

**KLASIFIKASI BIDANG ILMU PADA PUBLIKASI TERINDEKS
SCOPUS MENGGUNAKAN METODE NAÏVE BAYES**

LAPORAN TUGAS AKHIR

Laporan ini Disusun Guna Memenuhi Salah Satu Syarat Memperoleh Gelar Sarjana
Strata (S1) pada Program Studi Teknik Informatika Fakultas Teknologi Industri
Universitas Islam Sultan Agung Semarang



DISUSUN OLEH

KUSUMA NURNASIKHA

NIM 32601800015

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNOLOGI INDUSTRI

UNIVERSITAS ISLAM SULTAN AGUNG

SEMARANG

2022

FINAL PROJECT

**CLASSIFICATION OF SCOPUS INDEXED PUBLICATIONS USING THE
NAÏVE BAYES METHOD**

*Proposed to complete the requirement to obtain a bachelor's degree (S1) at
Informatics Engineering Departement of Industrial Technology Faculty Sultan
Agung Islamic University*



Arranged By:

KUSUMA NURNASIKHA

NIM 32601800015

**MAJORING OF INFORMATICS ENGINEERING
INDUSTRIAL TECHNOLOGY FACULTY
SULTAN AGUNG ISLAMIC UNIVERSITY
SEMARANG**

2022

LEMBAR PENGESAHAN PEMBIMBING

Laporan Tugas Akhir dengan judul “**Klasifikasi Bidang Ilmu pada Publikasi Terindeks Scopus Menggunakan Metode Naïve Bayes**” ini disusun oleh:

Nama : Kusuma Nurnasikha

NIM : 32601800015

Program Studi : Teknik Informatika

Telah disahkan oleh dosen pembimbing pada :


Hari : Senin


Tanggal : 13 Januari 2023

Mengesahkan,

Pembimbing I

Pembimbing II



Sam Farisa C.H, ST.M.Kom
NIDN. 0628028602


Imam Much Ibnu S, ST, M.Sc,Ph,D
NIDN. 0613037301

Mengetahui,

Ketua Program Studi Teknik Informatika
Fakultas Teknologi Industri
Universitas Islam Sultan Agung




Ir. Sri Mulyono, M.Eng
NIDN. 0626066601

LEMBAR PENGESAHAN PENGUJI

Laporan Tugas Akhir dengan judul “**Klasifikasi Bidang Ilmu pada Publikasi Terindeks Scopus Menggunakan Metode Naïve Bayes**” ini telah dipertahankan di depan dosen penguji Tugas Akhir pada :

Hari : Senin

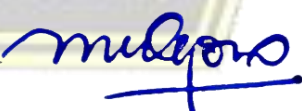
Tanggal : 09 Januari 2023



Penguji I

Penguji II


Andi Riansyah, ST, M.Kom
NIDN. 0609108802


Ir. Sri Mulyono, M.Eng
NIDN. 0626066601

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Kusuma Nurnasikha

NIM : 32601800015

Judul Tugas Akhir : Klasifikasi Bidang Ilmu pada Publikasi Terindeks Scopus
Menggunakan Metode Naïve Bayes

Dengan bahwa ini saya menyatakan bahwa judul dan isi Tugas Akhir yang saya buat dalam rangka menyelesaikan Pendidikan Strata Satu (S1) Teknik Informatika tersebut adalah asli dan belum pernah diangkat, ditulis ataupun dipublikasikan oleh siapapun baik keseluruhan maupun sebagian, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka, dan apabila di kemudian hari ternyata terbukti bahwa judul Tugas Akhir tersebut pernah diangkat, ditulis ataupun dipublikasikan, maka saya bersedia dikenakan sanksi akademis. Demikian surat pernyataan ini saya buat dengan sadar dan penuh tanggung jawab.

Semarang, 03 Februari 2023

Yang menyatakan,



Kusuma Nurnasikha

SURAT PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan dibawah ini :

Nama : Kusuma Nurnasikha
NIM : 32601800015
Program Studi : Teknik Informatika
Fakultas : Teknologi Industri
Alamat Asal : Jl. Kejaksan Desa Pedagangan RT 05/RW 01 Kec.
Dukuhwaru Kab. Tegal Jawa Tengah

Dengan ini menyatakan Karya Ilmiah berupa Tugas Akhir dengan Judul :
Klasifikasi Bidang Ilmu pada Publikasi Terindeks Scopus Menggunakan Metode
Naïve Bayes.

Menyetujui menjadi hak milik Universitas Islam Sultan Agung serta memberikan
Hak bebas Royalti Non-Eksklusif untuk disimpan, dialihmediakan, dikelola dan
pangkalan data dan dipublikasikan diinternet dan media lain untuk kepentingan
akademis selama tetap menyantumkan nama penulis sebagai pemilik hak cipta.
Pernyataan ini saya buat dengan sungguh-sungguh. Apabila dikemudian hari
terbukti ada pelanggaran Hak Cipta/Plagiarisme dalam karya ilmiah ini, maka
segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa
melibatkan Universitas Islam Sultan Agung.

Semarang, 03 Februari 2023

Yang menyatakan,



Kusuma Nurnasikha

KATA PENGANTAR

Dengan mengucapkan syukur alhamdulillah atas kehadiran Allah SWT yang telah memberikan rahmat dan karunianya kepada penulis, sehingga dapat menyelesaikan Tugas Akhir dengan judul “Klasifikasi Bidang Ilmu Pada Publikasi Terindeks Scopus Menggunakan Metode Naïve Bayes” ini untuk memenuhi salah satu syarat menyelesaikan studi serta dalam rangka memperoleh gelar sarjana (S-1) pada Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang.

Tugas Akhir ini disusun dan dibuat dengan adanya bantuan dari berbagai pihak, materi maupun teknis, oleh karena itu saya selaku penulis mengucapkan terima kasih kepada:

1. Rektor UNISSULA Bapak Prof. Dr. H. Gunarto, SH., M.Hum yang mengizinkan penulis menimba ilmu di kampus ini.
2. Dekan Fakultas Teknologi Industri Ibu Dr. Novi Marlyana, ST., MT.
3. Dosen pembimbing I penulis Sam Farisa Chaerul Haviana, ST., M.Kom. yang telah meluangkan waktu dan memberi ilmu.
4. Dosen pembimbing II penulis Imam Much Ibnu Subroto, ST, M.Sc., Ph.D. yang telah memberikan banyak nasehat dan saran.
5. Orang tua penulis yang telah mengizinkan untuk menyelesaikan laporan ini.
6. Dan kepada semua pihak yang tidak dapat saya satu persatu.

Dengan segala kerendahan hati, penulis menyadari masih banyak terdapat banyak kekurangan-kekurangan dari segi kualitas atau kuantitas maupun dari ilmu pengetahuan dalam penyusunan laporan, sehingga penulis mengharapkan adanya saran dan kritikan yang bersifat membangun demi kesempurnaan laporan ini.

Semarang, 03 Februari 2023



Kusuma Nurnasikha

DAFTAR ISI

LAPORAN TUGAS AKHIR	i
FINAL PROJECT	ii
KATA PENGANTAR	iii
DAFTAR ISI.....	iv
DAFTAR GAMBAR	vi
DAFTAR TABEL.....	vii
ABSTRAK.....	viii
BAB I.....	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	2
1.3 Pembatasan Masalah	2
1.4 Tujuan.....	2
1.5 Manfaat.....	2
1.6 Sistematika Penulisan.....	3
BAB II.....	4
TINJAUAN PUSTAKA DAN DASAR TEORI	4
2.1 Tinjauan Pustaka	4
2.2 Dasar Teori.....	12
2.2.1 Data Mining	12
2.2.2 Klasifikasi	12
2.2.3 Naïve bayes	12
2.2.3.1 Pengertian Naïve Bayes	12
2.2.3.2 Keuntungan Metode Naïve Bayes	13
2.2.3.3 Rumus Naïve Bayes	13
2.2.4 Scopus	13
2.2.5 Macam-macam 5 bidang ilmu.....	13

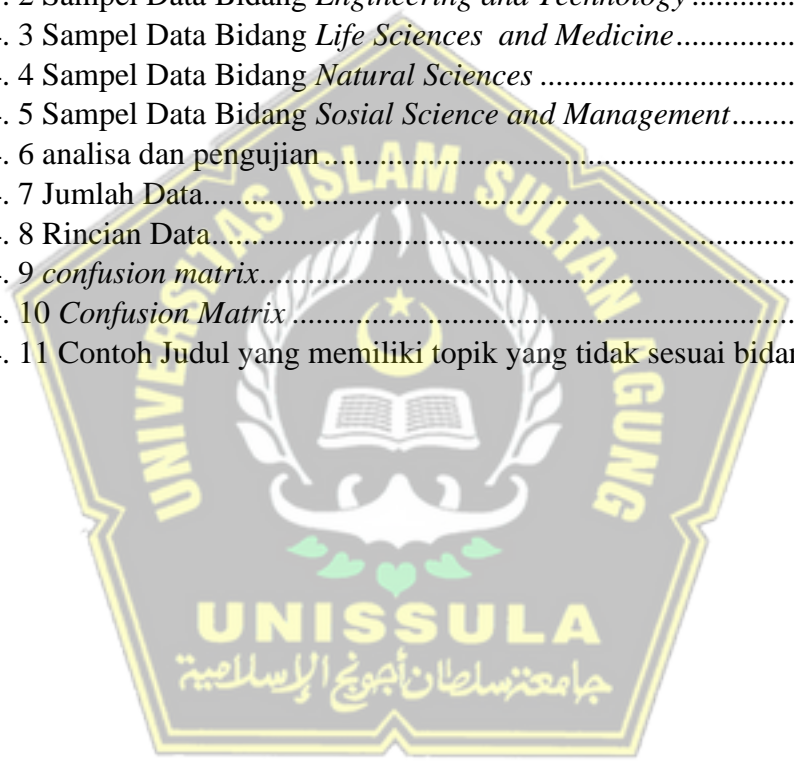
BAB III	15
METODE PENELITIAN.....	15
3.1 Metodologi Penelitian	15
3.1.1 Pengumpulan data Scopus	17
3.1.2 <i>Pre-Processing</i>	17
3.1.3 Penerapan Metode.....	19
3.1.4 Evaluasi.....	20
3.2 Metodologi Perancangan Alur Sistem.....	21
3.2.1 Analisis Kebutuhan	21
3.2.2 Implementasi Sistem	22
3.2.3 Alur Implementasi Sistem.....	22
3.3 Perancangan Interface	25
BAB IV	29
HASIL DAN ANALISIS PENELITIAN.....	29
4.1 Hasil.....	29
4.1.1 Pengumpulan data Scopus	29
4.1.2 <i>Pre-Processing</i>	33
4.1.3 Penerapan Metode Naïve Bayes.....	34
4.1.4 Evaluasi.....	35
4.2 <i>User Interface</i> dan Penggunaan Sistem.....	35
4.3 Analisa dan Pengujian.....	39
4.4 Analisis Akurasi	39
4.5 Pembahasan Hal Yang Mempengaruhi Hasil Akurasi	42
BAB V	45
KESIMPULAN DAN SARAN.....	45
5.1 Kesimpulan.....	45
5.2 Saran.....	45
DAFTAR PUSTAKA	46
LAMPIRAN.....	

DAFTAR GAMBAR

Gambar 3. 1 Tahap Penelitian	15
Gambar 3. 2 Tahap <i>Preprocessing</i>	16
Gambar 3. 3 Alur perancangan sistem	23
Gambar 3. 4 <i>Flowchart</i> sistem	24
Gambar 3. 5 halaman <i>home</i>	25
Gambar 3. 6 halaman <i>dataset</i>	26
Gambar 3. 7 halaman <i>predict</i>	27
Gambar 3. 8 <i>Output</i> dari halaman <i>predict</i>	28
Gambar 4. 1 Tahapan <i>Preprocessing</i>	34
Gambar 4. 2 Tahapan Evaluasi	35
Gambar 4. 3 Tampilan halaman <i>Home</i>	36
Gambar 4. 4 Tampilan halaman <i>dataset</i>	36
Gambar 4. 5 Tampilan fitur <i>download</i>	37
Gambar 4. 6 Tampilan halaman prediksi	38
Gambar 4. 7 Tampilan hasil prediksi	38
Gambar 4. 8 <i>Output</i> hasil nilai <i>confusion matrix</i>	41
Gambar 4. 9 <i>Output</i> hasil (<i>True Positive, True Negative, False Positive, False Negative</i>)	42

DAFTAR TABEL

Tabel 3. 1 Pengumpulan Data	17
Tabel 3. 2 Contoh proses data <i>input</i> dan <i>output</i> pada <i>cleaning</i>	17
Tabel 3. 3 Contoh proses data <i>input</i> dan <i>output</i> pada <i>Case folding</i>	18
Tabel 3. 4 Contoh proses data <i>input</i> dan <i>output</i> pada <i>Tokenizing</i>	18
Tabel 3. 5 Contoh proses data <i>input</i> dan <i>output</i> pada <i>Stopword</i>	19
Tabel 3. 6 Contoh proses data <i>input</i> dan <i>output</i> pada <i>stemming</i>	19
Tabel 4. 1 Sampel Data Bidang <i>Art and Humanities</i>	29
Tabel 4. 2 Sampel Data Bidang <i>Engineering and Technology</i>	30
Tabel 4. 3 Sampel Data Bidang <i>Life Sciences and Medicine</i>	31
Tabel 4. 4 Sampel Data Bidang <i>Natural Sciences</i>	32
Tabel 4. 5 Sampel Data Bidang <i>Sosial Science and Management</i>	32
Tabel 4. 6 analisa dan pengujian.....	39
Tabel 4. 7 Jumlah Data.....	40
Tabel 4. 8 Rincian Data.....	40
Tabel 4. 9 <i>confusion matrix</i>	40
Tabel 4. 10 <i>Confusion Matrix</i>	41
Tabel 4. 11 Contoh Judul yang memiliki topik yang tidak sesuai bidannya.....	43



ABSTRAK

Saat ini publikasi menjadi tuntutan sumber daya manusia untuk menunjukkan kreativitas dan kemampuan sebagai prodak dan buah karya pikir. Publikasi adalah pusat informasi atau berita, baik lama maupun baru, yang disediakan oleh seseorang kepada media massa sesuai dengan peraturan perundang-undangan yang berlaku di negara tersebut. Di Indonesia, publikasi hasil karya ilmiah itu sendiri sangat penting, dan salah satu kegiatan yang berkaitan dengan status peneliti adalah publikasi hasil karya tersebut. Berdasarkan hasil observasi SINTA, diketahui judul-judul yang ada tidak diklasifikasikan menurut lima bidang ilmu yaitu Arts & Humanities, Engineering & Technology, Life Sciences & Medicine, Natural Sciences, and Social Sciences & Management. Tujuan penelitian ini adalah untuk menerapkan dan menguji kinerja algoritma Naive Bayes dalam klasifikasi lima bidang ilmu pada publikasi terindeks Scopus. Naïve Bayes merupakan metode klasifikasi dan prediksi karena hasilnya mudah diinterpretasikan. Metode Naïve Bayes dapat memudahkan peneliti mengklasifikasikan lima bidang ilmu dalam publikasi terindeks Scopus karena modelnya sederhana, mudah dipahami, dan cukup fleksibel untuk meningkatkan kualitas keputusan yang dihasilkan.

Kata Kunci : Klasifikasi, publikasi, Naïve Bayes, SINTA

ABSTRACT

Currently, publication is a human resource guide to show creativity and ability as a product and work of thought. Publication is a center for information or news, old and new, that is provided by someone to the media in accordance with the laws in force in that country. In Indonesia, publishing the results of scientific work itself is very important, and one of the activities related to the status of a researcher is publishing the results of that work. Based on SINTA's observations, it is known that the existing titles are not classified according to five disciplines, namely Arts & Humanities, Engineering & Technology, Life Sciences & Medicine, Natural Sciences, and Social Sciences & Management. The purpose of this study was to apply and test the performance of the Naive Bayes algorithm in the classification of five disciplines in Scopus indexed publications. Naïve Bayes is a method of classification and prediction because the results are easy to interpret. The Naïve Bayes method can make it easier for researchers to classify five fields of knowledge in Scopus-indexed publications because the model is simple, easy to understand, and flexible enough to improve the quality of the resulting decisions.

Keywords: Classification, publication, Naïve Bayes, SINTA

BAB I

PENDAHULUAN

1.1. LATAR BELAKANG

Publikasi saat ini menjadi tuntutan bagi sumber daya manusia untuk menunjukkan kreativitas dan kemampuan sebagai prodak dan buah karya pikir. Publikasi adalah informasi atau berita lama dan baru, yang diberikan seseorang ke khalayak media berdasarkan dengan undang-undang yang berlaku di Negara.

Selain itu, pengertian publikasi menurut kamus besar Bahasa Indonesia menyatakan bahwa media informasi digunakan untuk menyebarluaskan informasi kepada masyarakat. (Mustopa dkk., 2021)

Menurut (Amaliyah, 2019). Jurnal internasional yang telah bereputasi yaitu jurnal yang terindeks Scopus. Sedangkan jurnal nasional merupakan jurnal yang telah terakreditasi oleh Kementerian Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia melalui Science and Technology Index (SINTA) dengan tingkat akreditasi yang tinggi yaitu jurnal SINTA. (Firmansyah dkk., 2020)

Sejauh ini, pencarian jurnal dalam web SINTA belum bisa mengkategorisasikan sesuai dengan 5 bidang keilmuan sehingga kurangnya efektif dalam melakukan pencarian artikel. Untuk itu diperlukan suatu metode klasifikasi semua artikel dengan menggunakan metode naïve bayes.

Scopus merupakan salah satu pengindeks jurnal terbesar di dunia dan banyak digunakan sebagai referensi publikasi penelitian pada jurnal. Scopus memang tidak mudah, harus melalui beberapa tahapan review oleh pakar (Ibrahim, 2019).

Beberapa metode dalam klasifikasi yang umum digunakan adalah KNN, ANN, Decision Tree, Naïve bayes dan lain sebagainya. Naive Bayes Classifier merupakan metode klasifikasi sederhana, bekerja untuk mengklasifikasi teks dengan teknik jaringan bayesian, dimana semua atribut diberi nilai variabel kelas.

Berdasarkan latar belakang di atas penelitian ini, diusulkan metode Naive Bayes untuk mengimplementasikan dan menguji kinerja algoritma Naive Bayes dalam klasifikasi 5 bidang ilmu yaitu: Arts & Humanities, engineering &

Technology, Life Sciences & Medicine, Natural Sciences, and Social Sciences & Management dalam publikasi terindeks Scopus.

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang tersebut, maka rumusan masalah dalam penelitian ini adalah bagaimana menerapkan metode Naïve bayes pada publikasi terindeks Scopus untuk mengklasifikasikan data SINTA ke dalam lima bidang ilmu.

1.3 BATASAN MASALAH

Batasan masalah dalam penelitian ini adalah :

1. Sumber data yang diperoleh dalam penelitian ini dari artikel terindeks scopus yang tersimpan dalam database SINTA.
2. Klasifikasi menggunakan Naive Bayes dibatasi pada lima bidang ilmu dalam penelitian ini.
3. Sumber data yang didapatkan dalam penelitian ini hanya dengan judul publikasi.

1.4 TUJUAN TUGAS AKHIR

Penelitian ini bertujuan untuk merancang sebuah sistem klasifikasi publikasi terindeks Scopus menggunakan algoritma Naïve Bayes.

1.5 MANFAAT

Manfaat yang diharapkan dari sistem ini adalah :

1. Mempermudah pembaca pada saat melakukan pencarian artikel sesuai dengan bidang ilmu.
2. Penerapan metode Naïve Bayes diharapkan mampu mengklasifikasikan data Scopus yang terindeks pada database SINTA sesuai dengan 5 bidang ilmu.

1.6 SISTEMATIKA LAPORAN

Sistematika penulisan yang digunakan oleh penulis dalam laporan tugas akhir adalah :

BAB I : PENDAHULUAN

Bab ini menjelaskan latar belakang, pemilihan judul, rumusan masalah, batasan masalah, tujuan penelitian, metodologi penelitian, dan sistematika penulisan.

BAB II : TINJAUAN PUSTAKA DAN DASAR TEORI

Bab ini menjelaskan tentang penelitian serupa yang telah dilakukan sebelumnya, serta topik yang mendukung mengenai permasalahan yang dibahas dalam Tugas Akhir.

BAB III : METODE PENELITIAN

Bagian bab ini menjelaskan mengenai metode penelitian. Dimana pada bab ini akan dijelaskan metode dan langkah-langkah yang harus dilakukan dalam penyelesaian Tugas Akhir.

BAB IV : HASIL PENELITIAN

Pada bab ini memuat hasil penelitian yang dilakukan untuk menyelesaikan permasalahan dalam tugas akhir, khususnya hasil klasifikasi bidang ilmu pada publikasi terindeks Scopus dengan menggunakan metode Naïve bayes.

BAB V : KESIMPULAN DAN SARAN

Dalam Bab ini, penulis akan memaparkan hasil kesimpulan dan saran disampaikan sebagai pelengkap untuk menyempurnakan tugas akhir.

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Klasifikasi adalah proses pengelompokan data ke dalam kelompok-kelompok kelas yang menggambarkan konsep atau kelas data untuk tujuan tertentu. Klasifikasi berisi informasi tentang bagaimana data dikelompokkan, sistem kemudian di training pada data yang diberi label (dalam kelompok mana data tersebut berada) dan sistem mengklasifikasikan data baru ke dalam kelompok yang ada (klasifikasi). Publikasi adalah bukti utama kebenaran penelitian yang dilakukan. Publikasi dibuat dengan menggunakan media massa untuk menyebarkan informasi sehingga diperoleh informasi secara efektif dan efisien. Terdapat berbagai publikasi antara lain: publikasi konvensional, publikasi elektronik, dan publikasi interpersonal (Kartikasari & Dianti, 2021).

Terdapat beberapa penelitian terdahulu tentang algoritma Naïve bayes yang pernah dilakukan Devita dan kawan-kawan dalam mengklasifikasikan barang atau benda tertentu, termasuk Perbandingan kinerja Metode Naïve Bayes dan K-Nearest Neighbor untuk mengklasifikasikan artikel Bahasa Indonesia. Pada penelitian ini, metode yang digunakan adalah Naive Bayes dan metode K-Nearest Neighbor. Metode Naive Bayes dipilih karena dapat memberikan akurasi yang tertinggi dengan data training paling sedikit. Pada saat yang sama metode KNN terpilih karena metode tersebut tangguh terhadap data noise. Performa kedua metode tersebut dibandingkan untuk mengetahui metode mana yang terbaik untuk mengklasifikasi dokumen. Berdasarkan penelitian ini diperoleh hasil bahwa metode Naive Bayes mempunyai efisiensi yang lebih baik dengan akurasi sebesar 70%, sedangkan metode K-Nearest Neighbor memiliki tingkat akurasi yang cukup rendah yaitu sebesar 40% (Devita dkk., 2018).

Dalam penelitian sebelumnya oleh annur tahun 2018 tentang klasifikasi penduduk miskin dengan menggunakan metode naïve bayes. Naïve bayes adalah metode yang digunakan dalam penelitian ini, yang merupakan salah satu teknik klasifikasi dalam data mining. Berdasarkan penelitian ini dapat disimpulkan

bahwa sistem klasifikasi masyarakat miskin di Kecamatan Tibawa Gorontalo dapat dibangun, dan berdasarkan hasil matriks konfusi yang dilakukan dengan metode split-validation, dengan menggunakan metode klasifikasi Bayesian, yang diuji dataset mencapai akurasi 73% dan tergolong dalam kategori baik. Nilai presisi dan perolehan juga masing-masing 92% dan 86% (Annur, 2018).

Pada penelitian sebelumnya oleh putra & putri tahun 2022 tentang Klasifikasi Jurusan Siswa Kelas XI menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor. Pada penelitian ini metode yang digunakan yaitu Algoritma Naive Bayes dan K-Nearest Neighbor. Metodologi penelitian dimulai dengan tahap pre-processing, training data, klasifikasi menggunakan algoritma Naïve Bayes dan K-Nearest Neighbor, model data, dan pengukuran kinerja atau evaluasi. Berdasarkan penelitian ini memperoleh hasil bahwa akurasi algoritma Naïve Bayes sebesar 81.82% untuk 55 sampel data dari 277 sampel. Algoritma K-Nearest Neighbor menghasilkan akurasi 92.73% pada sampel data yang sama. Hasil dari kedua algoritma tersebut menunjukkan algoritma terbaik adalah K-Nearest Neighbor dan kedua Naïve Bayes, nilai alpha kurang dari 0,05, jadi tidak banyak perbedaan. Secara umum dapat disimpulkan bahwa penggunaan algoritma K-Nearest Neighbor dibandingkan dengan Algoritma Naïve Bayes menunjukkan akurasi yang lebih tinggi dalam proses klasifikasi penjurusan IPA maupun IPS pada kelas XI SMA. Menentukan menggunakan algoritma klasifikasi yang sesuai dengan minat, kemampuan, dan potensi siswa menawarkan keuntungan dan membantu sekolah membuat klasifikasi lebih cepat (Putra & Putri, 2022).

Dalam penelitian sebelumnya oleh Ericha Apriliyani & Salim pada tahun 2022 tentang menganalisis kinerja pengklasifikasi Naïve Bayes Classifier pada Unbalanced Dataset. Naïve bayes classifier digunakan dalam penelitian ini. Berdasarkan hasil analisis dengan metode Naïve Bayes Classifier, dataset Glass mendapatkan akurasi 46%, presisi 47%, recall 46% dan f-measure 43%, dalam dataset Heart Disease diperoleh akurasi 88%, presisi 88%, recall 88% dan f-measure 88%, dalam dataset Kidney Disease diperoleh akurasi 100%, presisi 100%, recall 100% dan f-measure 100%, dalam dataset Liver Disease diperoleh akurasi 78%, presisi 82%, recall 78% dan f-measure 79%, dalam dataset Diabetes

diperoleh akurasi 77%, presisi 76%, recall 77% dan f-measure 76%, dan dalam dataset Breast Cancer diperoleh akurasi 94%, presisi 94%, recall 94% dan f-measure 94%. Hasil penelitian ini menunjukkan bahwa metode Naive Bayes Classifier menghasilkan nilai kinerja yang tidak valid untuk dataset yang tidak seimbang (Ericha Apriliyani & Salim, 2022).

Dalam penelitian sebelumnya oleh Utami dan kawan-kawan pada tahun 2022 tentang Metode weighted naive bayes dengan Laplace smoothing digunakan untuk mengklasifikasikan kelayakan penerima bantuan Program Keluarga Harapan (PKH). Dalam penelitian ini metode yang akan digunakan adalah naïve bayes. Pada penelitian tersebut dapat memberikan prediksi tentang klasifikasi kelayakan masyarakat penerima bantuan PKH, sehingga perangkat desa dapat menentukan siapa yang berhak mendapatkan bantuan dan memberikan hasil yang lebih cepat, akurat, dan tepat sasaran bagi penerima. Untuk mengklasifikasikan kelayakan penerima manfaat PKH pada penelitian ini digunakan Metode Weighted Naïve Bayes dengan Laplace smoothing. Dengan menambahkan bobot pada atribut kelas dari algoritma naïve bayes, akurasi klasifikasi pada algoritma naïve bayes berbobot tidak hanya didasarkan pada probabilitas tetapi juga pada bobot fitur kelas, yang ditambahkan ke atribut kelas dalam algoritma naïve bayes. Dapat menghindari nilai probabilitas 0 dengan menggunakan laplace smoothing. Berdasarkan hasil perhitungan klasifikasi, dengan menggunakan metode naïve bayes laplace smoothing, dengan 56 data training dan 24 data uji, hasil uji performa dengan menggunakan confusion matrix menunjukkan akurasi sebesar 95,83%, tingkat kesalahan 4,17%, Sensitivitas 100,00%, dan Spesifitas 94,12% (Utami dkk., 2022).

Dalam penelitian sebelumnya oleh Hozairi dan kawan-kawan 2021 tentang penerapan orange data mining digunakan untuk klasifikasi kelulusan mahasiswa dengan model k-nearest neighbor, decision tree, dan naïve bayes. Metode K-nearest Neighbor, Pohon keputusan, dan Naïve Bayes digunakan dalam penelitian ini. Penelitian dilakukan pada tahun 2016 sebagai bagian dari program studi Teknik Informatika Universitas Islam Madura, data mahasiswa kemudian dievaluasi menggunakan orange data mining dengan metode K-NN, Pohon

Keputusan, serta Naive Bayes. K-Fold Cross digunakan dalam proses uji validasi data ($K=5$), dan model evaluasi yang digunakan adalah confusion matrix dan ROC. Hasil perbandingan ketiga menggunakan metode K-NN menunjukkan akurasi 77%, akurasi Pohon Keputusan 74%, dan akurasi Naive Bayes 89%. Oleh karena itu, disarankan menggunakan metode Naive Bayes yang lebih akurat dibandingkan dengan K-NN dan Pohon Keputusan untuk klasifikasi tingkat kelulusan mahasiswa Prodi Teknik Informatika di Universitas Islam Madura (Hozairi dkk., 2021).

Dalam penelitian sebelumnya oleh Yuwono dan kawan-kawan tahun 2021 tentang metode Naive Bayes digunakan untuk mengklasifikasi penghasilan pedagang kaki lima dan bisnis online yang terkena dampak pandemi COVID-19. Metode yang digunakan dalam penelitian ini adalah Naive Bayes, dimana Naive Bayes merupakan salah satu metode klasifikasi dalam data mining. Menurut sebuah penelitian, pedagang kaki lima dan bisnis online mengalami penurunan penghasilan sebesar 89%. Selain itu, hasil pengujian berdasarkan confusion matrix menggunakan teknik split validasi dengan metode naive bayes. Akurasi yang dicapai untuk dataset yang telah diambil sebagai objek penelitian adalah 75% atau termasuk dalam kategori yang sesuai. Sedangkan nilai Precision sebesar 100% dan Recall sebesar 66.7% (Yuwono dkk., 2021).

Dalam penelitian sebelumnya oleh Waliyansyah & Fitriyah tahun 2019 tentang Metode Naive Bayes dan k-nearest-neighbor (KNN) yang digunakan untuk membandingkan Akurasi Klasifikasi Citra Kayu Jati. Terdapat 3 jenis kayu jati dalam penelitian ini yaitu Semarang, Blora dan Sulawesi. Metode Naive Bayes dan k-Nearest Neighbor (k-NN) digunakan dalam Proses klasifikasi Jati. Analisis level GLCM menggunakan metode tekstur yang dibuat dengan model Gray, dengan jarak spasial 1 piksel. Secara umum berdasarkan hasil pengujian dan analisis, metode k-NN dapat mengklasifikasikan tiga jenis kayu jati yaitu Semarang, Blora dan Sulawesi dengan akurasi lebih dari 70%. Namun klasifikasi terbaik untuk Jati Sulawesi adalah metode Naive Bayes yang memiliki akurasi sebesar 82,7% (Waliyansyah & Fitriyah, 2019).

Dalam penelitian sebelumnya pada tahun 2019 oleh Nanang Ruhyana tentang metode klasifikasi Naïve Bayes digunakan untuk Analisis sentimen terhadap implementasi sistem plat nomor ganjil/genap di twitter. Dalam penelitian ini, kami menggunakan teknik data mining untuk menentukan peringkat menggunakan algoritma pemeringkatan Naive Bayes. Naive Bayes Classifier (NBC) adalah teknik pengelompokan data Bayesian probabilistik. Tingkat akurasi penelitian ini adalah 86,67%, presisi adalah 71,43%. dan recall adalah 80,00% (Ruhyana, 2019).

Dalam penelitian sebelumnya oleh Prakoso dan kawan-kawan tahun 2019 tentang Algoritma Naïve Bayes Dengan Seleksi Fitur Dan Boosting digunakan untuk mengklasifikasi Berita. Hasil evaluasi nilai akurasi, recall, dan presisi diperoleh 73.2%. Namun, dalam metode Naive Bayes Classifier, pengklasifikasi Bayesian Boosting memiliki skor yang sama yaitu 73.2%. Hasil evaluasi model implementasi menunjukkan bahwa penerapan Information Gain feature selection tidak berpengaruh signifikan terhadap peningkatan kinerja pada kondisi label Polynomial (Prakoso dkk., 2019).

Dalam penelitian sebelumnya oleh Rohmat Indra Borman & Mina Wati tahun 2020 tentang Algoritma Naïve Bayes digunakan untuk mengimplementasikan Data Mining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandar Lampung. Dimana naïve Bayes adalah sebuah metode pada probabilistic reasoning. Dari hasil pengujian WEKA dan Rapidminer memperoleh akurasi 70,33%, recall 70,33%, dan presisi 100%, atau 1, menggunakan 1064 catatan pelatihan dan 300 catatan pengujian. Semakin banyak data uji yang digunakan, semakin besar dampaknya terhadap akurasi (Indra Borman & Wati, 2020).

Selanjutnya, pada penelitian sebelumnya tahun 2018 oleh Adi tentang algoritma naïve bayes classifier digunakan untuk penerapan klasifikasi penerima beasiswa PPA di Universitas Amikom Yogyakarta. Naïve Bayes merupakan metode yang digunakan dalam penelitian ini. Penelitian ini membuat model klasifikasi probabilistik untuk mengidentifikasi kategori calon pendaftar beasiswa. Hasil pengujian akurasi model sistem yang dikembangkan menunjukkan bahwa akurasi minimum adalah 64% saat diuji dengan 100 sampel data, dan akurasi

output tertinggi adalah 97,66% saat diuji dengan sampel sebanyak 386 data (Adi, 2018).

Dalam penelitian sebelumnya oleh Imandasari dan kawan-kawan tahun 2019 tentang Algoritma Naive Bayes Dalam Klasifikasi Lokasi Pembangunan Sumber Air. Dalam penelitian ini khususnya dengan metode data mining klasifikasi menggunakan algoritma Naive Bayes dapat diprediksi kelayakan lokasi pembangunan sumber air bersih berdasarkan data yang ada. Berdasarkan hasil perhitungan yang dilakukan dengan menggunakan algoritma Naive Bayes, diperoleh hasil klasifikasi dari 19 alternatif yang digunakan menunjukkan 8 kelas memenuhi persyaratan dan 11 kelas yang gagal, dengan akurasi keseluruhan sebesar 78,95%. Berdasarkan hasil yang diperoleh, diharapkan dapat membantu PDAM Tirta Lihou menemukan lokasi yang cocok untuk mengembangkan sumber air untuk memenuhi kebutuhan masyarakat. Penelitian ini juga bertujuan untuk menginformasikan peneliti masa depan tentang pengguna algoritma yang digunakan (Imandasari dkk., 2019).

Dalam penelitian sebelumnya oleh Siddik dan kawan-kawan tahun 2020 tentang Penggunaan Algoritma Naive Bayes dalam Klasifikasi kepuasan mahasiswa terhadap pelayanan perguruan tinggi. Naive Bayes Classifier adalah metode data mining yang digunakan dalam penelitian untuk mengklasifikasi data. Metode naive bayes memiliki nilai kinerja yang sangat bagus. Keuntungan dari metode naive bayes adalah sangat sederhana, mudah dan cepat digunakan. Total ada 213 sumber data mahasiswa di perguruan tinggi di Riau. Hasil pengujian Naive Bayes diperoleh tingkat akurasi sebesar 96,24%, presisi sebesar 93,14 %, dan recall sebesar 98,96% (Siddik dkk., 2020).

Dalam penelitian sebelumnya oleh Fani Prasetya & Ferdiansyah tahun 2022 tentang penggunaan Algoritma Naive Bayes untuk Analisis Data Mining Klasifikasi Berita Hoax COVID 19. Naive Bayes merupakan metode yang digunakan dalam penelitian ini. Berdasarkan hasil penelitian menggunakan metode naive bayes dan cross validation mengklasifikasikan berita hoax dengan benar, hasil akurasi sebesar 86.3% dengan 80-90% termasuk dalam kriteria bagus.

Tidak banyak data yang salah diprediksi karena dalam 300 kumpulan data hanya 41 yang tidak diberi label dari 2% dari total kumpulan data, sehingga dapat disimpulkan bahwa model ini dapat digunakan untuk referensi apabila ingin dilanjutkan pada model prediksi yang lebih kompleks lagi, misalnya model prediksi menggunakan machine learning berbasis web (Prasetya & Ferdiansyah, 2022).

Dalam penelitian sebelumnya oleh Wahyugi Fadri tahun 2023 tentang Metode Naive Bayes Dalam Klasifikasi Penyakit Hati. Metode Naive Bayes digunakan dalam penelitian ini karena metode yang memiliki nilai akurasi cukup tinggi. Berdasarkan hasil pengujian dari 50 data rekam medis pasien penyakit hati, dengan penerapan teknik Naive Bayes dalam data mining, bermanfaat sebagai perbandingan data rumah sakit dengan data sistem. Tingkat akurasi dari 45 data training dan 5 data uji adalah 60% (Fadri, 2023).

Dalam penelitian sebelumnya oleh Dyah Ariyanti & Kurnia Iswardani tahun 2020 tentang Penggunaan Algoritma Naive Bayes Untuk Teks Mining untuk Klasifikasi Keluhan Masyarakat. Penelitian tersebut 95% akurat, memungkinkan Dewan Kota Probolinggo untuk mengklasifikasikan pengaduan masyarakat (Ariyanti & Iswardani, 2020).

Dalam penelitian sebelumnya pada tahun 2019 oleh Agus Sugianto & Rizky Maulana tentang Klasifikasi Penerima Bantuan Pangan Non Tunai (Studi Kasus Kelurahan Utama) Menggunakan Algoritma Naive Bayes. Metode Naive Bayes dan Pohon Keputusan digunakan dalam penelitian ini. Melakukan klasifikasi penerima dan penerima bahan pokok non pangan dalam penelitian ini Teknik klasifikasi data mining seperti Naive Bayes dan algoritma pohon keputusan digunakan untuk perbandingan. Model data mining yang didukung oleh RapidMiner menghasilkan nilai probabilitas untuk kelas penerima yaitu (0,481, dibulatkan menjadi 0,48) dan nilai probabilitas untuk kelas non-penerima (0,519, dibulatkan menjadi 0,52). Tingkat akurasi algoritma Naive Bayesian sebesar 58,29%, tingkat presisi sebesar 92,90%, tingkat recall sebesar 21,84%, AUC sebesar 0,765, dan F-measure sebesar 34,42%. Sementara tingkat Akurasi

algoritma pohon keputusan 73,97%, presisi 85,04%, recall 61,92%, AUC 0,746, dan F-measure 71,17%. Dengan menggunakan alpha kurang dari 0,000 untuk hasil uji-t antara algoritma Naive Bayes dan algoritma Decision Tree, dapat disimpulkan bahwa uji-t antara algoritma Naive Bayes dan keputusan algoritma Pohon Keputusan memiliki hasil yang signifikan (Agus Sugianto & Rizky Maulana, 2019).

Dalam penelitian sebelumnya oleh Antonio Ciputra dan kawan-kawan tahun 2018 tentang penggunaan Algoritma Naive Bayes dan citra digital untuk mengklasifikasi tingkat kematangan buah. Jika data yang digunakan terdiri dari 130 citra (100 citra latih dan 30 citra uji), maka hasil penelitian ini memiliki akurasi sebesar 63% (Ciputra dkk., 2018).



2.2 Dasar Teori

2.2.1 Data Mining

Data mining adalah proses mengekstraksi dan mengidentifikasi informasi yang berguna dan relevan dari basis data besar menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning. (Utomo & Mesran, 2020). Di dalam data mining terdapat klasifikasi.

2.2.2 Klasifikasi

Klasifikasi adalah proses menemukan sekumpulan model (fitur) yang dapat menggambarkan kelas data atau konsep dan membedakannya sehingga dapat digunakan untuk memprediksi kelas dari objek yang tidak diketahui. (Nasrullah, 2021). Naïve Bayes merupakan salah satu metode Klasifikasi.

2.2.3 Naïve Bayes

2.2.3.1 Pengertian Naïve Bayes

Pengklasifikasi Naive Bayes Classifier adalah proses klasifikasi probabilistik sederhana yang terkait dengan teori Bayesian. Menurut teori ini, probabilitas terjadinya suatu peristiwa sama dengan probabilitas internal (dihitung dari informasi yang tersedia saat ini) dikalikan dengan probabilitas peristiwa yang sama terjadi lagi di masa depan (berdasarkan peristiwa di masa lalu). Naïve bayes adalah algoritma pembelajaran probabilitas yang berasal dari teori keputusan Bayesian. Algoritma ini memakai teorema Bayes & mengasumsikan bahwa seluruh atribut tidak tergantung dalam nilai variabel kelas. Dalam penggunaan memiliki keuntungan menggunakan metode Naïve bayes.

2.2.3.2 Keuntungan Metode Naïve Bayes

Keuntungan dalam menggunakan Metode Naïve Bayes sebagai berikut :

1. Interpolation: Metode bayesian memungkinkan pemilihan tentang berapa waktu dan upaya yang dihabiskan manusia vs komputer.
2. Bahasa: Metode bayesian memiliki bahasanya sendiri untuk menentukan probabilitas sebelum dan sesudah.
3. Intuisi: Mencakup dua aktivitas yang sangat berguna prior dan integrasi.

2.2.3.3 Rumus Naïve Bayes

Dalam teorema bayes Rumus Naïve Bayes Classifier adalah sebagai berikut:

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

X = data dokumen dengan kategori tidak diketahui.

H = hipotesis dari data X merupakan kelas tertentu.

$P(H | X)$ = probabilitas keadaan X yang diberikan dari hipotesis H (posteriori probability).

$P(H)$ = probabilitas dari hipotesis H (prior probability).

$P(X | H)$ = Probabilitas X berdasarkan hipotesis H.

$P(X)$ = Probabilitas dari X

2.2.4 Scopus

Scopus adalah salah satu database literatur atau kutipan ilmiah terkemuka di dunia, yang dimiliki oleh Elsevier, salah satu penerbit terkemuka dunia. Scopus didirikan pada tahun 2004. Scopus bukan satu-satunya database literatur ilmiah terkemuka. Basis data scopus berisi lima jenis bidang ilmu yang berbeda.

2.2.5 Macam-macam 5 bidang ilmu

Ada beberapa referensi pembagian bidang ilmu publikasi diantaranya adalah QS RANKING, organisasi yang pemeringkatan universitas di seluruh dunia berdasarkan lima bidang ilmu yaitu :

a) *Arts & Humanities*

Arts & humanities dianggap sebagai dua bidang pengetahuan manusia tertua. Humanities adalah serangkaian ilmu yang menginginkan manusia menjadi lebih manusiawi. Pada hakikatnya Humanities atau ilmu budaya adalah ilmu yang membahas tentang nilai-nilai kemanusiaan, termasuk studi agama, filsafat, seni, sejarah, dan linguistik.

b) *Engineering & Technology*

Teknologi sangat erat kaitannya dengan dunia ilmu pengetahuan dan teknologi. Teknologi pada dasarnya merupakan ilmu yang dapat mempermudah kehidupan manusia. Teknologi juga sering disebut dengan rekayasa yang merupakan penerapan ilmu pengetahuan dan teknologi untuk memecahkan masalah manusia.

c) *Life Sciences & Medicine*

Life sciences & medicine merupakan bidang ilmu yang berhubungan dengan penelitian ilmiah tentang kehidupan mikroorganisme, tumbuhan dan hewan termasuk manusia. Ilmu ini adalah salah satu dari dua bidang ilmu alam, yang lainnya adalah fisika yang berurusan dengan benda mati.

d) *Natural Sciences*

Natural sciences atau ilmu alam adalah cabang ilmu yang berkaitan dengan deskripsi pemahaman dan prediksi fenomena dalam berdasarkan bukti empiris dari pengamatan dan percobaan. Mekanismenya seperti peer review dan temuan berulang berusaha untuk memastikan kecukupan kemajuan ilmiah.

e) *Social Sciences & Management*

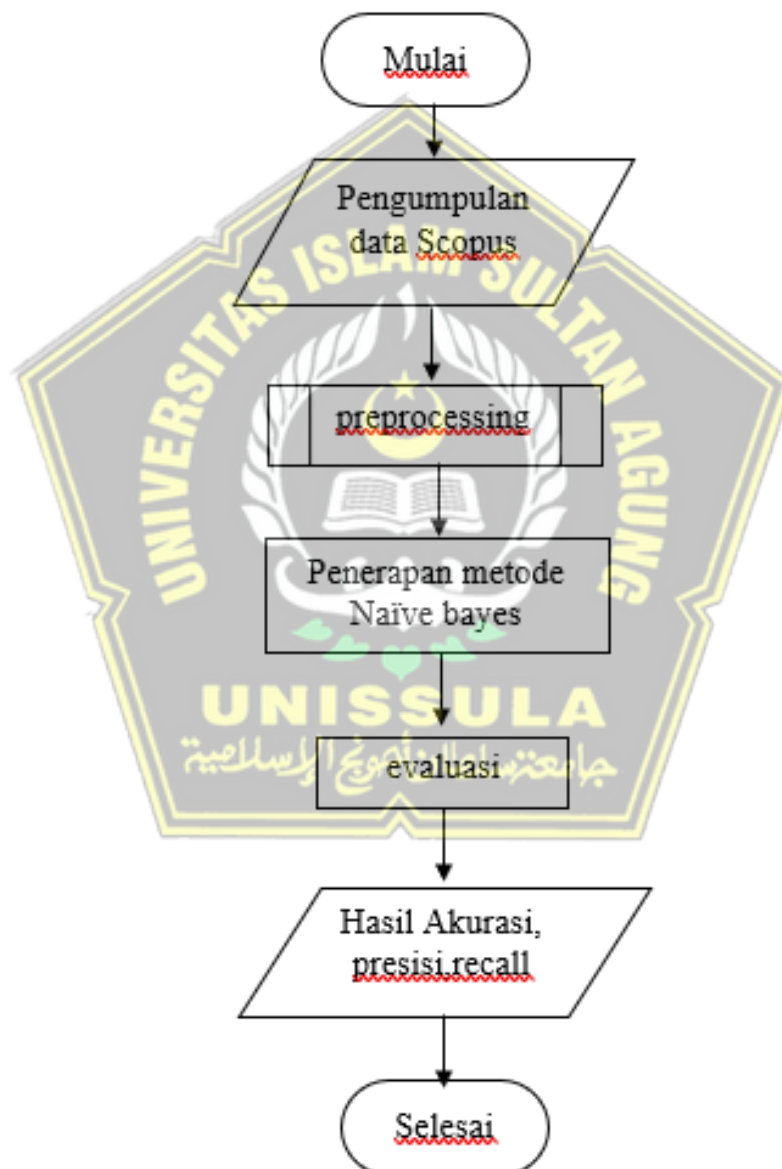
Ilmu sosial adalah sekelompok bidang ilmu yang berhubungan dengan aspek manusia dan lingkungan sosialnya. Sedangkan ilmu manajemen adalah proses dimana orang dapat mengelola apa yang individu atau kelompok lakukan.

BAB III

METODE PENELITIAN

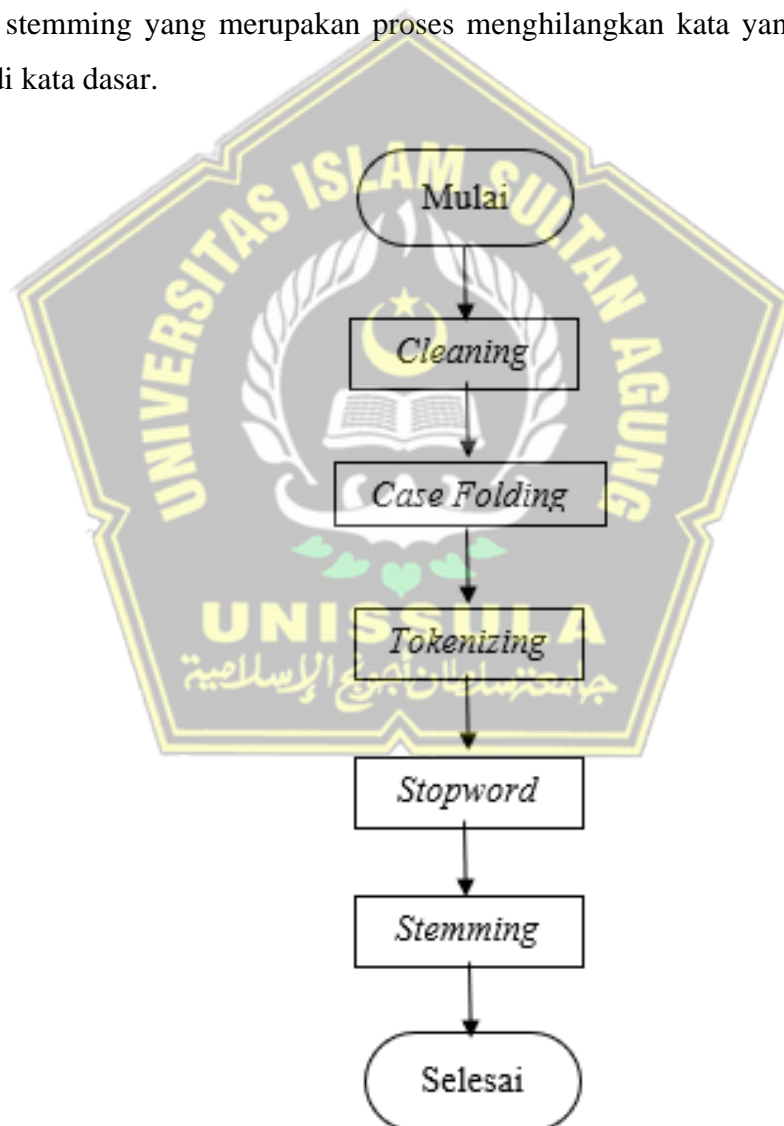
3.1 Metode Penelitian

Tahapan yang akan digunakan dalam penelitian tugas akhir ini ditunjukkan dalam flowchart gambar 3.1 dimulai dari langkah pertama yaitu pengumpulan data Scopus, *pre-processing*, penerapan metode Naïve Bayes, serta evaluasi.



Gambar 3. 1 Tahap Penelitian

Gambar 3.2 menunjukkan langkah preprocessing. Pada langkah preprocessing meliputi berbagai proses, seperti *Cleaning*, *Case folding*, *Tokenizing*, *Stopword* dan *Stemming*. Untuk langkah yang pertama *Cleaning* yaitu proses menghapus tanda baca pada teks contoh titik (.) , koma (,) , tanda petik (“), tanda kurung (), dan angka. Lalu tahap *Case folding* merupakan proses untuk mengganti huruf besar ke huruf kecil (*lowercase*). kemudian *tokenizing* merupakan proses pemisah kumpulan huruf menjadi sebuah kata. kemudian *stopword* adalah teknik menghilangkan kata sambung. Selanjutnya yang terakhir adalah *stemming* yang merupakan proses menghilangkan kata yang berimbuhan menjadi kata dasar.



Gambar 3. 2 Tahap *Preprocessing*

3.1.1 Pengumpulan data scopus

Tujuan dari langkah ini adalah memilih kumpulan data untuk digunakan dalam penelitian ini. Data yang dipilih harus sesuai dengan batasan yang ditetapkan dalam penelitian ini. Pengumpulan data Scopus pada penelitian ini diambil sesuai dengan 5 bidang ilmu yaitu Arts & Humanities, engineering & Technology, Life Sciences & Medicine, Natural Sciences, dan Social Sciences & Management. Dalam penelitian ini, sejumlah 1000 judul publikasi dari lima bidang ilmu berbeda diuji.

Tabel 3.1 Pengumpulan Data

No	Data
1.	Data Capital: A Systematic Literature Review
2.	Child Sex Tourism: A Case Study in Surabaya, Indonesia

3.1.2 Pre-processing

Merupakan tahap dimana terjadinya proses seleksi data agar data yang diambil lebih terstruktur dalam setiap dokumen. Tahapan *pre-processing* untuk menentukan data dengan nilai missing value yang berarti data yang diterima belum sesuai. *Pre-processing* teknik awalan data mining saat mengubah data yang dikumpulkan dari berbagai sumber menjadi data yang lebih bersih dan dapat digunakan untuk pemrosesan data berkelanjutan. Tahap preprocessing dalam penelitian ini yaitu pengkodean data pada judul artikel di SINTA yang terindeks Scopus. Tahapan-tahapan berikut ini dilakukan selama tahapan data preprocessing:

1. *Cleaning* merupakan proses memperbaiki atau menghapus karakter yang bukan huruf baca.

Tabel 3.2 Contoh proses data *input* dan *output* pada *cleaning*

No.	Data Input	Data Output
1.	Data Capital: A Systematic Literature Review	Data Capital A Systematic Literature Review

2.	Child Sex Tourism: A Case Study in Surabaya, Indonesia	Child Sex Tourism A Case Study in Surabaya Indonesia
----	--	--

2. *Case folding*, adalah proses mengubah semua huruf besar pada data menjadi huruf kecil (lowercase).

Tabel 3.3 Contoh proses data *input* dan *output* pada *Case folding*

No.	Data <i>Input</i>	Data <i>Output</i>
1.	Data Capital A Systematic Literature Review	data capital a systematic literature review
2.	Child Sex Tourism A Case Study in Surabaya Indonesia	child sex tourism case study in surabaya indonesia

3. *Tokenizing*, yaitu proses pemisahan sebuah kalimat menjadi satu kata dan lazimnya dipisahkan oleh karakter seperti tanda baca spasi, titik (.), dan koma (,).

Tabel 3.4 Contoh proses data *input* dan *output* pada *Tokenizing*

No.	Data <i>Input</i>	Data <i>Output</i>
1.	data capital a systematic literature review	data / capital / a / systematic / literature / review
2.	child sex tourism case study in surabaya indonesia	child / sex / tourism / case / study / in / surabaya / indonesia

4. *Stopword*, yaitu tahap pemilihan kata yang bernilai dari hasil token dengan menyimpan kata yang bernilai dan membuang kata yang tidak bernilai. Dapat menggunakan algoritma stoplist seperti (yang), (dan), (di), dan lain-lain.

Tabel 3.5 Contoh proses data *input* dan *output* pada *Stopword*

No.	Data <i>Input</i>	Data <i>Output</i>
1.	data capital a systematic literature review	data capital systematic literature review
2.	child sex tourism a case study in surabaya indonesia	child sex tourism case study surabaya indonesia

5. *Stemming*

Setelah melalui proses *stopword removal*, tahap selanjutnya adalah proses *stemming*. *Stemming* merupakan proses pemetaan dan penguraian berbagai bentuk (variasi) kata menjadi bentuk kata dasar.

Tabel 3.6 Contoh proses data *input* dan *output* pada *stemming*

No.	Data <i>Input</i>	Data <i>Output</i>
1.	data capital systematic literature review	data capital systematic literature review
2.	child sex tourism case study surabaya indonesia	child sex tourism case study surabaya indonesia

3.1.3 Penerapan metode

Data set diolah menggunakan metode Naïve Bayes. Pada penelitian ini mengimplementasikan metode naïve bayes. Naïve bayes adalah teknik klasifikasi data mining yang menerapkan teori Bayes. Kelebihannya adalah modelnya sederhana, cepat dan akurasi yang tinggi. Teorema keputusan Bayes adalah pendekatan statistik dasar untuk pengenalan pola. Naïve Bayes didasarkan pada asumsi penyederhanaan bahwa nilai output dan atribut tertentu secara kondisional independen. Artinya, nilai output yang diberikan, mengamati secara bersama adalah produk dari probabilitas masing-masing. Dengan mengganti Persamaan 1 menjadi Persamaan 2 memberikan pendekatan yang digunakan oleh NBC. Teorema Bayes didirikan oleh Thomas Bayes pada tahun 1763.

Teorema Bayes digunakan untuk menghitung probabilitas terjadinya suatu peristiwa yang terjadi berdasarkan pengaruh yang diperoleh dari hasil pengamatan. Perbedaan mendasar antara metode Bayesian dan statistik umumnya adalah bahwa dalam Bayesian parameter dianggap sebagai variabel acak sementara dalam statistik klasik, parameternya tidak diketahui. Nama teorema Bayes berasal dari Pendeta Thomas Bayes, yang menggambarkan hubungan antara probabilitas bersyarat dari dua peristiwa, H dan X.

3.1.4 Evaluasi

Tahap ini mencakup hasil eksperimen berdasarkan recall presisi dan akurasi. Tahapan ini digunakan untuk mendapatkan prediksi menggunakan algoritma yang sudah ditentukan. Tahapan ini merupakan langkah penting dalam proses membangun sebuah model. Menggunakan persamaan berikut untuk membuktikan kinerja algoritma yang digunakan:

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100 \% \quad (2)$$

Hasil akurasi dihitung berdasarkan confusion matrix. Perhitungan pada confusion matrix dilakukan berdasarkan prediksi positif (true positive), prediksi positif yang salah (false positive), prediksi negatif yang benar (True Negatif) dan prediksi negatif yang salah (False Negatif). Semakin tinggi nilai akurasi yang diperoleh maka semakin bagus metode yang dihasilkan.

Precision adalah tingkat ketelitian antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh system.

$$Precision = \frac{TP}{TP+FP} \times 100 \% \quad (3)$$

Recall adalah tingkat keberhasilan sistem dalam mengambil sebuah informasi.

$$Recall = \frac{TP}{TP+FN} \times 100 \% \quad (4)$$

F1 Score adalah perbandingan rata-rata presisi dan recall yang dibitikan.

$$F1 - score = 2x \frac{precision \times recall}{(Precision + Recall)} \quad (5)$$

Setelah hasil klasifikasi diperoleh, Cross Validation (confusion matrix) digunakan untuk mengevaluasi hasil akurasi, presisi dan recall yang dihasilkan oleh model yang diusulkan.

3.2 Metodologi Perancangan Alur Sistem

3.2.1 Analisis Kebutuhan

Untuk mengetahui Analisis kebutuhan, maka diperlukan untuk melakukan beberapa tahapan yaitu input judul, proses klasifikasi, melakukan prediksi, dan menampilkan hasil prediksi. Tahapan ini menganalisa tentang apa saja yang akan dilakukan oleh system, dalam melakukan proses input sampai dengan mengeluarkan hasil dari sebuah klasifikasi yang dilakukan oleh sistem ini, dan setidaknya ada beberapa proses atau fungsi yang harus ada pada sistem ini, meliputi:

A. Input Judul

Hal yang pertama sistem lakukan, yaitu sistem akan meminta masukan/input judul. Karena di dalam sistem ini membutuhkan sebuah judul yang dapat digunakan untuk melakukan klasifikasi, sebelum melakukan proses melakukan langkah selanjutnya, untuk melakukan input judul, sistem ini sudah menyimpan dataset publikasi judul.

B. Proses klasifikasi

Proses klasifikasi adalah fungsi kedua yang dimiliki oleh sistem. Karena maksud proses klasifikasi yaitu akan memprediksi suatu kategori dari input data sebelum proses prediksi dilakukan. Dimana sistem dapat mengolah judul publikasi menggunakan metode Naïve Bayes.

C. Melakukan prediksi

Pada fungsi ini, sistem bertugas melakukan prediksi terhadap judul-judul publikasi yang terkumpul. Sistem ini mencari tetangga terdekat dari prediksi judul publikasi, dengan menggunakan metode Naïve Bayes terdekat untuk mencari tetangga terdekat.

D. Menampilkan hasil prediksi

Dan fungsi yang terakhir dari sistem ini adalah untuk menampilkan hasil prediksi yang telah dibuat sebelumnya, fungsi menampilkan hasil prediksi bertujuan agar

pengguna dapat melihat hasil perhitungan yang dilakukan oleh sistem. Pengguna fungsi ini dapat melihat persentase tetangga terdekat dari setiap kelas yang memiliki 5 bidang informasi yang dapat diprediksi.

3.2.2 Implementasi Sistem

Pada tahap analisis sistem, akan menganalisa tools apa yang digunakan untuk membuat sistem dan dibawah ini adalah beberapa tools yang digunakan dalam pembuatan sistem :

A. Visual Studio Code

Di dalam Visual Studio Code mengambil fungsi code editor yang dikembangkan secara open-source. Visual studio code mendukung berbagai Bahasa pemrograman salah satu nya PHP. Pada code editor ini juga memiliki banyak fitur serta mendukung penggunaan diberbagai platform. Mulai dari MacOS, Linux, hingga Windows

B. Rubixml

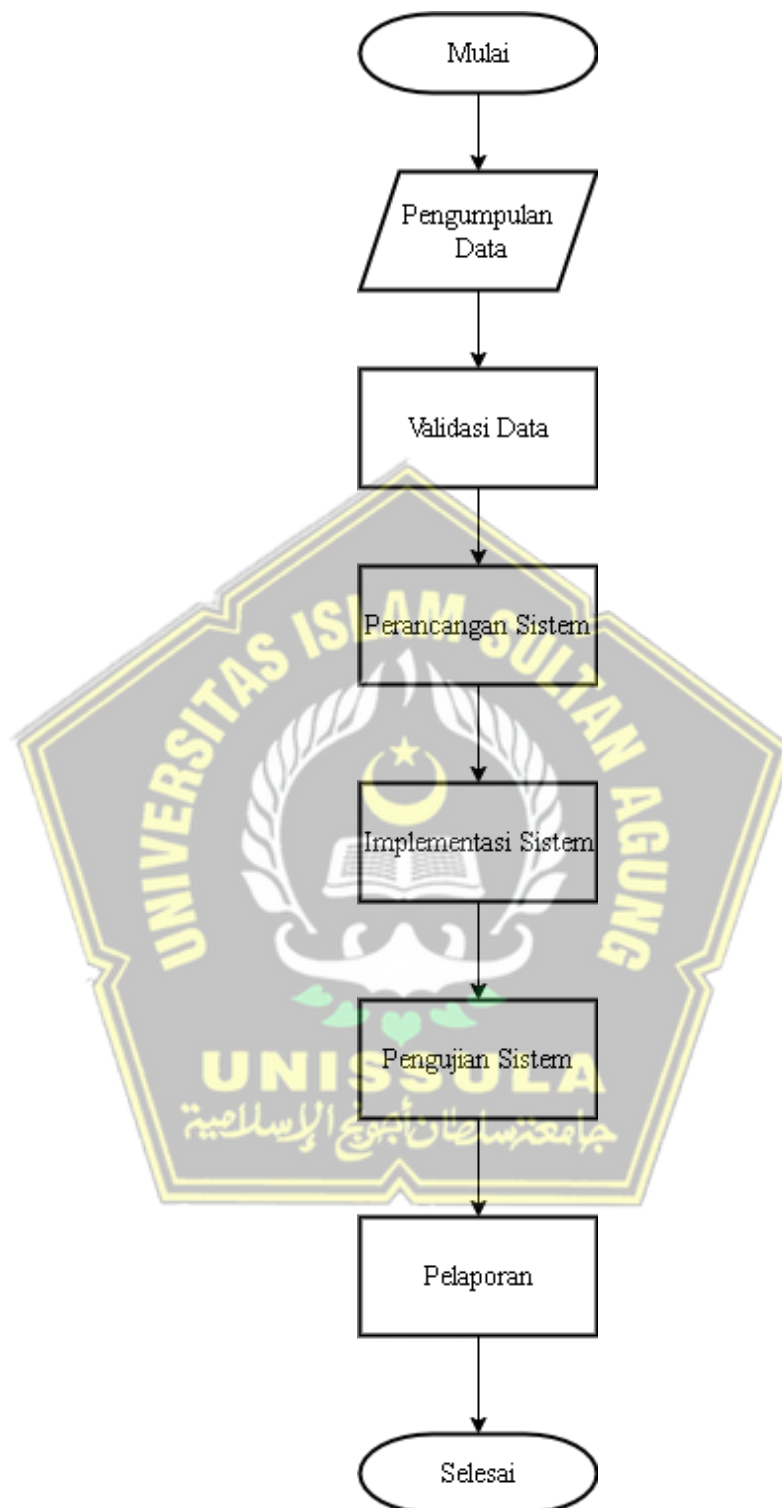
Library Rubixml menyediakan fungsi untuk membangun program dan sistem pembelajaran. Library ini menyediakan berbagai alat untuk membangun sistem, mulai dari pembacaan data, preprocessing, model training hingga testing. Dengan library Rubixml, pengguna bisa mempersingkat script atau Programkan kode tanpa membuat fungsi dan algoritmanya sendiri. pengguna cukup menggunakan function-function atau mungkin object yang sudah terdapat di dalam library.

C. Bootstrap

Fungsi dari bootstrap untuk membuat desain website. Semua situs web yang dibangun dengan menggunakan framework ini dengan cepat beradaptasi sesuai ukuran layar browser. Bootstrap juga menawarkan peluang untuk menambahkan CSS dan class. Sehingga, memungkinkan developer untuk membuat situs web yang lebih fleksibel. Framework ini juga bisa membantu menyesuaikan pengaturan layout dalam sebuah halaman dengan fitur yang disebut grid.

3.2.3 Alur Implementasi Sistem

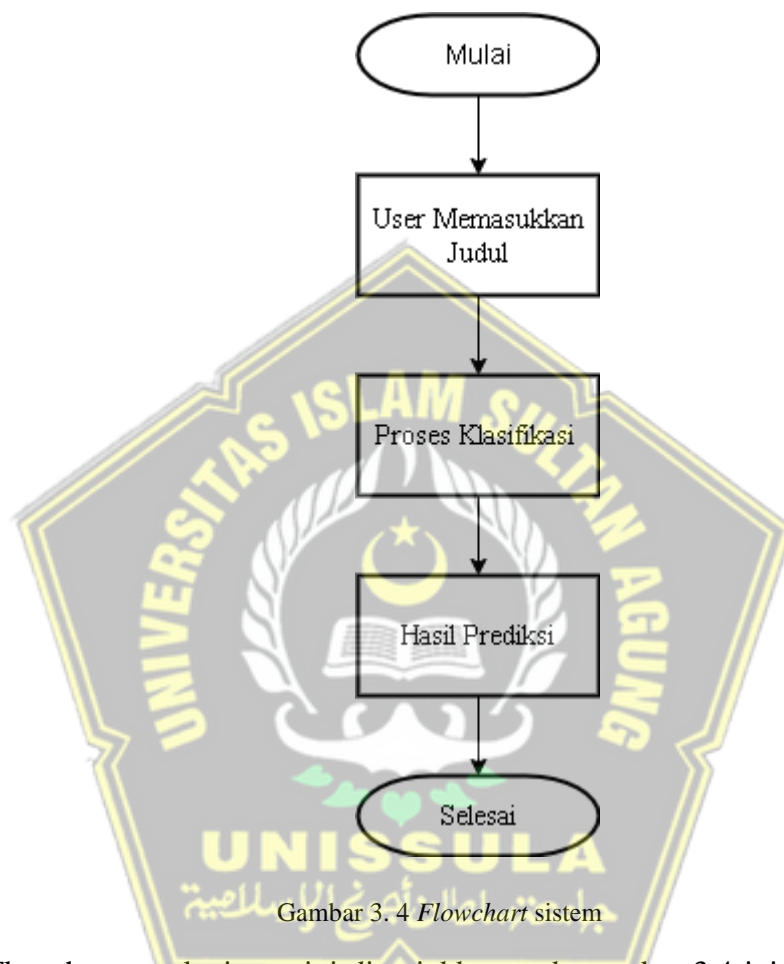
Pada alur implementasi sistem dibuat sebuah flowchart untuk menunjukkan alur perancangan dari sistem tersebut dan sekaligus alur kerja, dimana flowchart dari alur perancangan tersebut ditunjukkan pada Gambar 3.3.



Gambar 3. 3 Alur perancangan sistem

Gambar 3.3 menunjukkan urutan dalam perancangan sistem, dimana langkah pertama adalah pengumpulan data dari database SINTA, langkah

selanjutnya adalah validasi data, kemudian jika data sudah benar, langkah selanjutnya adalah perencanaan alur sistem dan dilanjutkan ke dalam implementasi *coding* , dan kemudian adalah penerapan metode. Kemudian pengujian sistem dan terakhir pelaporan.



Gambar 3.4 Flowchart sistem

Flowchart untuk sistem ini ditunjukkan pada gambar 3.4 ini nantinya, alur pengoperasian sistem ini dimulai dari pengguna mulai menggunakan perangkat yang digunakannya, dan pada penelitian ini penulis menggunakan laptop sebagai tahap utama di awal. Sebagai bagian dari sistem ini, pengguna harus membuka sistem ini setelah itu pengguna akan melihat beberapa menu seperti Home, Halaman Dataset, dan Halaman Predict. Pengguna kemudian memilih halaman prediksi. Halaman prediksi mengharuskan pengguna untuk memasukkan judul prediksi. Setelah user memasukkan judul, tombol prediksi selanjutnya ditekan untuk menyelesaikan proses klasifikasi, dimana sistem mengklasifikasikan judul

menggunakan metode Naïve Bayes. Sistem ini dapat memprediksi judul publikasi dari lima bidang ilmu tersebut, alur proses untuk sistem ini selesai.

3.3 Perancangan Interface

Pada bagian desain Perancangan interface yang dibuat pada penelitian ini berupa desain mockup, yang nantinya akan diimplementasikan dalam pembuatan sistem. Berikut adalah desain interface dari program klasifikasi bidang ilmu pada publikasi terindeks Scopus menggunakan metode Naive Bayes berbasis web.

1. Halaman Home

Halaman home adalah halaman pertama dari sebuah website yang berisi menu-menu website.

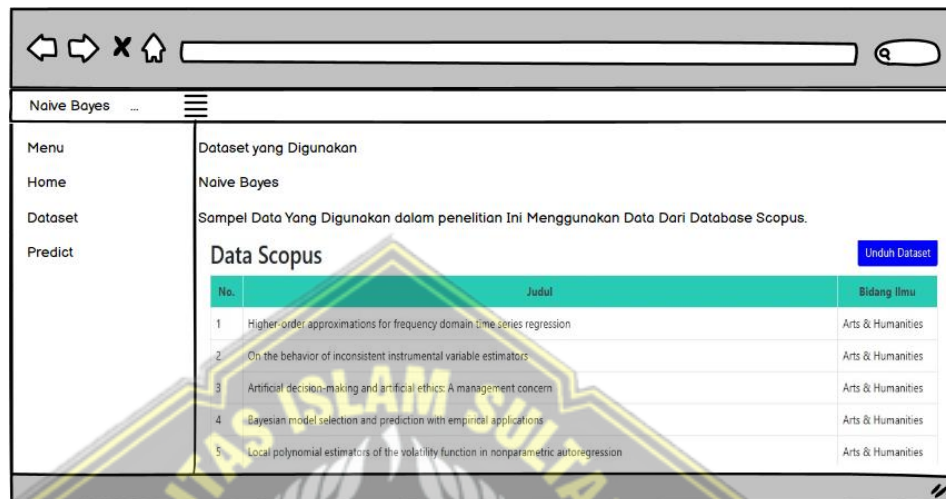


Gambar 3.5 halaman *home*

Gambar 3.5 diatas adalah tampilan halaman utama yang memuat judul tugas akhir dan deskripsi tugas akhir, serta beberapa gambar slide show tentang tema lima bidang ilmu tersebut.

2. Halaman Dataset

Halaman dataset adalah halaman yang menampilkan informasi yang nantinya akan digunakan dalam penelitian.

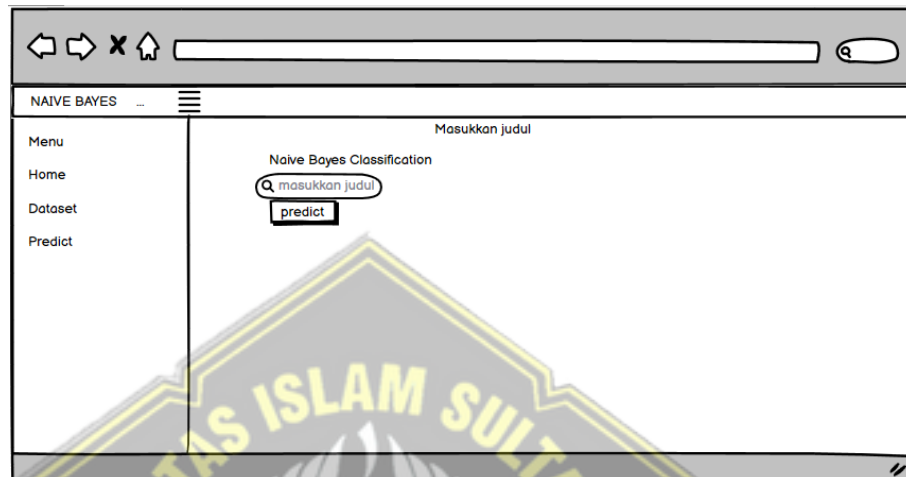


Gambar 3. 6 halaman *dataset*

Pada gambar 3.6 diatas merupakan gambar halaman dataset dimana halaman tersebut nantinya akan menampilkan data berupa judul artikel dan bidang ilmu dalam bentuk tabel. Kemudian website tersebut

3. Halaman Predict

Halaman predict merupakan halaman inti dari sistem yang dibangun kali ini, dimana halaman tersebut nantinya akan digunakan untuk menguji akurasi dari setiap judul yang diuji.

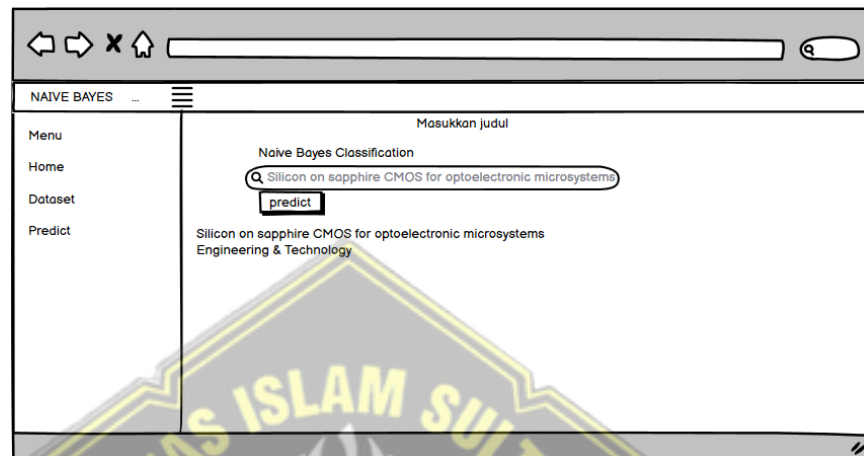


Gambar 3.7 halaman *predict*

Gambar 3.7 diatas menunjukkan desain gambar halaman prediksi yang akan diimplementasikan dalam pembuatan bab klasifikasi judul artikel. Pada halaman tersebut memiliki kolom untuk memasukkan judul artikel dan tombol untuk memproses klasifikasi tersebut.

4. *Output* dari halaman *predict*

Perancangan tampilan output halaman prediksi adalah tampilan sistem yang sangat penting di mana pengguna nantinya dapat mengetahui hasil akurasi judul artikel.



Gambar 3. 8 *Output* dari halaman *predict*

Gambar 3.8 diatas adalah desain gambar output yang nantinya akan digunakan untuk membuat sistem klasifikasi judul artikel. Kemudian dari desain tampilan output tersebut menampilkan judul artikel yang telah di input oleh pengguna dan diproses oleh sistem serta akurasinya di masing-masing 5 bidang ilmu.

BAB IV

HASIL DAN ANALISIS PENELITIAN

4.1 Hasil

Pengujian ini akan dihitung dan diklasifikasikan dengan menggunakan 5000 data yang terdiri dari masing-masing bidang terdapat 1000 data. pada pembahasan sebelumnya, telah disampaikan metode yang digunakan dalam penelitian ini adalah metode Naïve bayes. Dengan 5 tahapan yaitu pengumpulan data Scopus, *pre-processing*, penerapan metode naïve bayes, serta evaluasi. Hasil penelitian akan dijabarkan berdasarkan hasil akurasi

4.1.1. Pengumpulan data Scopus

Berdasarkan penelitian ini, peneliti menggunakan 5000 judul publikasi dimana masing-masing bidang terdapat 1000 judul publikasi yang diambil dari database SINTA. Adapun tabel 4.1 sampel data bidang ilmu *Art and Humanities*, tabel 4.2 sampel data bidang ilmu *Engineering and Technology*, tabel 4.3 sampel data bidang ilmu *Life Sciences and Medicine*, tabel 4.4 sampel data bidang ilmu *Natural Sciences*, tabel 4.5 sampel data bidang ilmu *Social Science and Management*. Tabel berikut adalah contoh beberapa sampel judul yang peneliti gunakan dari masing-masing 5 bidang ilmu.

1. Bidang Art and Humanities

Tabel 4. 1 Sampel Data Bidang Art and Humanities

No.	Judul
1.	The scale and nature of Stalinist repression and its demographic significance: on comments by Keep a
2.	Stable isotope and radiocarbon analyses of a black deposit associated with pictographs at Little Los
3.	Transient total tongue paralysis from simultaneous central and peripheral lesions
4.	The Development of Anxiety: The Role of Control in the Early Environment
5.	Medial efferent inhibition suppresses basilar membrane responses to near characteristic frequency to
6.	Nao and sea surface temperature signatures in tree-ring records from the North Atlantic sector

7.	Relationships between mental adjustment to HIV diagnosis, psychological morbidity and sexual behavior
8.	The Role of TGF β 's in Controlling Cell Adhesion and Differentiation of Colon Carcinoma Cells
9.	Disturbances in the system of hemostasis in patients suffering from non-insulin-dependent diabetes
10.	Soviet Industrialization Reconsidered: Some Preliminary Conclusions about Economic Development between
11.	Construct validity for measures of childhood depression: Application of multitrait-multimethod methodology
12.	On the relation of clinical research to clinical practice: Current issues, new directions
13.	Palaeoanthropological discoveries in Indonesia with special reference to the finds of the last two decades
14.	Verbal overshadowing effects on Raven's advanced progressive matrices: Evidence for multidimensional
15.	Parametric excitations of linear systems having many degrees of freedom

2. Bidang *Engineering and Technology*

Tabel 4. 2 Sampel Data Bidang *Engineering and Technology*

No.	Judul
1.	Evidence of lithium-nitrogen interaction in chitosan-based films from X-ray photoelectron spectroscopy
2.	An approach to functional synthesis of solutions in mechanical conceptual design. Part I: Introduction
3.	Analytical approach for economic risk quantification of large engineering projects
4.	Stabilization of Paramagnetic Metal(I) Complexes of the Nickel Subgroup. Electrochemical and Electro
5.	Synthesis of a Rigid Dimethoxynaphthalene-Spacer-Dithiol Which Spontaneously Attaches to Au and Pt E
6.	Some Newton-type methods for the regularization of nonlinear ill-posed problems
7.	The physical properties of oil palm empty fruit bunch (EFB) composites made from various thermoplast
8.	Characteristics of poly (ethylene oxide)-NaI polymer electrolyte and electrochemical cell performanc
9.	Electron-Transfer-Induced Isomerizations of Co-, Ni-, and Pd-Cyclooctatetraene Complexes: The Role
10.	An Alternate Path to Reductive Elimination for Group 4B Metals: Mechanism of Cyclopropane Formation
11.	Transition-Metal-Catalyzed Polymerization of Heteroatom-Functionalized Cyclohexadienes: Stereoregula

12.	Influence of varying cam profile and follower motion event types on parametric vibration and stability of flexible cam-follower systems
13.	Mechanism of Titanocene Metallacyclobutane Cleavage and the Nature of the Reactive Intermediate
14.	Single-mode optical fiber microbend loss modeling using the finite difference beam propagation method
15.	Synthesis and characterization of hexagonally packed mesoporous tantalum oxide molecular sieves

3. Bidang *Life Sciences and Medicine*

Tabel 4. 3 Sampel Data Bidang *Life Sciences and Medicine*

No.	Judul
1.	Trophic roles of heterotrophic nanoflagellates and ciliates among planktonic organisms in a hypereut
2.	Optimal timing of gibberellin A _{4/7} application to increase female strobilus numbers in a <i>Pinus radiata</i> seed orchard
3.	Effect of Vitamin A Levels in <i>Artemia Nauplii</i> on the Caudal Skeleton Formation of Japanese Flounder <i>Paralichthys olivaceus</i>
4.	Femtosecond pump-probe spectroscopy of conjugated polymers: coherent and sequential contributions
5.	Diurnal patterns of heat production and heart rate under thermoneutral conditions in Holstein Friesian cows differing in milk production
6.	Stereochemical Investigations of Coordinated Sulfur Stereocenters. X-Ray Structures of Diastereomers
7.	Cyclopentadienyl-ruthenium and -osmium chemistry. XXXIX. Syntheses of novel alkynyl-ruthenium comple
8.	Strained rings as a source of unsaturation: Polybenzvalene, a new soluble polyacetylene precursor
9.	Bis(ligand) Rhenium(V) and Technetium(V) Complexes of Two Naturally Occurring Binding Moieties (Oxaz)
10.	High-power, continuous-wave, current-tunable, single-mode quantum-cascade distributed-feedback laser
11.	Kinetics and Isotope Effects of the Aldol-Tishchenko Reaction between Lithium Enolates and Aldehydes
12.	Crystal structure of Myotoxin-II: A myotoxic phospholipase A2 -homologue from <i>Bothrops moojeni</i> veno
13.	Quantitative gas sensing by backscatter-absorption measurements of a pseudorandom code modulated \hat{I}
14.	Hydroxamate-Bridged Dinuclear Nickel Complexes as Models for Urease Inhibition 1
15.	Ion explosion and multi-mega-electron-volt ion generation from an underdense plasma layer irradiated

4. Bidang *Natural Sciences*Tabel 4. 4 Sampel Data Bidang *Natural Sciences*

No.	Judul
1.	Rapid thermal annealing of ion-implanted 6H-SiC by microwave processing
2.	Temperature-entropy diagram for an irreversible absorption refrigeration cycle
3.	Evaluation of calcium carbonate and calcium hydroxide treatments of prairie drinking water dugouts
4.	Evidence of lithium-nitrogen interaction in chitosan-based films from X-ray photoelectron spectroscopy
5.	Mineral and organic particles in astronom
6.	Ion explosion and multi-mega-electron-volt ion generation from an underdense plasma layer irradiated
7.	Homoepitaxial growth and hydrogen incorporation on the chemical vapor-deposited (111) diamond
8.	Study of overoxidized polypyrrole using X-ray photoelectron spectroscopy
9.	Surface effects on the magnetic properties of ultrafine cobalt particles
10.	Effect of hydrogen dilution on the deposition of carbon-rich SiC:H films by the electron cyclotron
11.	Effect of an annular pupil on confocal imaging through highly scattering media
12.	The iron-sulfur bond in cytochrome c
13.	Nonlinear interaction of a powerful laser with an electron plasma
14.	Analytical evaluation of three- and four-center electron-repulsion integrals for Slater-type orbital
15.	Effect of geometrical confinement on the interaction between charged colloidal suspensions

5. Bidang *Sosial Science and Management*Tabel 4. 5 Sampel Data Bidang *Sosial Science and Management*

No.	Judul
1.	Commentaries on "generalizing about univariate forecasting methods: Further empirical evidence"
2.	On the behavior of inconsistent instrumental variable estimators
3.	Culture-specific and cross culturally generalizable implicit leadership theories Are attributes of charismatic or transformational leadership universally endorsed?
4.	Dynamic models for fixed-income portfolio management under

	uncertainty
5.	Urban solid waste management in Bandung: Towards an integrated resource recovery system
6.	An interdisciplinary research initiative towards sustainable management of forest resources in lowla
7.	The demand for indonesian narrow money: Long-run equilibrium error correction and forward-looking behavior
8.	CONVERGENT ALGORITHMS FOR MINIMIZING A CONCAVE FUNCTION.
9.	Elaboration of GIS ecosystem monitoring and management of the Astrakhanskiy Biosphere Reserve
10.	Reengineering for human resource management in Japanese companies: is it important to be introduced?
11.	Scenario modeling for the management of international bond portfolios
12.	ustainable land management with rubber-based agroforestry: A Bangladeshi example of uplands communi
13.	Unfolding and refolding of the NAD-dependent isocitrate dehydrogenase from yeast
14.	Gender socioeconomic development and health-seeking behaviour in Bangladesh
15.	Teccon: Support systems for the process management

4.1.2. Preprocessing

Dalam tahapan ini merupakan tahap dimana terjadinya proses seleksi data agar data diambil lebih terstruktur dalam setiap dokumen. Tahapan *Pre-processing* terdiri dari Data Cleaning, tokenizing, stopword, & stemming. Data cleaning adalah proses pemecah deskripsi yang semula berupa kalimat menjadi kata dan menghilangkan delimiter seperti titik (.), koma (,), tanda petik (“), tanda kurung ({}), spasi dan angka. Case folding merupakan langkah untuk mengubah semua huruf menjadi huruf kecil. Tokenizing adalah proses memecah isi pengaduan menjadi kata. Stopword Removal merupakan langkah menghapus kata penting dari hasil tokenizing. Stemming merupakan langkah mengubah berbagai bentuk (variasi) kata menjadi bentuk kata dasarnya. Tahap preprocessing dalam penelitian ini yaitu pengkodean data pada judul artikel di SINTA yang terindeks Scopus.

4.1.3 Penerapan metode Naïve Bayes

Setelah melalui proses pengumpulan data dan preprocessing, tahapan selanjutnya adalah penerapan metode Naïve Bayes yang bermanfaat mengklasifikasikan data scopus dari hasil langkah preprocessing sebelumnya.

```
// pipeline untuk preprocessing dan klasifikasi
$estimator = new PersistentModel(
  new Pipeline([
    //PREPROCESSING
    new RegexFilter([
      RegexFilter::EXTRA_WHITESPACE,
      RegexFilter::EXTRA_WORDS
    ]), //text cleaning
    new MultibyteTextNormalizer(), //case folding - lower case
    new StopWordFilter($stopwords_en), //stopwords
    // new WordCountVectorizer(), //tokenizer (hapus untuk Bayes)

    // ALGORITMA KLASIFIKASI
  ], new NaiveBayes(null, 2.5)),
  // nama file model
  new Filesystem('./kusumata.model')
);
```

Gambar 4. 1 Tahapan *Preprocessing*



4.1.4 Evaluasi

Pada tahap ini, data melalui proses penerapan metode Naïve Baye selanjutnya data akan di evaluasi hasil akhir nya melalui pengukuran Akurasi, Presisi, dan *Recall*.

```
// testing menggunakan data testing dari dataset
$predictions = $estimator->predict($testing);

// validasi, confusion matrix, akurasi, dll
$report = new AggregateReport([
    'breakdown' => new MulticlassBreakdown(),
    'confussion_matrix' => new ConfusionMatrix(),
]);

$results = $report->generate($predictions, $testing->labels());

echo $results;
```

Gambar 4. 2 Tahapan Evaluasi

4.2 User Interface dan Penggunaan Sistem

Pada sistem ini memiliki tampilan menu yang terdiri dari halaman *home*, *dataset*, dan *predict* tampilan tersebut bisa disebut dengan *user interface*. Peran *user interface* bagian yang sangat penting karena sering dilihat dan diinteraksikan. Berikut adalah tampilan *user interface* dan penjelasan kegunaan dari sistem tersebut yang telah dibuat peneliti yaitu :

A. Halaman *Home*

Pada gambar 4.3 terdapat menu *Home* dimana tampilan berisi judul tugas akhir, dan penjelasan deskripsi tugas akhir. Pada halaman home ini *User* hanya dapat melihat penjelasan deskripsi tugas akhir dari sistem tersebut.

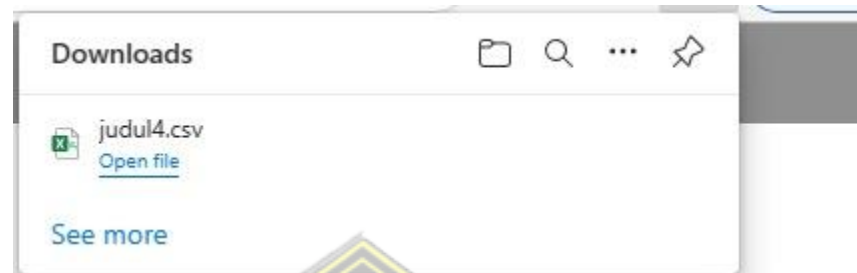
Gambar 4. 3 Tampilan halaman *Home*

B. Halaman Data Set

Pada gambar 4.4 adalah tampilan pada halaman *dataset* dimana halaman tersebut menampilkan tabel berisikan data *Scopus* yang telah diberi judul dari masing-masing 5 bidang ilmu. Data tersebut berjumlah 5000 judul dengan masing-masing bidang ilmu terdapat 1000 judul artikel jurnal publikasi. *User* dapat melihat hasil data pada tabel dan mengunduh *dataset* tersebut pada tombol yang telah disediakan. Hasil dari unduhan berupa format excel.

Gambar 4. 4 Tampilan halaman *dataset*

Pada gambar dibawah 4.5 tampilan fitur *download* fitur yang tersedia pada menu *dataset* yaitu *user* dapat melihat mengunduh dataset tersebut pada tombol yang telah disediakan. Hasil dari unduhan berupa format excel.

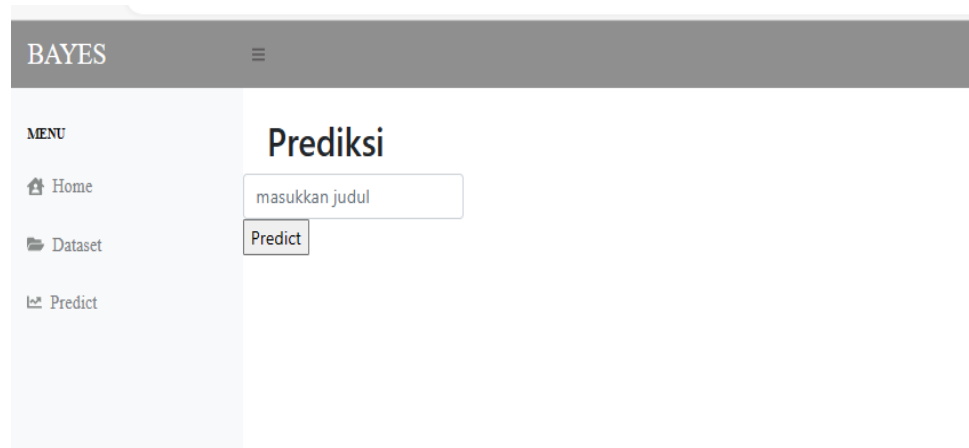


Unduh Dataset

Gambar 4. 5 Tampilan fitur *download*

C. Halaman *Predict*

Pada gambar 4.6 terdapat halaman *predict* merupakan menu terakhir, halaman ini menampilkan kolom input yang nantinya *user* akan memasukkan judul artikel ke dalam kolom tersebut, kemudian di dalam halaman tersebut juga terdapat *predict button* yang mana tombol tersebut berfungsi untuk memproses prediksi yang akan dilakukan oleh sistem tersebut.



Gambar 4. 6 Tampilan halaman prediksi

D. Hasil prediksi

Pada gambar dibawah 4.7 merupakan tampilan hasil dari pencarian data prediksi berupa judul artikel beserta persentase akurasi di setiap masing-masing bidang ilmu, dengan tampilan tersebut memudahkan *user* untuk mengetahui akurasi judul artikel tersebut masuk kedalam salah satu bidang ilmu.



Gambar 4. 7 Tampilan hasil prediksi

4.3 Analisa dan Pengujian

Pada tahapan ini, dilakukan pengujian sistem di mana metode yang digunakan tersebut diuji. Pengujian black box yang digunakan sebagai pengujian penelitian. Pengujian ini merupakan pengujian perangkat lunak dengan tujuan menguji fungsi aplikasi terlepas dari struktur atau cara kerjanya. Ada beberapa jenis pengujian seperti tes fungsional, tes non fungsional dan tes regresi. Pengujian sistem ini dioperasikan untuk menguji setiap fitur dari hasil uji tabel 4.6

Tabel 4. 6 analisa dan pengujian

Skenario Pengujian	Kasus Pengujian	Hasil Pengujian	Kesimpulan
Unduh Dataset	Mengunduh Sampel Data.	berhasil	sesuai
Memasukkan Judul	Memasukkan judul publikasi	berhasil	sesuai
Memprediksi Judul	Sistem memprediksi judul	berhasil	sesuai
Hasil Prediksi Judul	Output hasil prediksi judul	berhasil	sesuai

Dari hasil tabel 4.6 di atas terlihat bahwa pengujian dilakukan menggunakan sistem klasifikasi naive bayes pada publikasi terindeks Scopus sudah memadai dan dapat dijalankan untuk setiap fungsinya. Tahap pengujian berjalan dengan baik sesuai dengan fungsi utamanya.

4.4 Analisis Akurasi

Tahapan selanjutnya adalah tahap analisis akurasi, yang tujuannya adalah untuk mengetahui seberapa besar akurasi yang telah dicapai oleh sistem. Tahap pengujian menggunakan data testing, dilanjutkan dengan pembagian dari data testing dan data sampel. Tahapan ini merupakan bagian dari langkah untuk menentukan tingkat akurasi dari sistem yang dibuat

Tabel 4. 7 Jumlah Data

Nama Data	Jumlah (%)	Jumlah angka
Data Sampel	90%	4500
Data Testing	10%	500

Total Data : 5000

Tabel 4. 8 Rincian Data

class	Arts & Humanities	Engineering & Technology	Life Science & Medicine	Natural Sciences	Social Sciences & Management
Data Sampel	900	900	900	900	900
Data Testing	100	100	100	100	100

Dari data tersebut menghasilkan confusion matrix, confusion matrix ini digunakan untuk menentukan akurasi, presisi, recall & skor F1. Dari hasil training yang diujikan, menghasilkan confusion matrix dengan nilai akurasi, presisi, recall dan F1 sebagai berikut:

Tabel 4. 9 confusion matrix

Accuracy	Precision	Recall	F1 score
0.42575	0.0558	0.20133	0.0742

```
PS C:\xampp\htdocs\kusumata> php training-single-csv.php
Training model.... Model disimpan.... OK {
  "breakdown": {
    "overall": {
      "accuracy": 0.42575101221745887,
      "balanced accuracy": 0.4638533137205617,
      "f1 score": 0.07426385232676765,
      "precision": 0.05586464843221599,
      "recall": 0.20133333333333336,
      "specificity": 0.7263732941077901,
    }
  }
}
```

Gambar 4. 8 Output hasil nilai confusion matrix

Gambar 4.8 menunjukkan hasil pengujian dengan data sampel 1000 judul artikel Naive Bayesian menghasilkan nilai akurasi sebesar 0,425 dengan presisi, recall, dimana F1 masing-masing memiliki nilai 0,055, 0,201 dan 0,074 yang menunjukkan bahwa metode Naive Bayesian belum optimal untuk klasifikasi judul artikel jurnal publikasi scopus dimana metode tersebut hanya memberikan akurasi sebesar 0,425, sedangkan hasil tersebut masih tergolong rendah. Nilai akurasi yang rendah juga dapat disebabkan oleh beberapa faktor seperti datanya, dimana data tersebut digunakan peneliti masih sangat kotor dan perlu banyak pembersihan, pembersihan yang lebih dalam akan merubah format nama data tersebut yang berupa judul, sehingga itu tidak bisa dilakukan oleh peneliti yang mana akan mengubah substansi dari arti judul tersebut, kemudian yang kedua banyak ditemukan kemiripan judul antara satu bidang ilmu kesamaan.

Tabel 4. 10 Confusion Matrix

	Positive	Negatif
True	302	1208
False	1198	1198


```
"true positives": 302,  
"true negatives": 1208,  
"false positives": 1198,  
"false negatives": 1198,
```

Gambar 4. 9 Output hasil (*True Positives, True Negatives, False Positives, False Negatives*)

Berdasarkan pengujian yang dilakukan, dapat disimpulkan bahwa pengujian menggunakan total 5000 data, dimana rincian pembagian data training mencapai 90%=4500 dan data testing mencapai 10%=500. Hasilnya adalah akurasi keseluruhan sebesar 0,425 (42%).

4.5 Pembahasan Hal Yang Mempengaruhi Hasil Akurasi

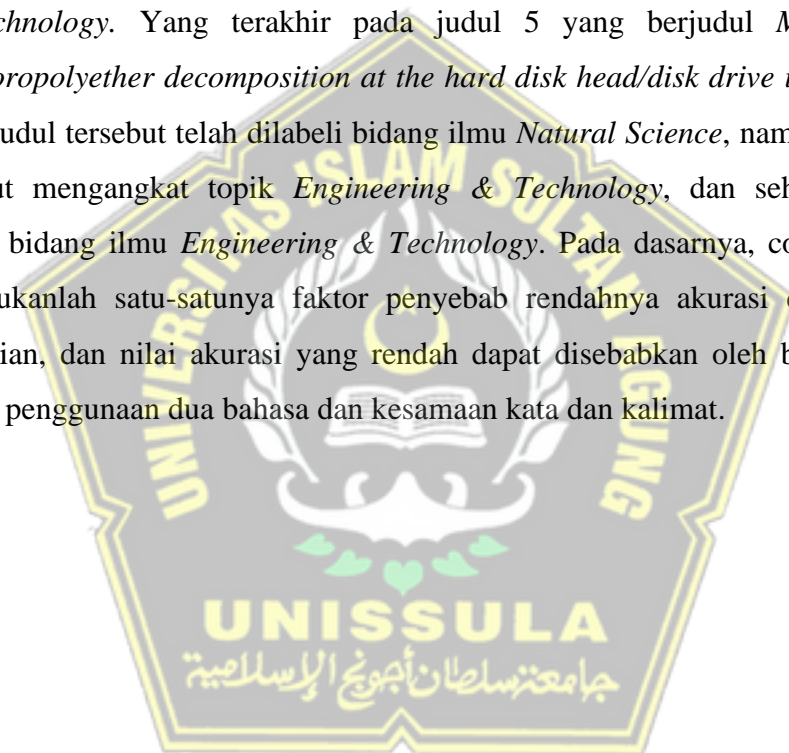
Hasil pengujian yang dilakukan menunjukkan tingkat akurasi yang cukup rendah yaitu 0,425 (42%). Oleh karena itu, ini adalah salah satu faktor yang dapat menyebabkan hasil rendahnya nilai akurasi tes yang dilakukan. Kata dan frasa dari satu bidang ilmu ke bidang ilmu lainnya memiliki banyak kesamaan, sehingga akurasinya rendah. Ada judul yang terkait dengan topik yang sama tetapi dibagi ke dalam bidang ilmu yang berbeda

Tabel 4. 11 Contoh Judul yang memiliki topik yang tidak sesuai bidannya

No.	Judul Artikel	Bidang Ilmu yang terlabel	Bidang Ilmu yang seharusnya
1.	Performance of neuropsychological tests in mildly hypoxemic patients with chronic obstructive pulmonary disease	Arts & Humanities.	Life science & medicine
2.	Bayesian model selection and prediction with empirical applications.	Social Sciences & Management	Engineering & Technology
3.	Digital data networks design using genetic algorithms.	Social Sciences & Management	Engineering & Technology
4.	Artificial neural networks: a new approach for treating electrical engineering problems	Social Sciences & Management	Engineering & Technology
5.	Mechanism of perfluoropolyether decomposition at the hard disk head/disk drive interface	Natural Sciences	Engineering & Technology

Pada Tabel tersebut merupakan tabel yang berisi contoh judul yang mempengaruhi nilai akurasi. tabel tersebut mencakup judul yang topiknya tidak sesuai dengan bidang ilmunya, seperti tabel pertama dimana pada tabel tersebut berjudul *Performance of neuropsychological tests in mildly hypoxemic patients with chronic obstructive pulmonary disease* yang mana judul tersebut telah dilabeli bidang ilmu *Arts & Humanities*, namun pada judul tersebut mengangkat topik *Life science & medicine*, dan seharusnya juga masuk bidang ilmu *Life science & medicine*. Kemudian pada judul 2 yang berjudul *Bayesian model selection and prediction with empirical applications.* yang mana judul tersebut telah dilabeli bidang ilmu *Social Sciences & Management*, namun pada judul tersebut mengangkat topik *Engineering & Technology*, dan seharusnya juga masuk bidang ilmu *Engineering & Technology*. Kemudian pada judul 3 yang

berjudul *Digital data networks design using genetic algorithms*. yang mana judul tersebut telah dilabeli bidang ilmu *Social Sciences & Management*, namun pada judul tersebut mengangkat topik *Engineering & Technology*, dan seharusnya juga masuk bidang ilmu *Engineering & Technology*. selanjutnya pada judul 4 yang berjudul *Artificial neural networks: a new approach for treating electrical engineering problems*. yang mana judul tersebut telah dilabeli bidang ilmu *Social Sciences & Management*, namun pada judul tersebut mengangkat topik *Engineering & Technology*, dan seharusnya juga masuk bidang ilmu *Engineering & Technology*. Yang terakhir pada judul 5 yang berjudul *Mechanism of perfluoropolyether decomposition at the hard disk head/disk drive interface* yang mana judul tersebut telah dilabeli bidang ilmu *Natural Science*, namun pada judul tersebut mengangkat topik *Engineering & Technology*, dan seharusnya juga masuk bidang ilmu *Engineering & Technology*. Pada dasarnya, contoh judul di atas bukanlah satu-satunya faktor penyebab rendahnya akurasi dalam sebuah penelitian, dan nilai akurasi yang rendah dapat disebabkan oleh banyak faktor, seperti penggunaan dua bahasa dan kesamaan kata dan kalimat.



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari penelitian ini dapat disimpulkan bahwa dengan mengklasifikasi bidang ilmu publikasi terindeks Scopus menggunakan metode Naïve bayes menghasilkan output yaitu kesesuaian judul artikel dalam SINTA yang sudah terindeks oleh Scopus dalam 5 bidang ilmu tersebut. Hasil dari pengujian sistem yang cukup dengan akurasi, recall, presisi. Diperoleh nilai akurasi sebesar 0,425, recall 0,20 dan presisi 0,05, Namun hasil tersebut masih cukup rendah jika digunakan pada sistem SINTA dikarenakan pada saat pelabelan dataset yang digunakan belum optimal. Dari percobaan diatas dataset yang diharapkan tidak sesuai, tetapi sistem bekerja dengan baik sebagaimana fungsinya mampu mengklasifikasi.

5.2 Saran

Saran untuk perangkat Sistem:

1. Untuk penelitian selanjutnya, perlu ditambahkan algoritma lain untuk mendukung hasil yang lebih baik agar nilai akurasi yang dihasilkan mendapatkan nilai yang lebih baik.
2. Untuk sistem yang tepat, UI dibuat semenarik mungkin.
3. Berbagai fungsi dapat ditambahkan ke pengembangan sistem, seperti fitur akses login pada tampilan utama.
4. Untuk penelitian selanjutnya perlu di upgrade versi mobile.

DAFTAR PUSTAKA

- Adi, S. (2018). Implementasi Algoritma Naive Bayes Classifier Untuk Klasifikasi Penerima Beasiswa PPA di Universitas Amikom Yogyakarta. *Jurnal Mantik Penusa*, 22(1), 11–16.
- Agus Sugianto, C., & Rizky Maulana, F. (2019). Algoritma Naïve Bayes Untuk Klasifikasi Penerima Bantuan Pangan Non Tunai. *Techno.COM*, 18(4), 321–331.
- Amaliyah. (2019). Program Kemitraan Masyarakat: Peningkatan Kemampuan Penulisan Karya Ilmiah dan Teknik Publikasi di Jurnal Internasional. *Intervensi Komunitas*, 1(1), 48–56.
- Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah*, 10(2), 160–165. <https://doi.org/10.33096/ilkom.v10i2.303.160-165>
- Ariyanti, D., & Iswardani, K. (2020). Teks Mining untuk Klasifikasi Keluhan Masyarakat Pada Pemkot Probolinggo Menggunakan Algoritma Naïve Bayes. *Jurnal IKRA-ITH Informatika*, 4(3), 125–132.
- Ciputra, A., Setiadi, D. R. I. M., Rachmawanto, E. H., & Susanto, A. (2018). Klasifikasi Tingkat Kematangan Buah Apel Manalagi Dengan Algoritma Naive Bayes Dan Ekstraksi Fitur Citra Digital. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 9(1), 465–472. <https://doi.org/10.24176/simet.v9i1.2000>
- Devita, R. N., Herwanto, H. W., & Wibawa, A. P. (2018). Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 5(4), 427. <https://doi.org/10.25126/jtiik.201854773>
- Ericha Apriliyani, & Salim, Y. (2022). Analisis performa metode klasifikasi

- Naïve Bayes Classifier pada Unbalanced Dataset. *Indonesian Journal of Data and Science*, 3(2), 47–54. <https://doi.org/10.56705/ijodas.v3i2.45>
- Fadri, W. (2023). Klasifikasi Penyakit Hati dengan Menggunakan Metode Naive Bayes. *Jurnal Informasi dan Teknologi*, 5(1), 32–36. <https://doi.org/10.37034/jidt.v5i1.230>
- Firmansyah, A., Qadri, R. A., & Arham, A. (2020). Pelatihan melalui Web Seminar terkait Publikasi Artikel untuk Menembus Jurnal Sinta 2 dan Scopus. *Abdimas: Jurnal Pengabdian Masyarakat Universitas Merdeka Malang*, 5(2), 131–138. <https://doi.org/10.26905/abdimas.v5i2.4244>
- Hozairi, H., Anwari, A., & Alim, S. (2021). Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes. *Network Engineering Research Operation*, 6(2), 133. <https://doi.org/10.21107/nero.v6i2.237>
- Ibrahim, C. (2019). *Tren Kepenulisan Publikasi Penelitian*. 6003, 13–28.
- Imandasari, T., Irawan, E., Windarto, A. P., & Wanto, A. (2019). Algoritma Naive Bayes Dalam Klasifikasi Lokasi Pembangunan Sumber Air. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1(September), 750. <https://doi.org/10.30645/senaris.v1i0.81>
- Indra Borman, R., & Wati, M. (2020). Penerapan Data Maining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandarlampung Dengan Algoritma Naïve Bayes. *Jurnal Ilmiah Fakultas Ilmu Komputer*, 09(01), 25–34.
- Kartikasari, E., & Dianti, A. (2021). Strategi Publikasi Unit Pengelolaan Informasi Majapahit Dalam Menyampaikan Informasi Di Masa Pandemi Covid-19. *eProceedings of Management*, 8(5).
- Mustopa, M. A., Famuzia, M. A., Budiawan, R., & Telkom, U. (2021). *Sema : Sistem Pendataan Dan Publikasi Elektronik Kabupaten*. 7(5), 1914–1922.
- Nasrullah, A. H. (2021). Implementasi Algoritma Decision Tree Untuk Klasifikasi

- Produk Laris. *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar*, 7(2), 45–51.
- Prakoso, B. S., Rosiyadi, D., Utama, H. S., & Aridarma, D. (2019). Klasifikasi Berita Menggunakan Algoritma Naive Bayes Classifier Dengan Seleksi Fitur Dan Boosting. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 3(2), 227–232. <https://doi.org/10.29207/resti.v3i2.1042>
- Prasetya, F., & Ferdiansyah, F. (2022). Analisis Data Mining Klasifikasi Berita Hoax COVID 19 Menggunakan Algoritma Naive Bayes. *Jurnal Sistem Komputer dan Informatika (JSON)*, 4(1), 132. <https://doi.org/10.30865/json.v4i1.4852>
- Putra, M. Y., & Putri, D. I. (2022). Pemanfaatan Algoritma Naive Bayes dan K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Kelas XI. *Jurnal Tekno Kompak*, 16(2), 176–187.
- Ruhyana, N. (2019). Analisis Sentimen terhadap Penerapan Sistem Plat Nomor Gnajil/Genap pada Twitter dengan Metode Klasifikasi Naive Bayes. *Jurnal IKRA-ITH Informatika*, 3(1), 94–99. www.situs.com
- Siddik, M., Hendri, H., Putri, R. N., Desnelita, Y., & Gustientiedina, G. (2020). Klasifikasi Kepuasan Mahasiswa Terhadap Pelayanan Perguruan Tinggi Menggunakan Algoritma Naive Bayes. *INTECOMS: Journal of Information Technology and Computer Science*, 3(2), 162–166. <https://doi.org/10.31539/intecom.v3i2.1654>
- Utami, D., Aisyiah, P., & Devi, R. (2022). *Klasifikasi kelayakan penerima bantuan program keluarga harapan (pkh) menggunakan metode weighted naïve bayes dengan laplace smoothing*. 07, 1373–1384.
- Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *Jurnal Media Informatika Budidarma*, 4(2), 437. <https://doi.org/10.30865/mib.v4i2.2080>
- Waliyansyah, R. R., & Fitriyah, C. (2019). Perbandingan Akurasi Klasifikasi

Citra Kayu Jati Menggunakan Metode Naive Bayes dan k-Nearest Neighbor (k-NN). *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 5(2), 157. <https://doi.org/10.26418/jp.v5i2.32473>

Yuwono, L., Fadillah, M. egi, Indrayani, M., & ... (2021). Klasifikasi Pendapatan Pedagang Kaki Lima Dan Pelaku Usaha Online Akibat Dampak Covid-19 Menggunakan Metode Naive Bayes. *Bulletin of Applied ...*, 2(1), 1–6. <http://jim.unindra.ac.id/index.php/baiet/article/view/3947>

