat.

5.2 Penerapan Algoritma Naïve Bayes Classifier

1. Data training yang telah melewati tahap preprocessing dan data uji.

Tabel 5.2 Data *Training* Penerapan Algoritma

No	Data Training	Ket
1	sistem pendukung keputusan penetapan penerima bantuan rumah rehab metode fuzzy mamdani	SE
2	prediksi hasil produksi jagung metode exponsial moving average	IS
3	analisis kualitas jaringan backbone nirkabel blok plan perkantoran standar quality of service qos	Networking

Tabel 5.3 Data *Testing* Penerapan Algoritma

No	Data Testing
1	Analisis Monitoring Jaringan Dengan Aplikasi The Dude di Universitas Ichsan Gorontalo

2. Pemecahan teks kalimat (*Tokenizing*) menjadi kumpulan kata agar mudah dalam melakukan pembobotan tiap kata.

Tabel 5.4 *Tokenizing*

Judul 1	Judul 2	Judul 3
Sistem	Prediksi	Analisis
Pendukung	Hasil	Kualitas
Keputusan	Produksi	Jaringan

3. Pembobotan menggunakan *term frekuensi* (jumlah kemunculan kata).

Tabel 5.5 *Term Frekuensi*

No	Kata	SE	IS	Networking
1	Sistem	1	0	0
2	Pendukung	1	0	0
3	Keputusan	1	0	0
4	Prediksi	0	1	0
5	Hasil	0	1	0
6	Produksi	0	1	0
7	Analisis	0	0	1
8	Kualitas	0	0	1
9	Jaringan	0	0	1

Dari tabel diatas diperoleh nilai term SE=3, IS=3, dan Networking=3.

E-ISSN : 2827-9425
P-ISSN : 2828-4666

4. Hitung probabilitas prior setiap kategori. Yang menjadi kategori ada 3 kategori yaitu kelas SE,IS, dan Networking.

   P(SE) = $\frac{1}{3}$ = 0.333333333

   P(IS) = $\frac{1}{3}$ = 0.333333333

   P(Net) = $\frac{1}{3}$ = 0.333333333

5. Perhitungan Probabilitas *likelihood.* Pada data uji, kata yang termasuk dalam data training adalah kata "Analisis" dan "Jaringan". Sehingga hanya kata tersebut yang dihitung probabilitasnya.

   P(Analisis|SE) = $\frac{0+1}{3+11}$ = 0.0714285714

   P(Analisis|IS) = $\frac{0+1}{3+11}$ = 0.0714285714

   P(Analisis|Net) = $\frac{1+1}{3+11}$ = 0.1428571429

   P(Jaringan|SE) = $\frac{0+1}{3+11}$ = 0.0714285714

   P(Jaringan|IS) = $\frac{0+1}{3+11}$ = 0.0714285714

   P(Jaringan|Net) = $\frac{1+1}{3+11}$ = 0.1428571429

6. Menghitung probabilitas pada data *testing*.

   P(*Testing*|SE)

   = P(SE) * P(Analisis|SE) * P(Jaringan|SE)

   = 0.333333333 * 0.0714285714 * 0.0714285714

   = 0.0017006803

   P(*Testing*|IS)

   = P(IS) * P(Analisis|IS) * P(Jaringan|IS)

   = 0.333333333 * 0.0714285714 * 0.0714285714

   = 0.0017006803

   P(*Testing*|Net)

   = P(Net) * P(Analisis|Net) * P(Jaringan|Net)

   = 0.333333333 * 0.1428571429 * 0.1428571429

   = 0.0068027211

   Dari perhitungan diatas, nilai probabilitas tertinggi yaitu sebesar 0.0068027211 pada P(Uji|Net) sehingga judul skripsi tersebut diklasifikasikan ke dalam kelas "Networking".

7. KESIMPULAN

Berdasarkan hasil penelitian yang sudah diuraikan di atas tentang Klasifikasi judul skripsi dengan menggunakan metode *Naive Bayes Classifier*, maka dapat disimpulkan sebagai berikut:

E-ISSN : 2827-9425
P-ISSN : 2828-4666

1. Hasil dari penerapan metode *Naive Bayes Classifier* pada klasifikasi judul skripsi memperoleh hasil yang cukup baik sehingga judul skripsi dapat terbagi sesuai dengan konsentrasinya masing-masing. Karena, walaupun masih ada beberapa hasil klasifikasi judul skripsi yang belum sesuai dengan konsentrasinya, tetapi sudah lebih banyak hasil klasifikasi yang sesuai.
2. Penerapan metode *Naive Bayes Classifier* pada Klasifikasi Judul Skripsi dengan menggunakan 404 dataset didapatkan hasil akurasi sebesar 80%. Dengan demikian, metode *Naive Bayes Classifier* dapat digunakan untuk mengklasifikasikan judul skripsi berdasarkan konsentrasi.

DAFTAR PUSTAKA

[1] M. Suhendri and Y. Afrilia, "Klasifikasi Karya Ilmiah (Tugas Akhir) Mahasiswa Menggunakan Metode Naive Bayes Classifier (Nbc)." [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[2] "Evaluasi Ekstraksi Fitur Klasifikasi Teks Untuk Peningkatan Akurasi Klasifikasi Menggunakan Naive Bayes".

[3] U. Pujianto, T. Widiyaningtyas, D. D. Prasetya, and B. Romadhon, "Penerapan algoritma naïve bayes classifier untuk klasifikasi judul skripsi dan tugas akhir berdasarkan Kelompok Bidang Keahlian," 2017.

[4] M. Azhar Mujahid and M. Syafrullah, "Implementasi Algoritma Naïve Bayes Clasifier untuk Mengelompokkan Naskah Berita Pendidikan dan berita Covid-19," *KRESNA: Jurnal Riset dan Pengabdian Masyarakat*, vol. 1, no. 1, pp. 34–43, 2021, [Online]. Available: https://jurnaldrpm.budiluhur.ac.id/index.php/Kresna/

[5] E. I. Program, S. Sistem, I. A. Kampus, and K. Bogor, "Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes," vol. VII, no. 1, 2019.

[6] E. Dyar Wahyuni and A. Anjani Arifiyanti, "Klasifikasi Berita Pada Akun Twitter Suara Surabaya Menggunakan Metode Naive Bayes," 2020.

[7] R. Aditya Nugroho and I. Cholissodin, "Implementasi Naïve Bayes Classifier Untuk Klasifikasi Emosi Tweet Berbahasa Indonesia Pada Spark," 2021. [Online]. Available: http://j-ptiik.ub.ac.id

[8] W. P. Nurmayanti, "Penerapan Naive Bayes dalam Mengklasifikasikan Masyarakat Miskin di Desa Lepak," *Geodika: Jurnal Kajian Ilmu dan Pendidikan Geografi*, vol. 5, no. 1, pp. 123–132, Jun. 2021, doi: 10.29408/geodika.v5i1.3430.

[9] "Kreativitas Berfikir, Teknik Penulisan Dan Penguasaan Metodologi Penilitian : Analisis Terhadap Kualitas Skripsi Mahasiswa STAIN Jurai Siwo Metro".

[10] "Analisa Kepuasan Mahasiswa Terhadap E-Learning Menggunakan Algoritma Support Vector Machine".

[11] "Penerapan Text Mining Pada Sistem Klasifikasi Email Spam Menggunakan Naive Bayes".

[12] H. Naparin, "Klasifikasi Peminatan Siswa SMA Menggunakan Metode Naive Bayes," 2016.

[13] "Input Dan Output Pada Bahasa Pemrograman Python," 2018.