

**PENERAPAN ALGORITMA NAIVE BAYES
CLASSIFIER UNTUK KLASIFIKASI JUDUL
SKRIPSI BERDASARKAN KONSENTRASI**
(Studi Kasus : Fakultas Ilmu Komputer, Universitas Ichsan Gorontalo)

Oleh
SALMIN DANIA
T3119032

SKRIPSI

**Untuk Memenuhi salah satu syarat ujian
Guna memperoleh gelar Sarjana**



**PROGRAM SARJANA
TEKNIK INFORMATIKA
UNIVERSITAS ICHSAN GORONTALO
GORONTALO
2023**

PERSETUJUAN SKRIPSI

PENERAPAN ALGORITMA NAIVE BAYES CLASSIFIER UNTUK KLASIFIKASI JUDUL SKRIPSI BERDASARKAN KONSENTRASI

(Studi Kasus : Fakultas Ilmu Komputer, Universitas Ichsan Gorontalo)

Oleh
SALMIN DANIA
T3119032

SKRIPSI

Untuk memenuhi salah satu syarat ujian
guna untuk memenuhi gelar Sarjana
Program Studi Teknik Informatika,
Ini telah disetujui oleh Tim Pembimbing

Gorontalo, November 2023

Pembimbing I



Rezqiwati Ishak, M.Kom

NIDN: 0903087901

Pembimbing II



Hastuti Dalai, M.Kom

NIDN: 0918038803

PENGESAHAN SKRIPSI

**PENERAPAN ALGORITMA NAÏVE BAYES
CLASSIFIER UNTUK KLASIFIKASI JUDUL
SKRIPSI BERDASARKAN KONSENTRASI**
(Studi Kasus: Fakultas Ilmu Komputer, Universitas Ichsan Gorontalo)

Oleh

SALMIN DANIA

T3119032

Diperiksa oleh Panitia Ujian Strata Satu (S1)
Universitas Ichsan Gorontalo

1. Ketua Penguji
Zohrahayaty, M.Kom
2. Anggota
Irma Surya Kumala, M.Kom
3. Anggota
Kartika Chandra Pelangi, M.Kom
4. Anggota
Rezqiwati Ishak, M.Kom
5. Anggota
Hastuti Dalai, M.Kom

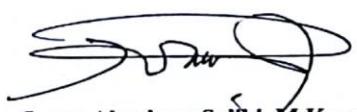


Handwritten signatures corresponding to the committee members listed above.

Mengetahui,

Dekan Fakultas Ilmu Komputer

Ketua Program Studi



Irvan Abraham Salih, M.Kom
NIDN: 0928028101



Sudirman S. Panna, M.Kom
NIDN: 0924038205

PERNYATAAN SKRIPSI

Dengan ini saya menyatakan bahwa:

1. Karya tulis (Skripsi) saya ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (Sarjana) baik di Universitas Ichsan Gorontalo maupun di perguruan tinggi lainnya.
2. Karya tulis (Skripsi) saya ini adalah murni gagasan, rumusan, dan penelitian saya sendiri, tanpa bantuan pihak lain, kecuali arahan dari Tim Pembimbing.
3. Dalam karya tulis (Skripsi) saya ini tidak terdapat karya atau pendapat yang telah dipublikasikan orang lain, kecuali secara tertulis dicantumkan sebagai acuan/sitasi dalam naskah dan dicantumkan pula dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya tulis ini, serta sanksi lainnya sesuai dengan norma-norma yang berlaku di Universitas Ichsan Gorontalo.

Gorontalo, November 2023

Yang Membuat Pernyataan,

Materai
6000

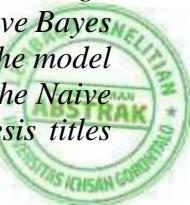
Salmin Dania

ABSTRACT

SALMIN DANIA. T3119032. THE APPLICATION OF THE NAIVE BAYES CLASSIFIER ALGORITHM FOR THE CLASSIFICATION OF THESIS TITLE BASED ON CONCENTRATION

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*



ABSTRAK

SALMIN DANIA. T3119032. PENERAPAN ALGORITMA NAIVE BAYES CLASSIFIER UNTUK KLASIFIKASI JUDUL SKRIPSI BERDASARKAN KONSENTRASI

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*



KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT, karena hanya dengan izin dan kuasa-Nya penulis dapat menyelesaikan penyusunan skripsi ini dengan judul **“Penerapan Algoritma Naive Bayes Classifier Untuk Klasifikasi Judul Skripsi Berdasarkan Konsentrasi”** Studi Kasus Fakultas Ilmu Komputer Universitas Ichsan Gorontalo. Skripsi ini dibuat untuk memenuhi salah satu syarat memperoleh gelar sarjana. Penulis menyadari dalam penyusunan skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, penulis mengharapkan masukan untuk kesempurnaan penulisan skripsi ini.

Penulis juga mengucapkan terima kasih kepada pihak-pihak yang telah membantu terselesainya skripsi ini :

1. Ibu Dr. Dra. Juriko Abdussamad, M.Si, selaku Ketua Yayasan Pengembangan Ilmu Pengetahuan dan Teknologi (YPIPT) Ichsan Gorontalo;
2. Bapak Dr. Abdul Gaffar La Tjokke, M.Si, selaku Rektor Universitas Ichsan Gorontalo;
3. Bapak Irvan A. Salihi, S.Kom, M.Kom, selaku Dekan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
4. Bapak Sudirman Melangi, M.Kom, selaku Wakil Dekan I Bidang Akademik Fakultas Ilmu Komputer;
5. Ibu Irma Surya Kumala, M.Kom, selaku Wakil Dekan II bidang Administrasi Umum dan KeuanganFakultas Ilmu Komputer;
6. Bapak Sudirman S. Panna, M.Kom, selaku Ketua Jurusan Teknik Informatika Fakultas ilmu Komputer Universitas Ichsan Gorontalo;
7. Ibu Rezqiwati Ishak, M.Kom, selaku Pembimbing I yang telah banyak membimbing penulis selama ini;
8. Ibu Hastuti Dalai, M.Kom, selaku Pembimbing II yang telah banyak membimbing penulis selama ini;
9. Bapak dan Ibu Dosen Universitas Ichsan Gorontalo yang sudah mendidik dan mengajarkan berbagai ilmu disiplin kepada penulis;

10. Kedua Orang Tua saya bapak Abdul Jamin Dania dan Ibu Haera Satali yang selalu mencerahkan kasih sayang, selalu memberikan dukungan baik secara moral maupun materiil, dan selalu mendoakan yang terbaik untuk saya selama ini ;
11. Keluarga yang selalu mendoakan dan mendukung demi kelancaran studi dan memberikan semangat untuk menyelesaikan pendidikan. Terlebih khusus Family Group yang selalu memberikan dukungan secara langsung maupun tidak langsung;
12. Sahabat Ana Fitriana dan Destira Laila yang sudah sangat banyak membantu penulis selama kuliah di Universitas Ichsan Gorontalo;
13. Semua pihak yang telah membantu dalam menyelesaikan pendidikan dan dalam menyelesaikan skripsi ini baik secara langsung maupun tidak langsung;
14. Untuk diri sendiri, terima kasih sudah bertahan dan berusaha sampai sejauh ini.

Semoga segala amal dan perbuatan yang telah diberikan mendapatkan rahmat dan balasan yang lebih baik dari Allah SWT. Akhir kata penulis berharap skripsi ini dapat berguna dan bermanfaat bagi semua. Amin.

Gorontalo, November 2023

Penulis

DAFTAR ISI

PERSETUJUAN SKRIPSI.....	i
PENGESAHAN SKRIPSI.....	ii
PERNYATAAN SKRIPSI.....	iii
ABSTRACT.....	iv
ABSTRAK	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL	xi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah.....	4
1.2 Rumusan Masalah.....	4
1.2 Tujuan Penelitian	4
1.2 Manfaat Penelitian	4
BAB II LANDASAN TEORI	6
2.1 Tinjauan Studi.....	5
2.2 Tinjauan Pustaka.....	7
2.2.1 Skripsi	7
2.2.2 Konsentrasi.....	7
2.2.3 <i>Text Mining</i>	8
2.2.4 Klasifikasi	8
2.2.5 <i>Naive Bayes Classifier</i>	9
2.2.6 Penerapan <i>Naive Bayes Classifier</i>	11
2.2.7 <i>Python</i>	15
2.2.8 <i>Confusion Matrix</i>	16
2.3 Perangkat Lunak	17
2.4 Kerangka Pikir	18
BAB III METODE PENELITIAN	19
3.1 Jenis, Metode, Subjek, Objek, Waktu dan Lokasi Penelitian	19
3.2 Pengumpulan Data.....	19

3.3 Pemodelan.....	20
3.3.1 Pra Pengolahan Data	20
3.3.2 Validasi	20
3.3.3 Pengembangan Model.....	21
3.3.4 Evaluasi Model.....	21
BAB IV HASIL PENELITIAN.....	22
4.1 Hasil Pengumpulan Data	22
4.2 Tahapan Naive Bayes	23
4.3 Hasil Pengembangan Model	27
4.3.1 Pengumpulan Data Latih.....	27
4.3.2 Preprocessing Data.....	29
4.3.3 Ekstraksi Fitur	31
4.3.4 Pelatihan Model Naive Bayes	33
4.3.5 Evaluasi Model.....	33
4.3.6 Klasifikasi Judul Skripsi	35
BAB V PEMBAHASAN PENELITIAN	39
5.1 Pembahasan Model	39
5.2 Penerapan Algoritma <i>Naive Bayes Classifier</i>	40
5.3 Pembahasan Hasil Pemodelan	43
BAB VI KESIMPULAN DAN SARAN.....	47
6.1 Kesimpulan	47
6.2 Saran	47
DAFTAR PUSTAKA	48
LAMPIRAN	

DAFTAR GAMBAR

Gambar 2.2 Kerangka Pikir.....	18
Gambar 3.1 Pengembangan Model Klasifikasi Judul Skripsi.....	21
Gambar 4.1 Evaluasi Model.....	27
Gambar 4.2 Hasil Kinerja Model Klasifikasi.....	35
Gambar 5.1 <i>Confusion Matrix</i>	39

DAFTAR TABEL

Tabel 1.1 Data Judul Skripsi Berdasarkan Konsentrasi	1
Tabel 2.1 Penelitian tentang Klasifikasi Menggunakan Metode <i>Naive Bayes</i>	5
Tabel 2.2 Tabel Data <i>Training</i>	11
Tabel 2.3 Tabel Data <i>Testing</i>	12
Tabel 2.4 Data Baru	13
Tabel 2.5 Tabel Proses Perhitungan	14
Tabel 2.6 Tabel Hasil.....	15
Tabel 2.7 Confusion Matrix.....	16
Tabel 2.8 Perangkat Lunak Pendukung.....	17
Tabel 3.1 Variabel/Atribut Data.....	19
Tabel 4.1 Hasil Pengumpulan Dataset.....	22
Tabel 4.2 Dataset Pemodelan	23
Tabel 4.3 <i>Lowercasing</i>	25
Tabel 4.4 Penghapusan Tanda Baca	25
Tabel 4.5 <i>Stopwords Removal</i>	25
Tabel 4.6 Hasil Preprocessing Data.....	26
Tabel 5.1 Hasil Akurasi	39
Tabel 5.2 Data <i>Training</i> Penerapan Algoritma	40
Tabel 5.3 Data <i>Testing</i> Penerapan Algoritma.....	40
Tabel 5.4 <i>Tokenizing</i>	41
Tabel 5.5 <i>Term Frekuensi</i>	41
Tabel 5.6 Data <i>Testing</i>	43
Tabel 5.7 Hasil Klasifikasi.....	44
Tabel 5.8 Hasil Klasifikasi Sesuai dan Belum Sesuai.....	45

BAB I

PENDAHULUAN

1.1 Latar Belakang

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].

Universitas Ichsan Gorontalo merupakan salah satu universitas yang ada di provinsi Gorontalo. Universitas tersebut memiliki 8 fakultas dan 13 program studi. Salah satunya adalah Fakultas Ilmu Komputer yang terdapat 3 jurusan yaitu Teknik Informatika, Sistem Informasi, dan Desain Komunikasi Visual. Banyaknya lulusan mahasiswa di Fakultas Ilmu Komputer dari tahun ke tahun tentunya berbanding lurus dengan banyaknya penelitian tugas akhir yang dihasilkan. Berikut adalah data judul skripsi berdasarkan konsentrasi:

Tabel 1.1 Data Judul Skripsi Berdasarkan Konsentrasi

No	Judul	Keterangan
1.	Sistem pengaman Pintu Menggunakan RFID Tak Berbasis <i>Arduino Uno</i> (Studi Kasus Universitas Ichsan Gorontalo)	Networking
2.	Prediksi Penjualan Air Mineral Boneva Menggunakan Metode <i>Least Square</i>	IS
3.	Implementasi Metode Regresi Linear Berganda Untuk Prediksi Harga Cabai Rawit di Kabupaten Gorontalo	IS
4.	Rancang Bangun Game Edukasi Membersihkan Sampah Berbasis Android	SE
5.	Pengenalan Kualitas Cabai Menggunakan Metode Artificial Neural Network	IS

No	Judul	Keterangan
6	Analisa dan Implementasi Metode Port Knocking Untuk Keamanan Jaringan Pada Router Mikrotik Universitas Ichsan Gorontalo	Networking
7	Rancang Bangun Sistem Penyiraman Bunga Berbasis Android Menggunakan Sensor Kelembapan Tanah dan Nodemcu	Networking
8	Prediksi Jumlah Orderan Nujek (Nusantara Ojek) Menggunakan Metode Simple Moving Average	IS
9	Sistem pengontrol Pintu Otomatis Ruangan Fakultas Ilmu Komputer Berbasis IOT	Networking
10	Prediksi Produksi Minyak Kelapa Sawit Menggunakan Metode <i>Exponential Moving Average</i>	IS
...
20	Analisa dan Perbandingan Kualitas Internet jaringan 4G Pada <i>Provider Selular</i> di Kota Gorontalo Menggunakan Standar Pengukuran QOS (<i>Quality Of Service</i>)	Networking

Sumber : Fakultas Ilmu Komputer

Skripsi atau karya ilmiah sebelumnya yang pernah diseminarkan dapat dijadikan referensi untuk dipelajari oleh mahasiswa sebelum memulai penelitian. Tetapi, hingga saat ini masih sulit bagi mahasiswa dalam mengetahui bidang penelitian dari judul penelitian atau skripsi karena belum terkласifikasi dengan baik. Oleh karena itu, peneliti melakukan klasifikasi terhadap judul skripsi sehingga informasi mengenai penelitian mahasiswa menjadi lebih terstruktur dan efisien.

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 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 Pujiyanto, Triyanna Widyaningtyas, 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 kedalam 5 kelas didapatkan nilai rata-rata akurasi yang cukup baik yaitu 86,68%.

Berdasarkan pemaparan diatas, peneliti bermaksud untuk membuat penelitian dengan judul **“Penerapan Algoritma Naive Bayes Classifier Untuk Klasifikasi Judul Skripsi Berdasarkan Konsentrasi”**.

1.2 Identifikasi Masalah

Berdasarkan uraian latar belakang diatas maka yang menjadi permasalahannya adalah bagaimana mengklasifikasikan data judul skripsi mahasiswa di fakultas teknik informatika menggunakan metode *Naive Bayes Classifier*.

1.3 Rumusan Masalah

Berdasarkan latar belakang masalah diatas, maka diperoleh beberapa rumusan masalah sebagai berikut :

1. Bagaimana hasil penerapan metode *Naive Bayes Classifier* untuk klasifikasi judul skripsi di Fakultas Ilmu Komputer, Universitas Ichsan Gorontalo?
2. Bagaimana tingkat akurasi penerapan metode *Naive Bayes Classifier* untuk klasifikasi judul skripsi di Fakultas Ilmu Komputer, Universitas Ichsan Gorontalo?

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, maka tujuan penelitian ini sebagai berikut :

1. Menerapkan metode *Naive Bayes Classifier* untuk klasifikasi judul skripsi di Fakultas Ilmu Komputer, Universitas Ichsan Gorontalo.
2. Mengetahui tingkat akurasi penerapan metode *Naive Bayes Classifier* untuk klasifikasi judul skripsi di Fakultas Ilmu Komputer, Universitas Ichsan Gorontalo.

1.5 Manfaat Penelitian

Penelitian ini semoga memberikan manfaat secara teoritis dan praktis kepada pihak terkait, sebagai masukan, pertimbangan, serta evaluasi.

1. Secara teoritis, penelitian ini diharapkan dapat mengembangkan ilmu pengetahuan sehingga dapat meningkatkan metode yang digunakan.
2. Secara praktis, penelitian ini diharapkan dapat mengembangkan ilmu pengetahuan bagi instansi sebagai bahan informasi agar dapat mengklasifikasikan judul skripsi berdasarkan konsentrasi.

BAB II

LANDASAN TEORI

2.1 Tinjauan Studi

Beberapa penelitian tentang klasifikasi dan penerapan metode *Naive bayes Classifier*, seperti di bawah ini:

Tabel 2.1 Penelitian tentang Klasifikasi Menggunakan Metode *Naive Bayes*.

No	Peneliti	Judul	Hasil
1.	Qonita, Eka Dyar Wahyuni, Amalia Anjani Arifiyanti (2020)[6].	Klasifikasi Berita Pada Akun Twitter Suara Surabaya Menggunakan Metode <i>Naive Bayes</i> .	Model yang dibuat dapat bekerja dengan baik. Dengan menggunakan algoritma multinomial <i>Naive Bayes</i> dengan penerapan SMOTE, didapatkan model dengan akurasi yang cukup tinggi yaitu 89%. Pengekstrasi informasi waktu dapat dilakukan dengan baik sedangkan untuk informasi lokasi belum menghasilkan hasil yang baik dikarenakan dataset tweet Suara Surabaya tidak menggunakan Bahasa Indonesia yang baik dan benar [6].
2.	Rizal Aditya Nugroho, Imam Cholissodin, Indriati (2021)[7].	Implementasi <i>Naive Bayes Classifier</i> untuk klasifikasi emosi tweet berbahasa Indonesia pada Spark	Akurasi dari metode <i>Naive Bayes Classifier</i> untuk klasifikasi emosi tweet berbahasa Indonesia pada Spark memiliki nilai rata-rata tertinggi sebesar 0,892 saat persentase 90% data latih dan 10% data uji. Lalu selanjutnya

Peneliti	Judul	Hasil
	Indonesia pada Spark.	nilai rata-rata tertinggi sebesar 0,880 saat menggunakan <i>smoothing</i> . Dan terakhir nilai rata-rata tertinggi sebesar 0,888 saat menggunakan prior konstan[7].
3. Wiwit Pura Nurmayanti dkk (2021)[8].	Penerapan <i>Naive Bayes</i> Dalam Mengklasifikasikan Masyarakat Miskin di Desa Lepak.	Berdasarkan hasil pengujian <i>confusion matrix</i> dengan menggunakan metode klasifikasi <i>Naive Bayes</i> terhadap dataset yang telah diambil pada objek penelitian diperoleh klasifikasi masyarakat miskin di desa Lepak yang memang miskin adalah 148 record dari 156 record yang artinya terdapat 8 record yang error, dimana ia lebih mirip dengan yang tidak miskin. Sedangkan untuk klasifikasi masyarakat tidak miskin terdapat 110 record dari 111 record yang memang tidak miskin dan sisanya 1 record yang error, dimana ia lebih mirip yang miskin. Nilai akurasi yang diperoleh adalah 96.63%[8].

2.2 Tinjauan Pustaka

2.2.1 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].

Menurut Machdar Helmi, skripsi merupakan karya tulis ilmiah berdasarkan hasil penelitian lapangan dan atau studi kepustakaan yang disusun mahasiswa sesuai dengan bidang studinya sebagai tugas akhir dalam studi formalnya di perguruan tinggi. Skripsi dalam dunia pendidikan berarti suatu hasil penyusunan tulisan ilmiah yang telah dibuktikan kebenarannya berdasarkan data-data yang telah dikumpulkan dan tentunya data yang dikumpulkan diolah untuk kemudian menjadi data yang valid sebagai bahan acuan untuk membuktikan kebenaran suatu tulisan tersebut[9].

2.2.2 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.

Di jurusan Teknik Informatika yang terdapat pada fakultas Ilmu Komputer di Universitas Ichsan Gorontalo terdapat 3 konsentrasi yaitu, *Intelligent System* (IS), *Software Engineering* (SE), dan *Networking*. Mahasiswa bebas memilih konsentrasi yang diinginkan sesuai dengan minat dan bakat yang dimiliki. Penentuan konsentrasi mahasiswa tidak mudah. Hal ini disebabkan akan berpengaruh terhadap kegiatan akademik mahasiswa.

2.2.3 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 retrieval*. Secara umum *text mining* terdiri dari tiga langkah yaitu: teks preprocessing, operasi penggalian teks, postprocessing. Tugas teks preprocessing adalah termasuk didalamnya pemilihan data, klasifikasi dan ekstraksi fitur untuk mengubah dokumen menjadi bentuk perantara, yang harus cocok dengan tujuan pencarian yang berbeda. Bagian utama dari pekerjaan operasi *text mining* mencakup pengelompokan, penemuan aturan asosiasi, tren analisis, pola penemuan, serta algoritma penemuan pengetahuan. Pekerjaan selanjutnya sebelum pemrosesan memanipulasi data atau informasi terbaru yang berasal dari proses *text mining*, seperti evaluasi dan pemilihan informasi yang ditemukan, interpretasi dan visualisasi informasi yang dihasilkan. Sampai saat ini, sudah banyak model atau pola penggalian *text mining* yang telah digunakan.

2.2.4 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*. Proses klasifikasi biasanya dibagi menjadi dua fase: *learning* dan *test*. Pada *vase learning*, sebagian data yang telah diketahui kelas datanya diumpamakan untuk membentuk model prediksi. Karena menggunakan data yang telah diberikan label terlebih dahulu sebagai contoh data yang benar maka klasifikasi sering disebut juga sebagai metode diawasi (*supervised method*). Kemudian pada fase *testing*, model prediksi yang sudah terbentuk

diuji dengan sebagian data lainnya untuk mengetahui akurasi dari model tersebut. Bila akurasinya mencukupi model ini dapat dipakai untuk prediksi kelas data yang belum diketahui[11].

2.2.5 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].

Dalam menggunakan metode *Naive Bayes*, keuntungan yang bisa didapat adalah metode ini hanya membutuhkan jumlah data training yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Karena diasumsikan sebagai variabel independen, maka hanya varian dari suatu variabel dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians (Santoso, 2017). *Naive Bayes* merupakan teknik prediksi berbasis probabilitas sederhana yang berdasarkan pada penerapan teorema Bayes dengan asumsi independensi yang kuat (Saleh, 2015). *Naive Bayes* didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output (Santoso, 2017). Pengklasifikasian *Naive Bayes* dilakukan dengan memilih probabilitas akhir (posterior) tertinggi dari masing-masing kelas (Simatupang, Wuryandari, & Suparti, 2016)[8].

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*)

$P(X|H)$: Probabilitas berdasarkan kondisi pada hipotesis

$P(X)$: Probabilitas

Agar menjelaskan teorema *Naive Bayes* perlu kita ketahui bahwa proses dari klasifikasi membutuhkan sejumlah petunjuk untuk menentukan kelas apa yang cocok untuk sampel yang akan dianalisis tersebut[12]. Karena itu teorema bayes diatas disesuaikan sebagai berikut :

$$P(C|F_1, \dots, F_n) \frac{P(C)P(F_1, \dots, F_n)|C}{P(F_1, \dots, F_n)}$$

Yang dimana suatu variabel C mempresentasikan kelas, sementara variabel $F_1 \dots F_n$ mempresentasikan karakteristik yang digunakan untuk melakukan klasifikasi. Sehingga rumus tersebut menjelaskan peluang masuknya sampel karakteristik tertentu dalam kelas C (posterior) merupakan peluang munculnya kelas C (sebelum masuknya sampel tersebut, sehingga disebut prior), dikali dengan peluang kemunculan karakteristik sampel secara global (disebut juga *evidence*)[12]. Oleh karena itu, rumus diatas dapat pula ditulis secara sederhana sebagai berikut :

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

Nilai dari *evidence* akan selalu tetap untuk setiap kelas pada suatu sampel. Nilai dari suatu posterior tersebut nantinya akan dibandingkan dengan nilai-nilai posterior kelas lainnya agar menentukan kelas apa suatu sampel akan diklasifikasikan[12].

2.2.6 Penerapan Naive Bayes Classifier

Berikut ini adalah contoh penerapan algoritma *Naive Bayes* yang diambil dari penelitian yang dilakukan oleh Muhammad Azhar Mujahid, Windarto, dan Mohammad Syafrullah dengan judul “Implementasi Algoritma *Naive Bayes Classifier* untuk Mengelompokkan Naskah Berita Pendidikan dan Berita Covid-19”.

1. Data Training

Penelitian ini menggunakan data berupa berita khusus berbahasa Indonesia yang terdapat dalam *Google Alerts* dengan topik Kesehatan dan Pendidikan. Data berita yang diambil berjumlah 295 data dan mengambil 236 total data sampel dengan pembagian 80:20. Penulis membuat database *MySQL* yang bernama kkp dan untuk menyimpan data tersebut, yang nantinya akan dipakai untuk proses *Stemming*, *Case Folding*, *Remove Punctuation* dan *Tokenizing*[4]. Dibawah ini adalah tabel contoh data *training*.

Tabel 2.2 Tabel Data *Training*

Sumber	Sumber	Text	Kategori	Kata Kunci
Google Alerts	Liputan6.com	Vaksin covid telah tersedia pada bulan Januari 2020	Kesehatan	Vaksin Covid
Google Alerts	Tribunnews.com	Saat ini seluruh siswa dan siswi se jabodetabek akan melakukan daring atau pembelajaran tatap muka pada awal December	Pendidikan	Pembelajaran Tatap Muka
Google Alerts	Kompas.com	Debat dua qosim alif terpapar visi misi dan program bangun lanjut di seluruh pelosok untuk pendidikan	Pendidikan	Pembelajaran Tatap Muka

2. Data Testing

Penelitian ini menggunakan data berupa berita khusus berbahasa Indonesia yang terdapat dalam *Google Alerts* dengan topik Kesehatan dan Pendidikan. Data berita yang diambil berjumlah 295 data dan mengambil 59 total data testing dengan pembagian 80:20. Penulis membuat database *MySQL* yang bernama kkp dan untuk menyimpan data tersebut, yang nantinya akan dipakai untuk proses *Stemming*, *Case Folding*, *Remove Punctuation* dan *Tokenizing*[4]. Dibawah ini adalah tabel contoh data *testing*.

Tabel 2.3 Tabel Data Testing

Sumber	Sumber	Text	Kategori	Nilai Probabilitas
Google Alerts	Liputan6.com	Vaksin covid percobaan di kabupaten penajam paser utara	Kesehatan	Kesehatan : 0.9898 Pendidikan : 0.5232
Google Alerts	Tribunnews.com	Sekolah sejabodetabek hampir sebulan, siswa sdn karyabakti tetap semangat belajar meski belajar tatap muka	Pendidikan	Kesehatan : 0.3203 Pendidikan : 0.8994
Google Alerts	Kompas.com	Debat dua qosim alif terpapar visi misi dan program bangun lanjut di seluruh pelosok untuk pendidikan	Pendidikan	Kesehatan : 0.5349 Pendidikan : 0.5232

Proses Data *Training* dilakukan secara otomatis untuk mengetahui seberapa akurat sistem dalam mengklasifikasikan objek tertentu. Proses tersebut dilakukan dengan cara sistem membaca data baru yang digunakan sebagai data *training* yang terdapat pada tabel lalu memprediksi pola kata dari model (*data training*) dan menghitung nilai probabilitas dari setiap kata yang ada pada berita.

Tabel 2.4 Data Baru

Text	Kategori
Catat ini dating jadwal vaksin corona di RI	Kesehatan
Kiat ajar tatap muka bupati Haryanto saya tidak mau murid jadi korban covid	Pendidikan

Pada tabel 4 terdapat data baru yang akan digunakan sebagai data testing untuk kemudian dilakukan perhitungan nilai probabilitas menggunakan algoritma *naive bayes classifier*. Berikut perhitungan nilai probabilitasnya. Pertama mencari probabilitas setiap kategori menggunakan persamaan pada rumus berikut ini.

$$P(Kesehatan) = \frac{1}{2} = 0.5$$

$$P(Pendidikan) = \frac{1}{2} = 0.5$$

1. P adalah probabilitas, Vj adalah kategori
2. docs adalah masing-masing dari kategori
3. training adalah keseluruhan dari kategori

Text Uji
Catat, bupati

Mencari nilai probabilitas setiap kata uji pada kesehatan menggunakan persamaan pada rumus yaitu sebagai berikut :

$$P(catat|pendidikan) = \frac{0 + 1}{14 + 22} = 0,02777$$

$$P(canggih|pendidikan) = \frac{1 + 1}{14 + 22} = 0,05555$$

Lalu setelah mengetahui hasil probabilitas pada setiap kata uji selanjutnya melakukan proses perhitungan pada tabel 6 yaitu sebagai berikut :

Tabel 2.5 Tabel Proses Perhitungan

Kategori	$P(vj) = docs\ j / training$	$P(ai vj)$ $= \frac{ ni + 1 }{ n + kosakata }$	$P(ai vj) / P(vj)$	$VMAP$ $= \frac{\text{argmax}}{vj \in V}$ $P(vj) \prod i P(ai Vj)$
kesehatan	0.5	P(catat kesehatan) = 0,09090 P(canggih kesehatan) = 0,04545	0,09090 * 0,04545 * 0,5 = 0,002065 7025	0,002065702 5
pendidikan	0.5	P(catat kesehatan) = 0,02777 P(canggih kesehatan) = 0,05555	0,02777 * 0,05555 * 0,5 = 0,000771 311	0,000771311 75

Pada tabel 6 proses perhitungan pada *Naive Bayes* bahwa probabilitas yang mempunyai nilai terbesar adalah kategori pendidikan dengan nilai probabilitas 0,00077131175.

3. Hasil Pengujian

Pada pengujian lebih lanjut untuk data testing, data yang digunakan yaitu sebanyak 295 data *training* dan 59 untuk data testing, kemudian data yang telah di *training* sebelumnya dijadikan acuan untuk mengukur akurasi pada data testing ini dan menghasilkan akurasi sebesar 74.58% dimana hasilnya dapat dilihat pada tabel 7 berikut[4].

Tabel 2.6 Tabel Hasil

Accuracy : 74.58%			
	True Kesehatan	True Pendidikan	Class Precision
Kesehatan	38	15	71.70%
Pendidikan	0	6	100.00%
Class Recall	100.00%	28.57%	

Pada hasil pengujian untuk data testing dan data *training* yang digunakan yaitu sebanyak 295 data, kemudian data yang telah di *training* sebelumnya dijadikan acuan untuk mengukur akurasi pada data testing ini dan menghasilkan akurasi sebesar 74.58% yang menghasilkan perhitungan *precision* dengan berita kesehatan memiliki nilai *precision* 71.70%, berita pendidikan memiliki nilai *precision* 100% dan hasil dari nilai *recall* berita kesehatan memiliki nilai *recall* sebesar 100%, berita pendidikan memiliki nilai *recall* sebesar 28.57%[4].

2.2.7 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 *komprehensif*. *Python* juga didukung oleh komunitas yang besar[13].

Python mendukung multi paradigma pemrograman, utamanya namun tidak dibatasi pada pemrograman berorientasi objek, pemrograman imperatif, dan pemrograman fungsional. Salah satu *fitur* yang tersedia pada *python* adalah sebagai bahasa pemrograman dinamis yang dilengkapi dengan manajemen memori otomatis. Seperti halnya pada bahasa pemrograman dinamis lainnya, *python* umumnya digunakan sebagai bahasa *script* meski pada praktiknya penggunaan bahasa ini lebih luas mencakup konteks pemanfaatan yang umumnya tidak dilakukan dengan menggunakan bahasa *script*. *Python* dapat digunakan untuk berbagai keperluan pengembangan perangkat lunak dan dapat berjalan di berbagai *platform* sistem operasi[13].

2.2.8 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	1	TP	FN
	0	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 1 yang salah diklasifikasikan sebagai kelas 0 Rumus confusion matrix untuk menghitung accuracy, precision, dan recall seperti berikut.

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

$$Accuracy = \frac{TP + TN}{\text{Total}}$$

$$Precision = \frac{TP}{TP + FP}$$

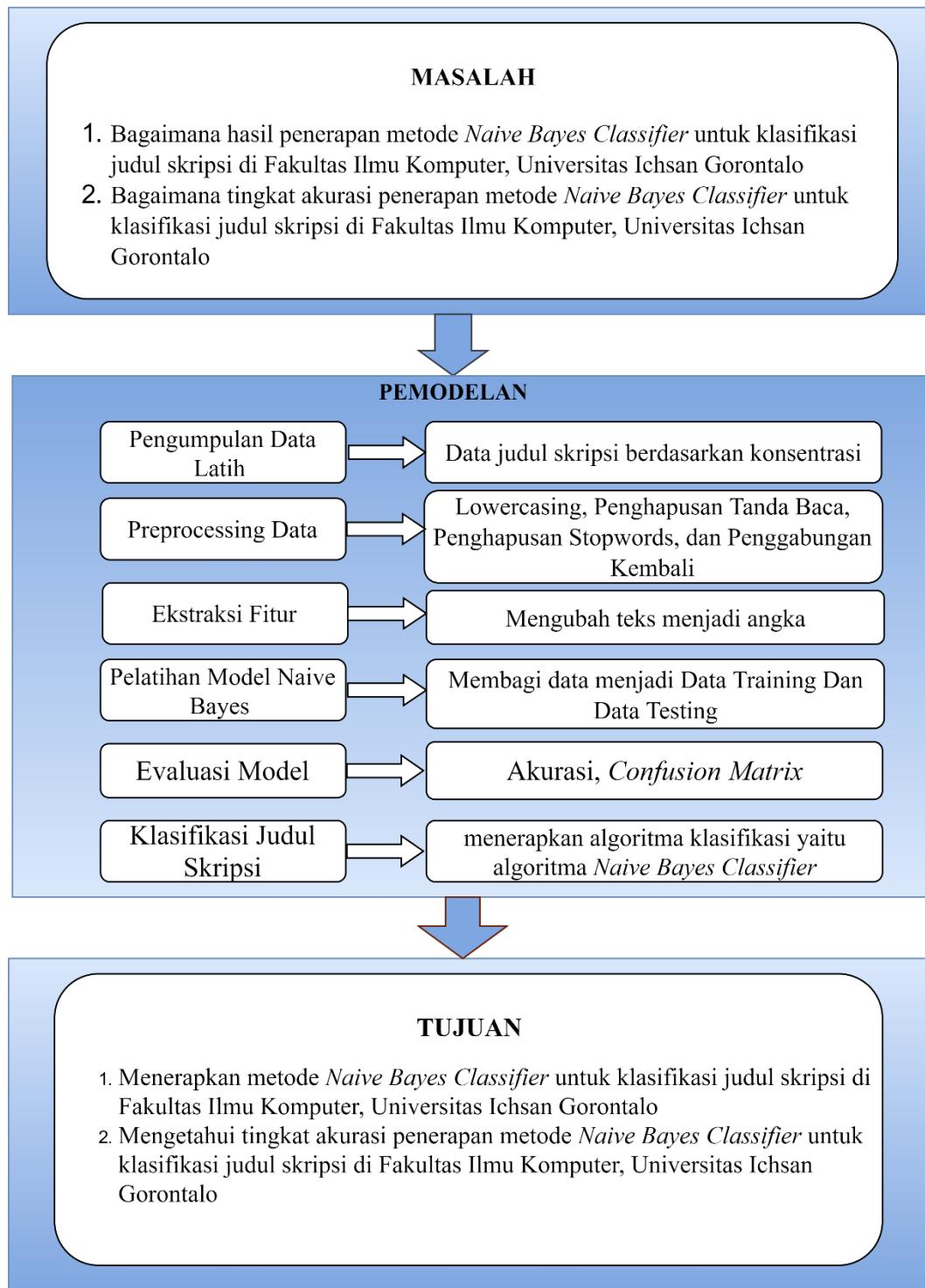
$$Recall = \frac{TP}{TP + FN}$$

2.3 Perangkat Lunak

Perangkat lunak yang digunakan dalam mengklasifikasi seperti pada tabel dibawah ini :

No	Perangkat Lunak	Keterangan
1.	<i>Python</i>	Python adalah bahasa pemograman yang digunakan untuk membuat aplikasi, perintah komputer, dan melakukan analisa data. Selain itu juga bisa digunakan untuk membuat program apa saja dan menyelesaikan berbagai masalah.
2.	<i>Google Colab</i>	Untuk digunakan menjalankan kode python tanpa perlu melakukan proses instalasi.

2.4 Kerangka Pikir



Gambar 2.2 Kerangka Pikir

BAB III

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. Adapun variabel/atribut dan tipe datanya masing-masing yang ditujukan dari tabel berikut:

Tabel 3.1. Variabel/Atribut Data

No.	Atribut	Jenis Data	Keterangan
1.	Judul skripsi	Varchar	<i>Input</i>
2.	Konsentrasi	Varchar	<i>Input</i>
3.	Hasil Klasifikasi	Integer	<i>Output</i>

3.3 Pemodelan

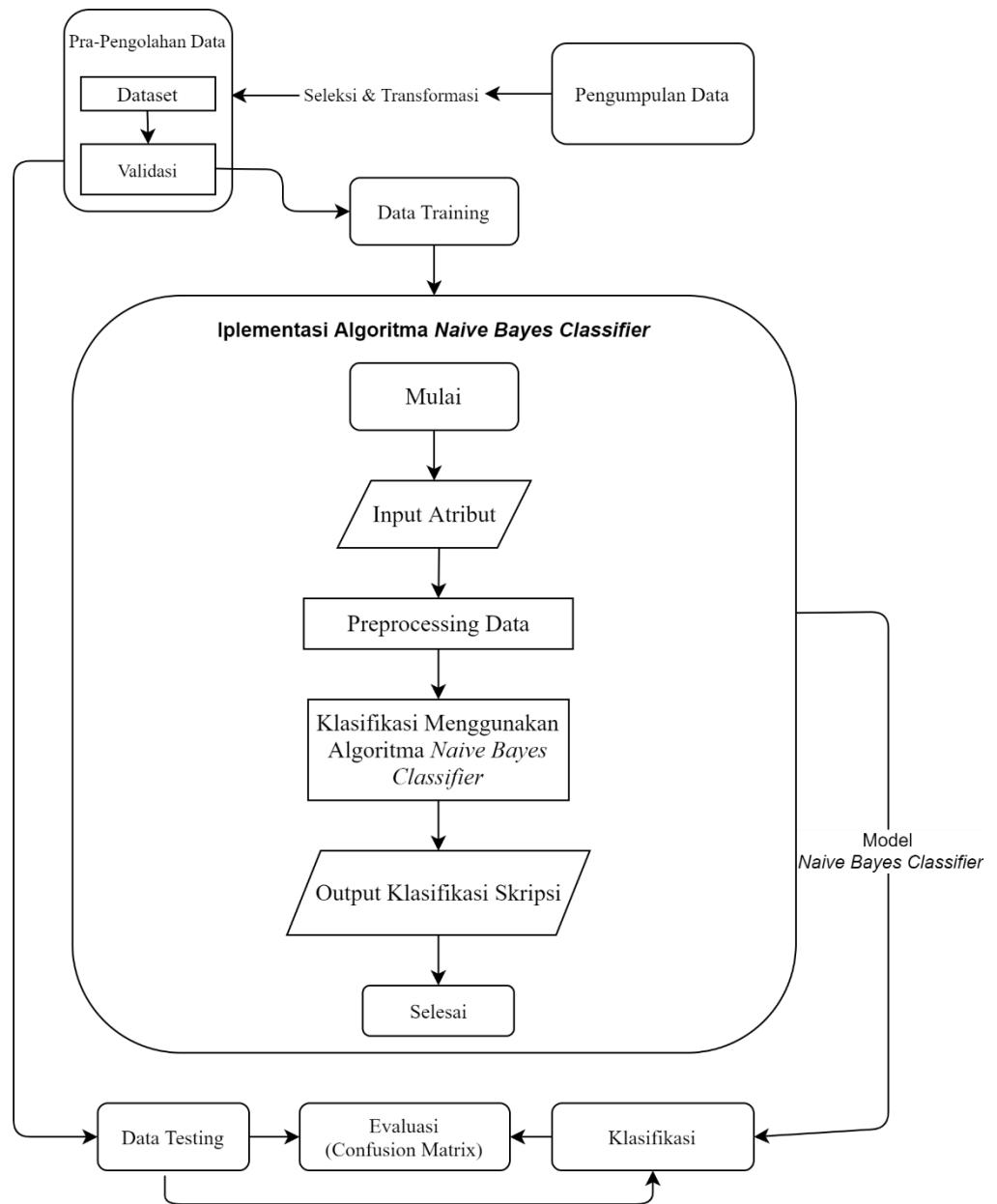
3.3.1 Pra Pengolahan Data

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.

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.

3.3.3 Pengembangan Model



Gambar 3.1 Pengembangan Model Klasifikasi Judul Skripsi

3.3.4 Evaluasi Model

Model yang telah dihasilkan kemudian dievaluasi dengan menggunakan *Confusion Matrix* untuk mengetahui akurasi.

BAB IV

HASIL PENELITIAN

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 PREFERENCE 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
3	T3117326	Sintia Hunowu	SISTEM PENDUKUNG KEPUTUSAN PENETAPAN PENERIMA BANTUAN PANGAN NON TUNAI (BPNT) MENGGUNAKAN METODE PREFERENCE SELECTION INDEX (PSI)	SE
4	T3115174	Sunardi R Ibrahim	KLASIFIKASI KELAYAKAN PEMBERIAN PINJAMAN MENGGUNAKAN METODE SUPPORT VEKTOR MANCHINE (STUDI KASUS KOPERASI MERTHA JAYA BUOL)	SE
5	T3117332	Sintia Lahay	SISTEM PENDUKUNG KEPUTUSAN PENERIMA BEASISWA PRESTASI DAERAH DENGAN MENERAPKAN METODE COMPOSITE PERFORMANCE INDEX (CPI)	SE
6	T3117306	Reska Ngabito	Sistem pendukung keputusan penetapan penerima bantuan rumah rehab menggunakan metode fuzzy mamdani	SE
7	T3117351	Rivandi Hulopi	sistem pendukung keputusan penetapan penerima bantuan program keluarga harapan(pkh) menggunakan metode moora	SE
8	T3117368	Merin Nurlaisa Abas	sistem pendukung keputusan penerima bantuan rumah rehab dengan	SE

No	NIM	Nama Lengkap	Judul Skripsi	Ket
			menerapkan metode composite performance index di desa limbula	
9	T3117254	Yasin K. Adam	penerapan Metode Single Moving Average Untuk Memprediksi Penjualan Air Minum Aqua Berdasarkan Jenis Kemasan	IS
10	T3117116	Ratna Kasim	Prediksi hasil produksi jagung menggunakan metode eksponensial moving average	IS
...
404	T3115214	Abdul Kadir Arsad	ANALISIS KUALITAS JARINGAN BACKBONE NIRKABEL DI BLOK PLAN PERKANTORAN MENGGUNAKAN STANDAR QUALITY OF SERVICE (QOS)	Networking

4.2 Tahapan *Naive Bayes Classifier*

Berdasarkan dataset pada tabel 4.1 di atas dilakukan beberapa pengolahan data sebelum diterapkan pada model *Naive Bayes Classifier* diantaranya dilakukan tahapan pengolahan sebagai berikut :

1. Pengumpulan *Data Training*

Hasil pengumpulan dataset pada tabel 4.1 di atas terdapat 4 atribut yaitu NIM, Nama, Judul Skripsi, dan Konsentrasi. Terdapat atribut yang tidak diperlukan dalam pemodelan yaitu NIM dan Nama, sehingga perlu dilakukan pemilihan atribut target yang sesuai dengan model yang akan digunakan. Adapun hasil pemilihan atribut target yang digunakan dapat dilihat pada tabel 4.2 berikut :

Tabel 4.2 Dataset Pemodelan

No	Judul Skripsi	Ket
1	PENERAPAN METODE PREFERENCE 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

No	Judul Skripsi	Ket
3	SISTEM PENDUKUNG KEPUTUSAN PENETAPAN PENERIMA BANTUAN PANGAN NON TUNAI (BPNT) MENGGUNAKAN METODE PREFERENCE SELECTION INDEX (PSI)	SE
4	KLASIFIKASI KELAYAKAN PEMBERIAN PINJAMAN MENGGUNAKAN METODE SUPPORT VEKTOR MANCHINE (STUDI KASUS KOPERASI MERTHA JAYA BUOL)	SE
5	SISTEM PENDUKUNG KEPUTUSAN PENERIMA BEASISWA PRESTASI DAERAH DENGAN MENERAPKAN METODE COMPOSITE PERFORMANCE INDEX (CPI)	SE
6	Sistem pendukung keputusan penetapan penerima bantuan rumah rehab menggunakan metode fuzzy mamdani	SE
7	sistem pendukung keputusan penetapan penerima bantuan program keluarga harapan(pkh) menggunakan metode moora	SE
8	sistem pendukung keputusan penerima bantuan rumah rehab dengan menerapkan metode composite performance index di desa limbula	SE
9	penerapan Metode Single Moving Average Untuk Memprediksi Penjualan Air Minum Aqua Berdasarkan Jenis Kemasan	IS
10	Prediksi hasil produksi jagung menggunakan metode eksponensial moving average	IS
...
404	ANALISIS KUALITAS JARINGAN BACKBONE NIRKABEL DI BLOK PLAN PERKANTORAN MENGGUNAKAN STANDAR QUALITY OF SERVICE (QOS)	Networking

2. *Preprocessing* Data

Pada tahapan *preprocessing* data dilakukan beberapa tahapan sebagai berikut :

1. Lowercasing

Pada tahapan ini dilakukan konversi huruf menjadi huruf kecil.

Tabel 4.3 Lowercasing

Input	Output
SISTEM PENDUKUNG KEPUTUSAN PENETAPAN PENERIMA BANTUAN PANGAN NON TUNAI (BPNT) MENGGUNAKAN METODE PREFERENCE SELECTION INDEX (PSI)	sistem pendukung keputusan penetapan penerima bantuan pangan non tunai (bpnt) menggunakan metode preference selection index (psi)

2. Menghapus Tanda Baca dan Karakter Khusus

Pada tahapan ini dilakukan penghapusan tanda baca dan karakter khusus.

Tabel 4.4 Penghapusan Tanda Baca

Input	Output
sistem pendukung keputusan penetapan penerima bantuan pangan non tunai (bpnt) menggunakan metode preference selection index (psi)	sistem pendukung keputusan penetapan penerima bantuan pangan non tunai menggunakan bpnt menggunakan metode preference selection index psi

3. Stopwords Removal

Pada tahapan ini dilakukan penghapusan kata-kata umum.

Tabel 4.5 Stopwords Removal

Input	Output
sistem pendukung keputusan penetapan penerima bantuan pangan non tunai bpnt menggunakan metode preference selection index psi	sistem pendukung keputusan penetapan penerima bantuan pangan non tunai bpnt metode preference selection index psi

Hasil dari tahap preprocessing data dapat dilihat pada **Tabel 4.6** berikut:

Tabel 4.6 Hasil Preprocessing Data

No	Judul Skripsi	Ket
1	penerapan metode preference selection index sistem pendukung keputusan penilaian kinerja aparatur desa	SE
2	sistem pendukung keputusan penerima bantuan rumah rehab desa botubilotahu metode vikor	SE
3	sistem pendukung keputusan penetapan penerima bantuan pangan non tunai bpnt metode preference selection index psi	SE
4	klasifikasi kelayakan pemberian pinjaman metode support vektor manchine studi koperasi mertha jaya buol	SE
5	sistem pendukung keputusan penerima beasiswa prestasi daerah menerapkan metode composite performance index cpi	SE
6	sistem pendukung keputusan penetapan penerima bantuan rumah rehab metode fuzzy mamdani	SE
7	sistem pendukung keputusan penetapan penerima bantuan program keluarga harapanpkh metode moora	SE
8	sistem pendukung keputusan penerima bantuan rumah rehab menerapkan metode composite performance index desa limbula	SE
9	penerapan metode single moving average memprediksi penjualan air minum aqua berdasarkan jenis kemasan	IS
10	prediksi hasil produksi jagung metode eksponensial moving average	IS
...
404	analisis kualitas jaringan backbone nirkabel blok plan perkantoran standar quality of service qos	Networking

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
(0, 92)	0.24708722209611186
(0, 591)	0.24708722209611186
(0, 151)	0.39150408081294574
(0, 578)	0.3306432468806023
(0, 359)	0.20578826501509506
(0, 586)	0.20578826501509506

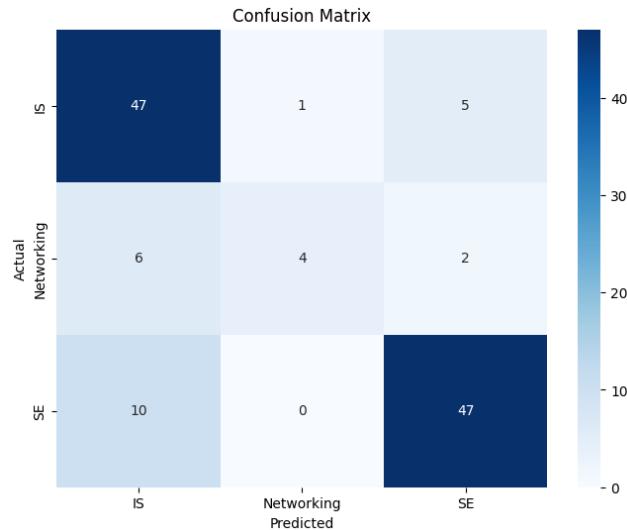
Kolom pertama menunjukkan nomor indeks, kolom kedua menunjukkan nomor indeks pada dokumen, dan kolom ketiga menunjukkan bobot yang diberikan kepada kata-kata pada indeks.

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. Evaluasi kinerja model *Naive Bayes* menggunakan metode *Confusion Matrix* dengan hasil perhitungan dapat dilihat pada **Gambar 4.1** berikut :



Gambar 4.1 Evaluasi model

6. Klasifikasi Judul Skripsi

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

4.3 Hasil Pengembangan Model

Pengembangan Model *Naive Bayes Classifier* menggunakan bahasa pemrograman *Python* dengan *tools Google Colaboratory* dengan tahapan-tahapan berikut :

4.3.1 Pengumpulan *Data Training*

Pada tahapan pengumpulan *data training* ini dilakukan beberapa proses sebagai berikut :

1. Mengaktifkan penggunaan *Google Drive* sebagai tempat penyimpanan, adapun coding dengan *Python* sebagai berikut:

```
#Proses Pengaitan dengan Google Drive
from google.colab import drive
drive.mount('/content/drive')
```

2. Membaca Dataset file excel :

```
#Membaca Dataset
df = pd.read_excel("/content/drive/MyDrive/Skripsi/filejudul.xlsx")
df.head()
```

Hasil outputnya sebagai berikut :

No.	Judul Skripsi	Keterangan
0	1 Sistem Pendukung Keputusan pemilihan Calon Pen...	SE
1	2 Prediksi Persediaan Jumlah Benih Padi Mengguna...	IS
2	3 Penerapan Algoritma Naive Bayes Pada Analisis ...	SE
3	4 Prediksi Penjualan Minyak kelapa Kampung Mengg...	IS
4	5 Prediksi Tingkat Produksi Kopi Pinogu Mengguna...	IS

3. Konversi Dataset Excel menjadi CSV

Dataset dalam format excel perlu dikonversi kebentuk CSV (*COMMA Separated Values*) agar bisa digunakan dalam pemodelan data dengan menggunakan perintah sebagai berikut:

```
#Mengkonversi File Excel .xlsx ke .csv
df.to_csv("/content/drive/MyDrive/Skripsi/filejudul.csv", index=False)
df.info()
```

Setelah dijalankan perintah diatas maka didapatkan hasil sebagai berikut:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 404 entries, 0 to 403
Data columns (total 3 columns):
 #   Column           Non-Null Count  Dtype  
---  --  
 0   No.              404 non-null    int64  
 1   Judul Skripsi    404 non-null    object  
 2   Keterangan       404 non-null    object  
dtypes: int64(1), object(2)
memory usage: 9.6+ KB
```

4.3.2 Preprocessing Data

Pada tahapan *preprocessing* data ini dilakukan beberapa proses sebagai berikut:

1. Mengimpor *library* kedalam *Python* untuk mengunduh *stopwords*

```
import nltk
nltk.download('stopwords')
```

Setelah dijalankan perintah di atas maka didapatkan hasil sebagai berikut:

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]  Unzipping corpora/stopwords.zip.
True
```

2. Melakukan pemrosesan teks dengan menggunakan beberapa *library*.

```
import pandas as pd
import re
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from sklearn.feature_extraction.text import TfidfVectorizer

# Fungsi untuk melakukan preprocessing pada teks
def preprocess_text(text):
    # Mengonversi huruf menjadi huruf kecil
    text = text.lower()

    # Menghapus tanda baca dan karakter khusus
    text = re.sub(r'[^a-zA-Z0-9\s]', '', text)

    # Menghapus kata-kata umum (stop words)
    stop_words = set(stopwords.words('indonesian')) # Menggunakan daftar kata-kata umum bahasa Indonesia
    words = text.split()
    words = [word for word in words if word not in stop_words]

    # Menggabungkan kata-kata kembali menjadi teks
    text = ' '.join(words)

    return text
```

Kemudian dilakukan pemanggilan fungsi *preprocessing* data.

```
# Memanggil Fungsi Preprocessing Data
df['Judul Skripsi'] = df['Judul Skripsi'].apply(preprocess_text)
df.head()
```

Setelah dijalankan maka didapatkan hasil sebagai berikut :

No.	Judul Skripsi	Keterangan
0 1	sistem pendukung keputusan pemilihan calon pen...	SE
1 2	prediksi persediaan benih padi algoritma regre...	IS
2 3	penerapan algoritma naive bayes analisis senti...	SE
3 4	prediksi penjualan minyak kelapa kampung metod...	IS
4 5	prediksi tingkat produksi kopi pinogu metode l...	IS

Setelah itu, dilakukan penghitungan jumlah setiap kategori judul dalam kolom keterangan.

```
# Menghitung jumlah setiap kategori Judul dalam kolom 'Keterangan'
df['Keterangan'].value_counts()
```

Setelah dijalankan maka didapatkan hasil sebagai berikut :

```
SE           196
IS           171
Networking    37
Name: Keterangan, dtype: int64
```

3. Selanjutnya dilakukan pengunduhan dan penginstalan modul *Python* yang disebut Sastrawi. Sastrawi adalah pustaka pemrosesan bahasa alami untuk bahasa Indonesia.

```
!pip install Sastrawi
```

Setelah dijalankan maka didapatkan hasil sebagai berikut:

```
Collecting Sastrawi
  Downloading Sastrawi-1.0.1-py2.py3-none-any.whl (209 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 209.7/209.7 kB 3.2 MB/s eta 0:00:00
Installing collected packages: Sastrawi
Successfully installed Sastrawi-1.0.1
```

4. Kemudian dilakukan beberapa tahap pemrosesan teks dalam bahasa Indonesia.

```
import pandas as pd
import re
from nltk.corpus import stopwords
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from sklearn.feature_extraction.text import TfidfVectorizer
```

Setelah itu, mengimpor modul untuk mengakses alat yang diperlukan untuk memproses data teks, melatih model klasifikasi, menguji model tersebut, dan mengevaluasi kinerjanya.

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, classification_report
```

5. Ekstraksi Fitur

Pada tahapan ini, dilakukan proses pengubahan teks judul skripsi menjadi representasi angka agar dapat digunakan dalam algoritma pemodelan.

```
# Ekstraksi Fitur
# mengubah teks judul skripsi menjadi representasi angka
vectorizer = TfidfVectorizer()
X = vectorizer.fit_transform(df['Judul Skripsi'])
y = df['Keterangan']
```

Kemudian dilakukan pencetakan variabel x.

```
print(X)
```

Setelah dijalankan maka didapatkan hasil sebagai berikut :

```
(0, 434)      0.38292243566428324
(0, 478)      0.11035805047823478
(0, 861)      0.38292243566428324
(0, 406)      0.43937947843996056
(0, 92)       0.24708722209611186
(0, 591)      0.24708722209611186
(0, 151)      0.39150408081294574
(0, 578)      0.3306432468806023
(0, 359)      0.20578826501509506
(0, 586)      0.20578826501509506
```

Setiap baris menunjukkan sebuah judul skripsi, dan setiap pasangan dalam tanda kurung, menunjukkan bahwa kata tertentu memiliki bobot tertentu dalam judul skripsi tersebut. Nilai bobot ini menggambarkan pentingnya kata dalam dokumen.

Setelah itu, dilakukan pencetakan variabel y.

```
print(y)
```

Setelah dijalankan maka didapatkan hasil sebagai berikut :

```
0           SE
1           IS
2           IS
3           IS
4           IS
...
320         SE
321         SE
322     Networking
323         IS
324         SE
Name: Keterangan, Length: 325, dtype: object
```

Setiap baris menunjukkan label atau kategori yang sesuai dengan judul skripsi tertentu dalam dataset. Label ini akan digunakan sebagai target untuk pemodelan klasifikasi atau analisis teks lebih lanjut.

6. Pelatihan Model *Naive Bayes*

Pada tahapan pelatihan model *naive bayes* dilakukan beberapa proses sebagai berikut :

1. Dilakukan pembagian data menjadi data uji dan data latih. Untuk digunakan dalam tahap selanjutnya, seperti pelatihan model dan pengujian model klasifikasi.

```
# Bagi data menjadi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
print ('Train set:', X_train.shape, y_train.shape)
print ('Test set:', X_test.shape, y_test.shape)
```

2. Selanjutnya adalah melakukan pelatihan model *Naive Bayes*.

```
# Pelatihan Model Naive Bayes
clf = MultinomialNB()
clf.fit(X_train, y_train)
```

Setelah dijalankan, maka didapatkan hasil sebagai berikut :

```
▼ MultinomialNB
MultinomialNB()
```

7. Evaluasi Model

Pada tahap ini dilakukan evaluasi model untuk mendapatkan informasi tentang sejauh mana model berhasil dalam melakukan klasifikasi teks pada data pengujian.

```
# Evaluasi Model
y_pred = clf.predict(X_test)

print("Akurasi:", accuracy_score(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

Hasil outputnya sebagai berikut :

```
Akurasi: 0.8032786885245902
      precision    recall  f1-score   support

          IS       0.75      0.89      0.81       53
  Networking     0.80      0.33      0.47       12
           SE       0.87      0.82      0.85       57

   accuracy                           0.80      122
    macro avg       0.81      0.68      0.71      122
 weighted avg       0.81      0.80      0.79      122
```

Untuk membantu mengevaluasi kinerja model klasifikasi secara lebih rinci, maka ditampilkan dalam *confusion matrix*.

```
from sklearn.metrics import confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt

# Membuat confusion matrix

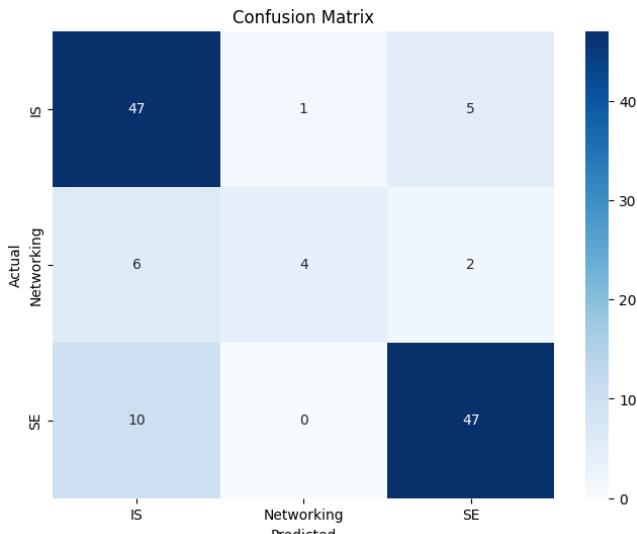
y_pred = clf.predict(X_test)
labels=['IS','Networking','SE']

cm = confusion_matrix(y_test, y_pred)

# Mencetak confusion matrix

plt.figure(figsize=(8, 6))
sns.heatmap(cm, annot=True, fmt='d', xticklabels=labels, yticklabels=labels, cmap='Blues')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()
```

Setelah dijalankan, maka diperoleh hasil sebagai berikut:



Gambar 4.2 Hasil Kinerja Model Klasifikasi

8. Klasifikasi Judul Skripsi

Pada tahapan ini dilakukan klasifikasi judul skripsi dengan menggunakan data testing. Tetapi sebelum itu, dilakukan klasifikasi pada satu judul skripsi terlebih dahulu.

```
# Klasifikasi Judul Skripsi untuk 1 judul
new_title = ["Sistem Pendukung Keputusan Pemilihan Anggota Paskibraka Menggunakan Metode Profile Matching"]
new_title = pd.Series(new_title)
new_title = new_title.apply(preprocess_text)
new_title_tfidf = vectorizer.transform(new_title)
predicted_class = clf.predict(new_title_tfidf)
print("Hasil Klasifikasi:", predicted_class)
```

Setelah dijalankan. Maka diperoleh hasil sebagai berikut :

```
Hasil Klasifikasi: ['SE']
```

Setelah dilakukan pengklasifikasian menggunakan satu judul skripsi, maka dapat pula dilakukan pengklasifikasian judul skripsi dengan jumlah yang banyak menggunakan excel.

```
# Klasifikasi Judul Skripsi dari File Testing Baru
dt = pd.read_excel("/content/drive/MyDrive/Skripsi/DataTesting.xlsx")
dt.head()
```

Setelah dijalankan, maka diperoleh hasil sebagai berikut :

No.	Judul Skripsi
0 1	sistem pendukung keputusan penentuan SMP terba...
1 2	Sistem Pendukung Keputusan Pemberian Beasiswa ...
2 3	SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN TENAGA KE...
3 4	SISTEM PENDUKUNG KEPUTUSAN PENENTUAN CALON PEN...
4 5	SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN AYAM PROD...

Dataset dalam format excel perlu dikonversi kebentuk CSV terlebih dahulu, agar dapat digunakan dalam pemodelan data dengan perintah sebagai berikut :

```
#Mengkonversi File Excel .xlsx ke .csv
dt.to_csv("/content/drive/MyDrive/Skripsi/DataTesting.csv", index=False)
dt.info()
```

Setelah dijalankan perintah diatas maka diperoleh hasil sebagai berikut :

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 2 columns):
 #   Column      Non-Null Count  Dtype  
---  --  
 0   No.          50 non-null    int64  
 1   Judul Skripsi 50 non-null   object 
dtypes: int64(1), object(1)
memory usage: 928.0+ bytes
```

Setelah itu, dilakukan klasifikasi pada file CSV dengan cara membaca data dari file CSV, melakukan pra-pemrosesan teks, melakukan ekstraksi fitur dengan metode TF-IDF, melakukan klasifikasi, dan menyimpan hasil klasifikasi dalam file CSV yang baru.

```
#Klasifikasi dari File yang Di-Load
new_data = pd.read_csv("/content/drive/MyDrive/Skripsi/DataTesting.csv") # Gantilah "data_baru.csv" dengan nama file yang di-load
new_data.info()

# Preprocessing data yang di-load
new_data['Judul Skripsi'] = new_data['Judul Skripsi'].apply(preprocess_text)

# Ekstraksi fitur dari data yang di-load menggunakan vectorizer yang sama
new_data_tfidf = vectorizer.transform(new_data['Judul Skripsi'])
predicted_classes = clf.predict(new_data_tfidf)

# Simpan hasil klasifikasi atau tampilkan sesuai kebutuhan
new_data['Klasifikasi Judul'] = predicted_classes
new_data.to_csv("/content/drive/MyDrive/Skripsi/hasil_klasifikasi.csv", index=False)
```

Hasil outputnya sebagai berikut:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 2 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   No.          50 non-null    int64  
 1   Judul Skripsi  50 non-null   object  
dtypes: int64(1), object(1)
memory usage: 928.0+ bytes
```

Selanjutnya, untuk menampilkan hasil klasifikasi maka digunakan kode program sebagai berikut :

```
pd.set_option('display.max_colwidth', None) # Untuk menampilkan seluruh isi kolom
print(new_data)
```

Hasil outputnya sebagai berikut :

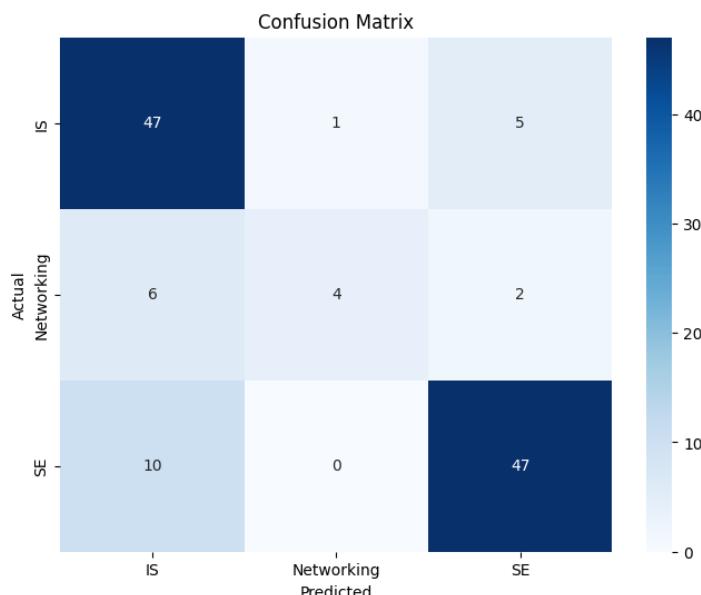
No.	Judul
0	SE
1	SE
2	SE
3	SE
4	SE
5	SE
6	SE
7	SE
8	SE
9	SE
10	SE
11	sistem pendukung keputusan penentuan smp sistem pendukung keputusan pemberian beasiswa peningkatan prestasi akademik ppa metode additive we: sistem pendukung keputusan pemilihan tenaga kesehatan telan sistem pendukung keputusan calon penerima bantuan pangan non tunai daerah bpntd metode addit sistem pendukung keputusan pemilihan ayam produktif metode saw simple sistem pendukung keputusan penentuan bantuan sembako dinas sosial sistem pendukung keputusan menentukan mitra kerja metode multi objective oopt sistem pendukung keputusan penerima bantuan bibit jagung metode simple sistem pendukung keputusan kelayakan anggota koperasi metode simple sistem pendukung keputusan penilaian kinerja karyawan kontrak sistem pendukung keputusan penerima bantuan masyarakat pengguna

BAB V

PEMBAHASAN PENELITIAN

5.1 Pembahasan Model

Berdasarkan hasil pengolahan 404 *data training* menggunakan algoritma *Naive Bayes* maka dibuat dalam model *confusion matrix* dengan hasil sebagai berikut :



Gambar 5.1 *Confusion Matrix*

Berdasarkan hasil *confusion matrix* diatas, menunjukkan bahwa nilai *true positive* untuk konsentrasi IS sebanyak 47 data, nilai *true positive* untuk konsentrasi *Networking* sebanyak 4 data, dan nilai *true positive* untuk konsentrasi SE sebanyak 47 data.

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

Untuk cara memperoleh nilai precision, recall, dan f1-score pada **Tabel 5.1** di atas, maka dilakukan perhitungan salah satu konsentrasi diatas yaitu konsentrasi IS. cara perhitungannya sebagai berikut:

$$TP = 47$$

$$FP = 6+10 = 16$$

$$FN = 1+5 = 6$$

$$1. \text{ Precision IS} = \frac{47}{47+16} = 0.75$$

$$2. \text{ Recall IS} = \frac{47}{47+6} = 0.89$$

$$3. \text{ F1-Score IS} = 2 \frac{0.75 * 0.89}{0.75 + 0.89} = 0.81$$

$$4. \text{ Accuracy} = \frac{47+4+47}{122} = 0.80$$

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 eksponensial 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.

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.

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

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

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

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

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

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

6. Menghitung probabilitas pada data *testing*.

$$\begin{aligned} P(\text{Testing|SE}) \\ &= P(\text{SE}) * P(\text{Analisis|SE}) * P(\text{Jaringan|SE}) \\ &= 0.333333333 * 0.0714285714 * 0.0714285714 \\ &= 0.0017006803 \end{aligned}$$

$$\begin{aligned} P(\text{Testing|IS}) \\ &= P(\text{IS}) * P(\text{Analisis|IS}) * P(\text{Jaringan|IS}) \\ &= 0.333333333 * 0.0714285714 * 0.0714285714 \\ &= 0.0017006803 \end{aligned}$$

$$\begin{aligned} P(\text{Testing|Net}) \\ &= P(\text{Net}) * P(\text{Analisis|Net}) * P(\text{Jaringan|Net}) \\ &= 0.333333333 * 0.1428571429 * 0.1428571429 \\ &= 0.0068027211 \end{aligned}$$

Dari perhitungan diatas, nilai probabilitas tertinggi yaitu sebesar 0.0068027211 pada $P(\text{Uji}|\text{Net})$ sehingga judul skripsi tersebut diklasifikasikan ke dalam kelas “Networking”.

5.3 Pembahasan Hasil Pemodelan

Setelah pengklasifikasian judul skripsi menggunakan *data training* telah dilakukan dan nilai akurasinya telah muncul, maka dapat dilakukan pengujian model *Naive Bayes* untuk melihat kinerja model tersebut dengan menggunakan 50 *data testing* sebagai berikut :

Tabel 5.6 Data Testing

No.	Judul Skripsi	Ket
1	sistem pendukung keputusan penentuan SMP terbaik di kabupaten Pohuwato menggunakan metode Smart	?
2	Sistem Pendukung Keputusan Pemberian Beasiswa Peningkatan Prestasi Akademik (PPA) menggunakan Metode Simple Additive Weighting (SAW) pada Universitas Ichsan Gorontalo	?
3	SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN TENAGA KESEHATAN TELADAN MENGGUNAKAN METODE SIMPLE ADDITIVE WEIGHTING (SAW)	?
4	SISTEM PENDUKUNG KEPUTUSAN PENENTUAN CALON PENERIMA BANTUAN PANGAN NON TUNAI DAERAH (BPNTD) MENGGUNAKAN METODE ADDITIVE RATIO ASSESMENT (ARAS) PADA DESA MUTIARA	?
5	SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN AYAM PRODUKTIF MENGGUNAKAN METODE SAW (SIMPLE ADDITIVE WEIGHTING) DI PETERNAKAN KIBAR JAYA	?
6	SISTEM PENDUKUNG KEPUTUSAN PENENTUAN BANTUAN SEMBAKO DI DINAS SOSIAL MENGGUNAKAN METODE PREFERENCE SELECTION INDEX (PSI)	?
7	SISTEM PENDUKUNG KEPUTUSAN MENENTUKAN MITRA KERJA MENGGUNAKAN METODE MULTI OBJECTIVE OPTIMIZATION ON THE BASIS OF RATIO ANALYSIS	?
8	SISTEM PENDUKUNG KEPUTUSAN PENERIMA BANTUAN BIBIT JAGUNG MENGGUNAKAN METODE SIMPLE MULTY ATTRIBUTE RATING TECHNIQUE (SMART)	?
9	sistem pendukung keputusan kelayakan anggota koperasi menggunakan metode simple multi attribute rating technique (SMART)	?

No.	Judul Skripsi	Ket
10	sistem pendukung keputusan Penilaian kinerja karyawan kontrak menggunakan metode Multy attribute utility Theory(MAUT)	?
...
50	Analisa Jaringan Internet Menggabungkan Local Area Network (LAN) Dan Wireless Local Area Network (WLAN) Untuk Mendukung Kinerja Pegawai	?

Setelah dilakukan klasifikasi dengan menggunakan model *Naive Bayes*, maka didapatkan hasil klasifikasi sebagai berikut :

Tabel 5.7 Hasil Klasifikasi

No.	Judul Skripsi	Ket
1	sistem pendukung keputusan penentuan smp terbaik kabupaten pohuwato metode smart	SE
2	sistem pendukung keputusan pemberian beasiswa peningkatan prestasi akademik ppa metode simple additive weighting saw universitas ichsan gorontalo	SE
3	sistem pendukung keputusan pemilihan tenaga kesehatan teladan metode simple additive weighting saw	SE
4	sistem pendukung keputusan penentuan calon penerima bantuan pangan non tunai daerah bpntd metode additive ratio assesment aras desa mutiara	SE
5	sistem pendukung keputusan pemilihan ayam produktif metode saw simple additive weighting peternakan kibar jaya	SE
6	sistem pendukung keputusan penentuan bantuan sembako dinas sosial metode preference selection index psi	SE
7	sistem pendukung keputusan menentukan mitra kerja metode multi objective ooptimization on the basis of ratio analysis	SE
8	sistem pendukung keputusan penerima bantuan bibit jagung metode simple multy attribute rating technique smart	SE
9	sistem pendukung keputusan kelayakan anggota koperasi metode simple multi attribute rating technique smart	SE

No.	Judul Skripsi	Ket
10	sistem pendukung keputusan penilaian kinerja karyawan kontrak metode multy attribute utility theorymaut	SE
...
50	analisa jaringan internet menggabungkan local area network lan wireless local area network wlan mendukung kinerja pegawai	IS

Berdasarkan hasil klasifikasi di atas, dengan memperoleh hasil klasifikasi 80% dapat disimpulkan bahwa model *Naive Bayes* tersebut, dapat mengklasifikasikan data judul skripsi dengan cukup baik sehingga judul skripsi dapat terbagi sesuai dengan konsentrasi masing-masing. Walaupun masih ada beberapa hasil klasifikasi yang belum sesuai, tetapi sudah lebih banyak hasil klasifikasi yang sesuai. Hasil yang sesuai dan belum sesuai dapat dilihat pada **Tabel 5.4** berikut:

Tabel 5.8 Hasil Klasifikasi Sesuai dan Belum Sesuai

No.	Judul Skripsi	Ket	Hasil Klasifikasi
1	sistem pendukung keputusan penentuan smp terbaik kabupaten pohuwato metode smart	SE	Sesuai
2	sistem pendukung keputusan pemberian beasiswa peningkatan prestasi akademik ppa metode simple additive weighting saw universitas ichsan gorontalo	SE	Sesuai
3	sistem pendukung keputusan pemilihan tenaga kesehatan teladan metode simple additive weighting saw	SE	Sesuai
4	sistem pendukung keputusan penentuan calon penerima bantuan pangan non tunai daerah bpntd metode additive ratio assesment aras desa mutiara	SE	Sesuai
5	sistem pendukung keputusan pemilihan ayam produktif metode saw simple additive weighting peternakan kibar jaya	SE	Sesuai
6	sistem pendukung keputusan penentuan bantuan sembako dinas sosial metode preference selection index psi	SE	Sesuai
7	sistem pendukung keputusan menentukan mitra kerja metode multi objective optimization on the basis of ratio analysis	SE	Sesuai

No.	Judul Skripsi	Ket	Hasil Klasifikasi
8	sistem pendukung keputusan penerima bantuan bibit jagung metode simple multy attribute rating technique smart	SE	Sesuai
9	sistem pendukung keputusan kelayakan anggota koperasi metode simple multi attribute rating technique smart	SE	Sesuai
10	sistem pendukung keputusan penilaian kinerja karyawan kontrak metode multy attribute utility theorymaut	SE	Sesuai
...	
50	analisa jaringan internet menggabungkan local area network lan wireless local area network wlan mendukung kinerja pegawai	IS	Belum sesuai

BAB VI

KESIMPULAN DAN SARAN

6.1 Kesimpulan

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

1. Hasil dari penerapan metode *Naive Bayes Classifier* pada klasifikasi judul skripsi memperoleh hasil yang cukup baik sehingga judul skripsi dapat terbagi sesuai dengan konsentrasi masing-masing. Karena, walaupun masih ada beberapa hasil klasifikasi judul skripsi yang belum sesuai dengan konsentrasi, 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.

6.2 Saran

Berdasarkan kesimpulan di atas, peneliti dapat memberikan saran kepada penelitian selanjutnya yaitu:

1. Penelitian selanjutnya bisa menggunakan dataset lebih banyak lagi. Karena hal tersebut dapat meningkatkan akurasi dari algoritma *Naive Bayes Classifier*.
2. Penelitian ini dapat dikembangkan dengan menggunakan metode klasifikasi lain seperti *Decision Tree*, *K-Nearest Neighbor*, dan *Support Vector Machine*, sehingga bisa membandingkan hasil klasifikasi yang didapatkan.

DAFTAR PUSTAKA

- [1] M. Suhendri and Y. Afrilia, “Klasifikasi Karya Ilmiah (Tugas Akhir) Mahasiswa Menggunakan Metode Naive Bayes Classifier (Nbc).” [Online]. Available: <http://sistemas.ftik.unisi.ac.id>
- [2] “Evaluasi Ekstraksi Fitur Klasifikasi Teks Untuk Peningkatan Akurasi Klasifikasi Menggunakan Naive Bayes”.
- [3] U. Pujiyanto, T. Widyaningtyas, 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 Arifyanti, “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.

DATA YANG DIKUMPULKAN

No	NIM	Nama Lengkap	Judul Skripsi	Ket
1	T3117369	Marsela Herawati Sarga	PENERAPAN METODE PREFERENCE 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
3	T3117326	SINTIA HUNOWU	SISTEM PENDUKUNG KEPUTUSAN PENETAPAN PENERIMA BANTUAN PANGAN NON TUNAI (BPNT) MENGGUNAKAN METODE PREFERENCE SELECTION INDEX (PSI)	SE
4	T3115174	SUNARDI R IBRAHIM	KLASIFIKASI KELAYAKAN PEMBERIAN PINJAMAN MENGGUNAKAN METODE SUPPORT VEKTOR MANCHINE (STUDI KASUS KOPERASI MERTHA JAYA BUOL)	SE
5	T3117332	SINTIA LAHAY	SISTEM PENDUKUNG KEPUTUSAN PENERIMA BEASISWA PRESTASI DAERAH DENGAN MENERAPKAN METODE COMPOSITE PERFORMANCE INDEX (CPI)	SE
6	T3117306	RESKA NGABITO	Sistem pendukung keputusan penetapan penerima bantuan rumah rehab menggunakan metode fuzzy mamdani	SE
7	T3117351	Rivandi Hulopi	sistem pendukung keputusan penetapan penerima bantuan program keluarga harapan(pkj) menggunakan metode moora	SE
8	T3117368	Merin Nurlaisa Abas	sistem pendukung keputusan penerima bantuan rumah rehab dengan menerapkan metode composite performance index di desa limbula	SE
9	T3117254	YASIN K. ADAM	penerapan Metode Single Moving Average Untuk Memprediksi Penjualan Air Minum Aqua Berdasarkan Jenis Kemasan	IS
10	T3117116	ratna kasim	Prediksi hasil produksi jagung menggunakan metode eksponensial moving average	IS
11	T3115214	abdul kadir arsad	ANALISIS KUALITAS JARINGAN BACKBONE NIRKABEL DI BLOK PLAN PERKANTORAN MENGGUNAKAN STANDAR QUALITY OF SERVICE (QOS)	Networking

No	NIM	Nama Lengkap	Judul Skripsi	Ket
12	T3117391	FAISAL B. ABAY	CLUSTERING DENGAN ALGORITMA K-MENS TERHADAP CALON PENERIMA BANTUAN RENOVASI RUMAH	SE
13	T3118146	MEGAWATI	SPK UNTUK MENENTUKAN MINAT SISWA DALAM MEMILIH KEGIATAN EKTRAKURIKULER MENGGUNAKAN METODE ANALYTICAL HIERARCHY DAN METODE SIMPLE PROCESS ADDITIVE	SE
14	T3115061	MOHAMAD RIFKI RAUF	PREDIKSI PENJUALAN PRODUK LIPSTIK FANTASTIC MENGGUNAKAN METODE LEAST SQUARE	IS
15	T3118280	T3118280	Sistem Pendukung Keputusan Pemilihan Obat Batuk Di Apotek Di Hidayah Mananggu Menggunakan Metode Maut	SE
16	T3118237	Alfian Usman	SISTEM PENDUKUNG KEPUTUSAN PENENTUAN KELOMPOK TANI PENERIMA BANTUAN BIBIT JAGUNG MENGGUNAKAN METODE MAUT	SE
17	T3118211	Nengsi Djafar	Sistem pendukung keputusan penerima bantuan pupuk menggunakan metode multi objective optimization on the basis ratio analisis(moora) pada desa teratai	SE
18	T3118044	MOHAMAD IKBAL HULUKATI	IMPLEMENTASI PRIVATE CLOUD STORAGE PADA JARINGAN LOKAL DI LAB KOMPUTER	Networking
19	T3117114	Rifdah Rofifah Faruk Abdullah	PENERAPAN CONVOLUTIONAL NEURAL NETWORK (CNN) UNTUK KLASIFIKASI PENGGUNAAN MASKER	SE
20	T3117140	Sintiya Dhewasri Habiba Lahabu	Penerapan metode least square untuk memprediksi harga pangan di kota gorontalo	IS
...
404	T3118324	RAHMAT HIDAYAT	Sistem Pengaman Pintu Menggunakan RFID TAG Berbasis Arduino Uno	IS

KODE PROGRAM

```
#Mengimport Library
import pandas as pd
import numpy as np

#Proses Pengaitan dengan Google Drive
from google.colab import drive
drive.mount('/content/drive')

#Membaca Dataset
df =
pd.read_excel("/content/drive/MyDrive/Skripsi/filejudul.xlsx")
df.head()

#Mengkonversi File Excel .xlsx ke .csv
df.to_csv("/content/drive/MyDrive/Skripsi/filejudul.csv",
index=False)
df.info()

import nltk
nltk.download('stopwords')


import pandas as pd
import re
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from sklearn.feature_extraction.text import TfidfVectorizer

# Fungsi untuk melakukan preprocessing pada teks
def preprocess_text(text):
    # Mengonversi huruf menjadi huruf kecil
    text = text.lower()

    # Menghapus tanda baca dan karakter khusus
    text = re.sub(r'^a-zA-Z0-9\s]', '', text)

    # Menghapus kata-kata umum (stop words)
    stop_words = set(stopwords.words('indonesian')) #
Menggunakan daftar kata-kata umum bahasa Indonesia
words = text.split()
```

```

# Menggabungkan kata-kata kembali menjadi teks
text = ' '.join(words)

return text

# Memanggil Fungsi Preprocessing Data
df['Judul Skripsi'] = df['Judul
Skripsi'].apply(preprocess_text)
df.head()

#Simpan Hasil klasifikasi ke Excel
df.to_excel("/content/drive/MyDrive/Skripsi/hurufkecil.xlsx",
index=False)

# Menghitung jumlah setiap kategori Judul dalam kolom
'Keterangan'
df['Keterangan'].value_counts()

!pip install Sastrawi

import pandas as pd
import re
from nltk.corpus import stopwords
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, classification_re

# Ekstraksi Fitur
# mengubah teks judul skripsi menjadi representasi angka
vectorizer = TfidfVectorizer()
X = vectorizer.fit_transform(df['Judul Skripsi'])
y = df['Keterangan']
print(X)
print(y)

```

```

# Bagi data menjadi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.3, random_state=42)
print ('Train set:', X_train.shape, y_train.shape)
print ('Test set:', X_test.shape, y_test.shape)
# Pelatihan Model Naive Bayes
clf = MultinomialNB()
clf.fit(X_train, y_train)
# Evaluasi Model
y_pred = clf.predict(X_test)

print("Akurasi:", accuracy_score(y_test, y_pred))
print(classification_report(y_test, y_pred))

from sklearn.metrics import confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt

# Membuat confusion matrix

y_pred = clf.predict(X_test)
labels=['IS','Networking','SE']

cm = confusion_matrix(y_test, y_pred)

# Mencetak confusion matrix

plt.figure(figsize=(8, 6))
sns.heatmap(cm, annot=True, fmt='d', xticklabels=labels,
yticklabels=labels, cmap='Blues')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()

# Klasifikasi Judul Skripsi untuk 1 judul
new_title = ["Clustering Kepuasan Masyarakat Untuk Pelayanan
Puskesmas Menggunakan Metode Naive Bayes"]
new_title = pd.Series(new_title)
new_title = new_title.apply(preprocess_text)
new_title_tfidf = vectorizer.transform(new_title)
predicted_class = clf.predict(new_title_tfidf)
print("Hasil Klasifikasi:", predicted_class)

```

```

# Klasifikasi Judul Skripsi dari File Testing Baru
dt =
pd.read_excel("/content/drive/MyDrive/Skripsi/DataTesting.xlsx")
dt.head()
#Mengkonversi File Excel .xlsx ke .csv
dt.to_csv("/content/drive/MyDrive/Skripsi/DataTesting.csv",
index=False)
dt.info()
#Klasifikasi dari File yang Di-Load
new_data =
pd.read_csv("/content/drive/MyDrive/Skripsi/DataTesting.csv")
) # Gantilah "data_baru.csv" dengan nama file yang ingin
Anda klasifikasikan.
new_data.info()

# Preprocessing data yang di-load
new_data['Judul Skripsi'] = new_data['Judul
Skripsi'].apply(preprocess_text)

# Ekstraksi fitur dari data yang di-load menggunakan
vectorizer yang sama
new_data_tfidf = vectorizer.transform(new_data['Judul
Skripsi'])
predicted_classes = clf.predict(new_data_tfidf)

# Simpan hasil klasifikasi atau tampilkan sesuai kebutuhan
new_data['Klasifikasi Judul'] = predicted_classes
new_data.to_csv("/content/drive/MyDrive/Skripsi/hasil_klasifikasi.csv", index=False)

pd.set_option('display.max_colwidth', None) # Untuk
menampilkan seluruh isi kolom
print(new_data)

#Simpan Hasil klasifikasi ke Excel
new_data.to_excel("/content/drive/MyDrive/Enric2023/hasiljdl
_klasifikasi.xlsx", index=False)

```

KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS ICHSAN GORONTALO
FAKULTAS ILMU KOMPUTER
SURAT KEPUTUSAN MENDIKNAS RI NOMOR 84/D/O/2001
Jl. Achmad Nadjamuddin No. 17 Telp (0435) 829975 Fax (0435) 829976 Gorontalo

SURAT KETERANGAN PENELITIAN

Nomor : 488/FIKOM-UIG/SKP/XI/2023

Yang bertanda tangan dibawah ini :

N a m a	:	Sudirman S. Panna, M.Kom
Jabatan	:	Ketua Program Studi Teknik Informatika

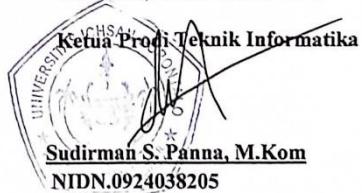
Dengan ini Menerangkan bahwa :

N a m a Mahasiswa	:	Salmin Dania
N I M	:	T3119032
Program Studi	:	Teknik Informatika

Bahwa yang bersangkutan benar-benar telah melakukan penelitian tentang “**Klasifikasi Judul Skripsi Berdasarkan Konsentrasi Pada Program Studi Teknik Informatika**” Guna untuk meyelesaikan Studi pada Program Studi Teknik Informatika Fakultas Ilmu Komputer, dan bersangkutan telah menyelesaikan penelitian Tersebut pada **TGL 18 November 2023** sesuai dengan waktu yang telah ditentukan.

Demikian Surat Keterangan ini dibuat dan digunakan untuk seperlunya.

Gorontalo, 23 November 2023





**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
 UNIVERSITAS IHSAN GORONTALO
 FAKULTAS ILMU KOMPUTER
 SURAT KEPUTUSAN MENDIKNAS RI NOMOR 84/D/O/2001
 Jl. Achmad Najamuddin No. 17 Telp. (0435) 829975 Fax (0435) 829976 Gorontalo**

SURAT REKOMENDASI BEBAS PLAGIASI
 No. 493/FIKOM-UIG/R/XI/2023

Yang bertanda tangan di bawah ini :

Nama : Irvan Abraham Salihi, M.Kom
 NIDN : 0928028101
 Jabatan : Dekan Fakultas Ilmu Komputer

Dengan ini menerangkan bahwa :

Nama Mahasiswa : Salmin Dania
 NIM : T3119032
 Program Studi : Teknik Informatika (S1)
 Fakultas : Fakultas Ilmu Komputer
 Judul Skripsi : Penerapan Algoritma Naive Bayes Classifier Untuk
 Klasifikasi Judul Skripsi Berdasarkan Konsentrasi

Sesuai hasil pengecekan tingkat kemiripan skripsi melalui aplikasi **Turnitin** untuk judul skripsi di atas diperoleh hasil *Similarity* sebesar **28%**, berdasarkan Peraturan Rektor No. 32 Tahun 2019 tentang Pendekripsi Plagiat pada Setiap Karya Ilmiah di Lingkungan Universitas Ichsan Gorontalo dan persyaratan pemberian surat rekomendasi verifikasi calon wisudawan dari LLDIKTI Wil. XVI, bahwa batas kemiripan skripsi maksimal 30%, untuk itu skripsi tersebut di atas dinyatakan **BEBAS PLAGIASI** dan layak untuk diujangkan.

Demikian surat rekomendasi ini dibuat untuk digunakan sebagaimana mestinya.

Mengetahui
 Dekan,

Irvan Abraham Salihi, M.Kom
 NIDN. 0928028101

Gorontalo, 25 November 2023
 Tim Verifikasi,

Zulfrianto Y. Lamasiqi, M.Kom
 NIDN. 0914089101

Terlampir:
 Hasil Pengecekan Turnitin

 **Similarity Report ID:** oid:25211:47242866

PAPER NAME	AUTHOR
SKRIPSI_T3119032_SALMIN_DANIA.pdf	SALMIN DANIA salmindania275@gmail.com
<hr/>	
WORD COUNT	CHARACTER COUNT
6475 Words	40850 Characters
PAGE COUNT	FILE SIZE
46 Pages	1.6MB
SUBMISSION DATE	REPORT DATE
Nov 23, 2023 8:08 PM GMT+8	Nov 23, 2023 8:09 PM GMT+8
<hr/>	
● 28% Overall Similarity	
The combined total of all matches, including overlapping sources, for each database.	
<ul style="list-style-type: none">• 28% Internet database• Crossref database• 3% Submitted Works database• 6% Publications database• Crossref Posted Content database	
● Excluded from Similarity Report	
<ul style="list-style-type: none">• Bibliographic material• Cited material• Quoted material• Small Matches (Less than 30 words)	
<hr/>	
Summary	

 **Similarity Report ID:** oid:25211:47242866

● **28% Overall Similarity**

Top sources found in the following databases:

- 28% Internet database
- Crossref database
- 3% Submitted Works database
- 6% Publications database
- Crossref Posted Content database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

Rank	Source	Category	Similarity (%)
1	jurnaldrpm.budiluhur.ac.id	Internet	8%
2	id.123dok.com	Internet	3%
3	core.ac.uk	Internet	3%
4	media.neliti.com	Internet	2%
5	ejournal.bsi.ac.id	Internet	2%
6	eprints.utdi.ac.id	Internet	2%
7	repositori.uma.ac.id	Internet	1%
8	docplayer.info	Internet	1%

Sources overview

 turnitin

Similarity Report ID: oid:25211:47242866

9	tunasbangsa.ac.id Internet	1%
10	e-journal.metrouniv.ac.id Internet	<1%
11	journal.stekom.ac.id Internet	<1%
12	repository.unmuhjember.ac.id Internet	<1%
13	andi.ddns.net Internet	<1%
14	mulok.library.um.ac.id Internet	<1%
15	repository.uinjkt.ac.id Internet	<1%
16	coursehero.com Internet	<1%
17	repository.ittelkom-pwt.ac.id Internet	<1%



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS ICHSAN GORONTALO
FAKULTAS ILMU KOMPUTER
UPT. PERPUSTAKAAN FAKULTAS
SK. MENDIKNAS RI NO. 84/D/0/2001
Jl. Achmad Nadjamuddin No.17 Telp(0435) 829975 Fax. (0435) 829976 Gorontalo

SURAT KETERANGAN BEBAS PUSTAKA

No : 020/Perpustakaan-Fikom/XI/2023

Perpustakaan Fakultas Ilmu Komputer (FIKOM) Universitas Ichsan Gorontalo dengan ini menerangkan bahwa :

Nama Anggota : Salmin Dania
No. Induk : T3119032
No. Anggota : M202382

Terhitung mulai hari, tanggal : Senin, 13 November 2023, dinyatakan telah bebas pinjam buku dan koleksi perpustakaan lainnya.

Demikian keterangan ini di buat untuk di pergunakan sebagaimana mestinya.

Gorontalo, 13 November 2023

Mengetahui,
Kepala Perpustakaan

Apriyanto Alhamad, M.Kom

NIDN : 0924048601



RIWAYAT HIDUP PENELITI



Nama	: Salmin Dania
NIM	: T3119032
Tempat, Tgl Lahir	: Bulagi, 30 Desember 2001
Agama	: Islam
Jenis Kelamin	: Perempuan
Fakultas/Jurusan	: Ilmu Komputer/Teknik Informatika
Konsentrasi	: Software Engineering
Alamat	: Jl. Jendral Sudirman
Email	: salmindania275@gmail.com

Riwayat Pendidikan:

1. SD Negeri 2 Bulagi (2007 – 2013)
2. SMP Negeri 1 Bulagi (2013 – 2016)
3. SMA Negeri 1 Bulagi (2016 – 2019)
4. Universitas Ichsan Gorontalo (2019 - Sekarang)