

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

LAPORAN TUGAS AKHIR

Laporan ini Disusun untuk Memenuhi Salah Satu Syarat Memperoleh Gelar Sarjana
Strata 1 (S1) pada Program Studi Teknik Informatika Fakultas Teknologi Industri
Universitas Islam Sultan Agung Semarang



DISUSUN OLEH :

KUSUMA HADI WIJAYA

NIM 32601800014

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM SULTAN AGUNG
SEMARANG**

DESEMBER 2022

FINAL PROJECT

***CLASSIFICATION OF SCIENCE FIELDS IN GARUDA 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 HADI WIJAYA

NIM 32601800014

MAJORING OF INFORMATICS ENGINEERING

INDUSTRIAL TECHNOLOGY FACULTY

SULTAN AGUNG ISLAMIC UNIVERSITY

SEMARANG

DECEMBER 2022

LEMBAR PENGESAHAN PEMBIMBING

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

Nama : Kusuma Hadi Wijaya

NIM : 32601800014

Program Studi : Teknik Informatika


Telah disahkan oleh dosen pembimbing pada :

Hari : Kamis

Tanggal : 29 Desember 2022

Mengesahkan,

Pembimbing I


Andi Riansyah, ST.M.Kom

NIDN. 0609108802


Pembimbing II


Sam Farisa C, ST.M.Kom

NIDN. 0628028602

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 Garuda Menggunakan Metode Naïve Bayes” ini telah dipertahankan di depan dosen penguji Tugas Akhir pada :


Hari : Kamis


Tanggal : 29 Desember 2022

TIM PENGUJI

Penguji I

Penguji II


Imam Much Ibnu Subroto, ST, M.Sc, Phd
NIDN.0613037301


Ghufron, ST, M.Kom
NIDN. 062079005

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Kusuma Hadi Wijaya

Nim : 32601800014

Nama Tugas Akhir : Klasifikasi Bidang Ilmu Pada Publikasi Terindeks
Garuda 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, 15 Februari 2023
Yang Menyatakan



Kusuma Hadi wijaya

SURAT PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan dibawah ini :

Nama : Kusuma Hadi Wijaya

Nim : 32601800014

Program Studi : Teknik Informatika

Fakultas : Teknologi Industri

Alamat Asal : RT 03/ RW 02 Ds. Sowan Kidul Kec. kedung Kab. Jepara
Jawa Tengah

Dengan ini menyatakan Karya Ilmiah berupa Tugas Akhir dengan Judul :
Klasifikasi Bidang Ilmu Pada Publikasi Terindeks Garuda 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/Plagiatisme dalam karya ilmiah ini, maka
segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa
melibatkan Universitas Islam Sultan agung.

Semarang, 15 Februari 2023

Yang Menyatakan



Kusuma Hadi wijaya

KATA PENGANTAR

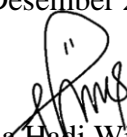
Dengan mengucapkan syukur alhamdulillah atas kehadiran Allah SWT yang telah memberikan rahmat dan karunianya kepada penulis, sehingga dapat menyelesaikan Tugas Akhir dengan judul “ Rancang Bangun Sistem Kepuasan Pelanggan Terhadap Kualitas Pelayanan Menggunakan Metode *Servqual* Di Rumah Makan Kalipening Kendal” 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 Andi Riansyah, ST, M.Kom yang telah meluangkan waktu dan memberi ilmu.
4. Dosen pembimbing II penulis Sam Farisa Chaerul Haviana, ST, M.Kom yang 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 dan masa mendatang.

Semarang, 29 Desember 2022


Kusuma Hadi Wijaya

DAFTAR ISI

HALAMAN JUDUL.....	ii
LEMBAR PENGESAHAN PEMBIMBING.....	iii
LEMBAR PENGESAHAN PENGUJI.....	iv
SURAT PERNYATAAN KEASLIAN TUGAS AKHIR.....	v
SURAT PERSETUJUAN PUBLIKASI KARYA ILMIAH.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL.....	xii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Batasan Masalah.....	3
1.4. Tujuan Tugas Akhir.....	3
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI.....	5
2.1. Tinjauan Pustaka.....	5
2.2. Dasar Teori.....	10
2.2.1 Garuda.....	10
a. <i>Preprocessing</i>	12
b. Klasifikasi.....	13
c. <i>Naïve Bayes</i>	13
BAB III METODE PENELITIAN.....	16
3.1. Tahapan Penelitian.....	16
3.1.1 Koleksi data Garuda.....	18

3.1.2	<i>Preprocessing</i>	20
3.1.3	Penerapan Metode.....	24
3.1.4	Evaluasi.....	29
3.2	Metodologi Perancangan Alur Sistem.....	30
3.2.1	Analisa Kebutuhan Fungsi Sistem	30
3.2.2	Analisa Kebutuhan Sistem	31
1.	Rubix ML	31
2.	Visual studio Code	31
3.	Bootstrap	31
3.3	Perancangan Interface	34
BAB IV HASIL DAN ANALISIS PENELITIAN		39
4.1	Implementasi User Interface (UI).....	39
a.	Tampilan Menu Bar.....	39
b.	Halaman Home.....	40
e.	Proses download dataset.....	42
f.	Tampilan halaman predict.....	43
4.2	Pengujian Sistem.....	45
4.3	Analisa Hasil Validasi Penerapan Algoritma Naive Bayes.....	46
4.4	Pembahasan hasil akurasi	48
BAB V KESIMPULAN DAN SARAN.....		51
5.1	Kesimpulan.....	51
5.2	Saran	51
DAFTAR PUSTAKA		52
LAMPIRAN.....		55

DAFTAR GAMBAR

Gambar 3. 1 Skema tahapan proses penelitian.....	16
Gambar 3. 2 Tahapan preprocessing.....	17
Gambar 3. 3 Data sampel 500 judul artikel jurnal garuda	20
Gambar 3. 4 Penerapan preprocessing pada kode program	21
Gambar 3. 5 Kode Program Cleaning	21
Gambar 3. 6 kode Program Case Folding	22
Gambar 3. 7 Kode Program Tokenizing	22
Gambar 3. 8 Kode Program Stopwords	23
Gambar 3. 9 Kode Program Stemming	23
Gambar 3. 10 Proses penerapan metode di dalam coding	28
Gambar 3. 11 Proses evaluasi dengan Confusion Matrix di dalam coding	29
Gambar 3. 12 Flowchart perancangan alur sistem	32
Gambar 3. 13 Flowchart alur fungsi sistem	33
Gambar 3. 14 Rancangan tampilan menu bar	34
Gambar 3. 15 Halaman Home.....	35
Gambar 3. 16 Halaman Home full screen.....	35
Gambar 3. 17 Halaman Dataset	36
Gambar 3. 18 halaman Dataset Full Screen	36
Gambar 3. 20 Halaman output predict	38
Gambar 4. 1 Tampilan Menu Bar	39
Gambar 4. 2 Tampilan halaman awal sistem klasifikasi.....	40
Gambar 4. 3 Tampilan halaman dataset	41
Gambar 4. 4 Tampilan halaman dataset full screen	41
Gambar 4. 5 Download button	42
Gambar 4. 6 Proses dataset yang telah terdownload.....	42
Gambar 4. 7 Tampilan halaman Predict.....	43
Gambar 4. 8 Tampilan halaman Predict full screen	43
Gambar 4. 9 Tampilan output hasil prediksi	44

Gambar 4. 10 Output hasil Confusion Matrix..... 47

Gambar 4. 11 Pengukuran Accuracy, Recall, Precision, F1 score..... 47



DAFTAR TABEL

Tabel 3. 1 Contoh sampel data judul dari garuda	19
Tabel 3. 2 Contoh proses data input dan output pada cleaning.....	21
Tabel 3. 3 Contoh Proses data input dan output pada case folding.....	22
Tabel 3. 4 Contoh proses data input dan output pada tokenizing	22
Tabel 3. 5 Contoh proses data input dan output pada stopword	23
Tabel 3. 6 contoh proses data input dan output pada stemming	23
Tabel 3. 7 Tabel pembagian contoh dataset yang dibagi menjadi data training dan testing	24
Tabel 3. 8 Pemecahan kata-kata pada teks dimasing-masing kategori kelas	25
Tabel 3. 9 Contoh data testing yang akan dicari kategori kelasnya	26
Tabel 3. 10 Proses perhitungan untuk class arts and humanities	26
Tabel 3. 11 Proses perhitungan untuk class engineering and technology	26
Tabel 3. 12 Proses perhitungan untuk class life sciences and medicine	27
Tabel 3. 13 Proses perhitungan untuk class natural sciences	27
Tabel 3. 14 Proses perhitungan untuk class social sciences and management	28
Tabel 3. 15 Hasil dari perhitungan dari masing-masing kelas dengan mengalikan hasil pada setiap pecahan kata	28
Tabel 4. 1 Pengujian Black Box.....	45
Tabel 4. 2 Jumlah Data.....	46
Tabel 4. 3 Rincian Data.....	46
Tabel 4. 4 Confusion Matrix	47
Tabel 4. 5 Hasil Pengukuran Accuracy, Recall, Precision, F1 score	47
Tabel 4. 6 Judul yang memiliki tema yang tidak sesuai bidangnya.....	49

ABSTRAK

Garuda adalah portal akademik dan mesin pencari untuk mengakses artikel, jurnal dan publikasi serta karya ilmiah. Garuda telah resmi diluncurkan oleh kemenristekdikti pada tahun 2018, banyaknya artikel maupun jurnal dan penerbit dengan berbagai macam bidang ilmu menjadikan para pengakses lebih membutuhkan pemetaan klasifikasi sesuai bidang ilmu yang dicari, penelitian ini bertujuan untuk menyelesaikan masalah bagaimana membuat klasifikasi pada bidang ilmu dengan metode *Naive Bayes* untuk mengetahui indeks kategori bidang ilmu yang berada di portal Garuda, *Naive Bayes* menjadi satu diantara metode yang sering digunakan untuk melakukan klasifikasi data, metode ini memprediksi probabilitas keanggotaan *class*, untuk evaluasi pengukuran hasil kinerja metode tersebut menggunakan *Precision*, *Recall*, *F-1 Score*, dan *Accuracy*. Penelitian ini menggunakan 500 data sampel data Garuda yaitu dibagi menjadi 400 data *training* dan 100 data *testing* dari lima kategori bidang ilmu yaitu *Arts & Humanities*, *Engineering & Technology*, *Life Science & Medicine*, *Natural Sciences*, *Social Sciences & Management* dan masing-masing bidang ilmu mengambil 100 sampel. Dari percobaan yang dilakukan hanya mendapatkan nilai akurasi sebesar 0.44, *recall* 0.2, presisi 0.04 dan *F-1 Score* mendapat nilai 0.06, percobaan tersebut dilakukan beberapa kali dengan konsistensi hasil yang sama, hal tersebut menunjukkan bahwa metode *Naive Bayes* untuk penelitian kali ini belum dapat menghasilkan nilai akurasi yang diharapkan, faktor penyebabnya antara lain data yang digunakan kurang akurat dalam pelabelan sehingga nilai akurasi yang didapatkan masih jauh dari yang diharapkan.

Kata kunci: Garuda, *Naive Bayes*, Klasifikasi, *Precision*, *Recall*, *F-1 Score*, dan *Accuracy*.

ABSTRACT

Garuda is an academic portal and search engine for accessing articles, journals and publications and scientific works. Garuda has been officially launched by the Ministry of Research, Technology and Higher Education in 2018, the large number of articles and journals and publishers with various fields of science makes accessors more in need of classification mapping according to the field of science being sought, this study aims to solve the problem of how to make a classification in the field of science with the Naive Bayes method to determine the index of the category of science in the Garuda portal, Naive Bayes is one of the methods that is often used to classify data, this method predicts the probability of class membership, to evaluate the measurement of the performance of the method using Precision, Recall, F-1 Score, and Accuracy. This study uses 500 data samples of Garuda data, which are divided into 400 training data and 100 testing data from five categories of science fields, namely Arts & Humanities, Engineering & Technology, Life Science & Medicine, Natural Sciences, Social Sciences & Management and each science field takes 100 samples. From the experiments carried out only get an accuracy value of 0.44, recall 0.2, precision 0.04 and F-1 Score gets a value of 0.06, the experiment was carried out several times with the same consistency of results, this shows that the Naive Bayes method for this research has not been able to produce the expected accuracy value, the factors that cause it include the data used is less accurate in labeling so that the accuracy value obtained is still far from expected.

Keywords: Garuda, Naive Bayes, Classification, Precision, Recall, F-1 Score, and Accuracy.

BAB I

PENDAHULUAN

1.1. Latar Belakang

Garuda merupakan sebuah database akademik serta mesin pencari akademisi untuk peneliti di Indonesia, portal garuda ini awalnya adalah *indonesian publication index* (IPI) yang didirikan pada tahun 2011 oleh (IAES) dan kemudian pada tahun 2018 oleh kementerian Riset, Teknologi, dan Pendidikan Tinggi RI (RISTEKDIKTI) telah resmi mengakuisisi. Portal garuda ini adalah sebuah platform yang bermanfaat dalam lingkungan akademik untuk mengakses dan menemukan artikel, dalam jurnal akademik Indonesia, *repository*, arsip dan berbagai koleksi karya ilmiah lainnya. Agregator, pengindeks, dan layanan terdesentralisasi lainnya, seperti *ProQuest*, *EBSCOHost*, *IEEE Xplore*, *Ei Engineering Village*, *ScienceDirect*, *Scopus* dan *Web of Science*. memungkinkan peneliti menemukan artikel yang relevan berdasarkan pencarian mereka. Pada saat ini jumlah artikel terindeks garuda sebanyak 1.974.843 artikel dari 14.540 jurnal dan 183 *conferences*. Tujuan adanya garuda diharapkan menjadi sarana bagi para peneliti dalam berkomunikasi serta menghindari praktik plagiat, serta menjadi peningkatan standar kualitas publikasi di Indonesia. (Subroto dkk., 2014)

Dalam berjalannya waktu perkembangan publikasi dibedakan menjadi beberapa kategori bidang ilmu yang berbeda, setiap bidang ilmu tersebut memiliki beberapa ciri dan karakter keilmuan yang berbeda pula, dari beberapa kategori bidang tersebut meliputi lima kategori bidang ilmu utama yaitu bidang ilmu *Life Science & Medicine*, bidang ilmu *Social Sciences & Management*, bidang ilmu *arts & Humanities*, bidang ilmu *Engineering & Technology*, bidang ilmu *Natural Sciences*, namun untuk sekarang belum ada mekanisme yang efektif dalam proses pengklasifikasian semua artikel yang sesuai dengan kategori bidang ilmu tersebut secara cepat, efisien dan otomatis dalam garuda. Dan hal itu menjadi kelemahan bagi Garuda untuk sekarang ini.

Naive Bayes merupakan satu dari beberapa metode klasifikasi yang sering

digunakan dalam proses pengklasifikasian data. Ketika melakukan klasifikasi *Naive Bayes* memakai probabilitas sederhana yaitu dengan menghitung kumpulan data kemudian menjumlahkan campuran nilai dataset yang diberikan dengan frekuensi. Algoritma memakai *teorema Bayes* serta meringkas setiap atribut sebagai independen atau tidak signifikan terkait dengan manfaat yang didapatkan oleh nilai pada kelas variabel. pengertian lain menyatakan bahwa *Naive Bayes* sebagai klasifikasi berdasarkan probabilitas serta metode statistik yang dikembangkan oleh sarjana Inggris *Thomas Bayes*, yaitu memprediksi hasil masa depan berlandaskan kinerja masa lalu.(Putri dkk., 2021). Telah banyak penelitian yang menggunakan metode *Naive Bayes* yang mana berdasarkan beberapa peneliti *Naive bayes* mempunyai keunggulan, yaitu cara perhitungan yang lebih cepat serta memiliki kesederhanaan algoritma dengan hasil akurasi yang baik.

Dari penelitian sebelumnya yang menggunakan *Naive Bayes* dengan judul klasifikasi artikel berita bahasa Indonesia dengan *naive bayes classifier* dimana pada penelitian tersebut peneliti membuat rancangan sistem yang berupa aplikasi berbasis website yang dapat mengkategorikan artikel berbahasa Indonesia sesuai isi artikel tersebut secara otomatis. Aplikasi tersebut menggunakan fitur *N-Gram* dengan metode *Naive Bayes*. Hasil dari penelitian tersebut didapatkan kesimpulan yaitu *Naive Bayes* memiliki akurasi yang tinggi yaitu antara 88.7% sampai 94.7% dengan variasi rasio dataset *training* dan *testing* sebesar 80 : 20, 70 : 30, 60 : 40, 50 : 50. Serta Ngram juga memiliki hasil akurasi yang tinggi jika dibanding bigram.(Setiawan dkk., 2020).

Berdasarkan latar belakang diatas. Penelitian ini akan melakukan eksperimen mengenai bagaimana cara membuat klasifikasi mengenai bidang ilmu yang terdapat pada portal garuda sehingga seluruh publikasi di dalamnya dapat diklasifikasikan dengan rapi sesuai dengan bidang ilmu yang ditentukan

1.2 Rumusan Masalah

Bagaimana membuat klasifikasi judul artikel pada Garuda berdasarkan 5 bidang ilmu secara otomatis

1.3 Batasan Masalah

Untuk batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Klasifikasi pada penelitian ini hanya diambil pada 5 bidang ilmu menggunakan metode *Naive Bayes*
2. Dalam penelitian ini data yang digunakan hanya berdasarkan judul publikasi terindeks dari Garuda

1.4 Tujuan Tugas Akhir

Tujuan tugas akhir ini adalah merancang sebuah sistem untuk melakukan klasifikasi bidang ilmu pada publikasi terindeks Garuda dengan metode klasifikasi data mining menggunakan Metode *Naive Bayes*

1.5 Manfaat

Manfaat dari penelitian ini, memudahkan dalam pengklasifikasian artikel sesuai dengan bidang ilmu yang dicari pada publikasi terindeks dari lima bidang ilmu secara komputerisasi.

1.6 Sistematika Penulisan

Untuk sistematika penulisan yang digunakan oleh penulis dalam sebuah pembuatan laporan tugas akhir adalah:

BAB 1: PENDAHULUAN

Pada bagian bab ini peneliti menyampaikan latar belakang dari judul penelitian, rumusan masalah, batasan masalah, tujuan penelitian, serta sistematika pada penulisan.

BAB 2: TINJAUAN PUSTAKA DAN DASAR TEORI

Pada bagian bab ini merupakan bagian yang memuat tinjauan pustaka dari berbagai penelitian yang ada sebelumnya yang berkaitan dengan penelitian

yang akan dibuat serta dasar teori yang membantu peneliti dalam membuat penelitian ini sehingga dengan dasar teori ini dapat lebih muda dalam memahami konsep kerja sebuah teori yaitu teori klasifikasi dengan menggunakan *Naive Bayes*

BAB 3: METODE PENELITIAN

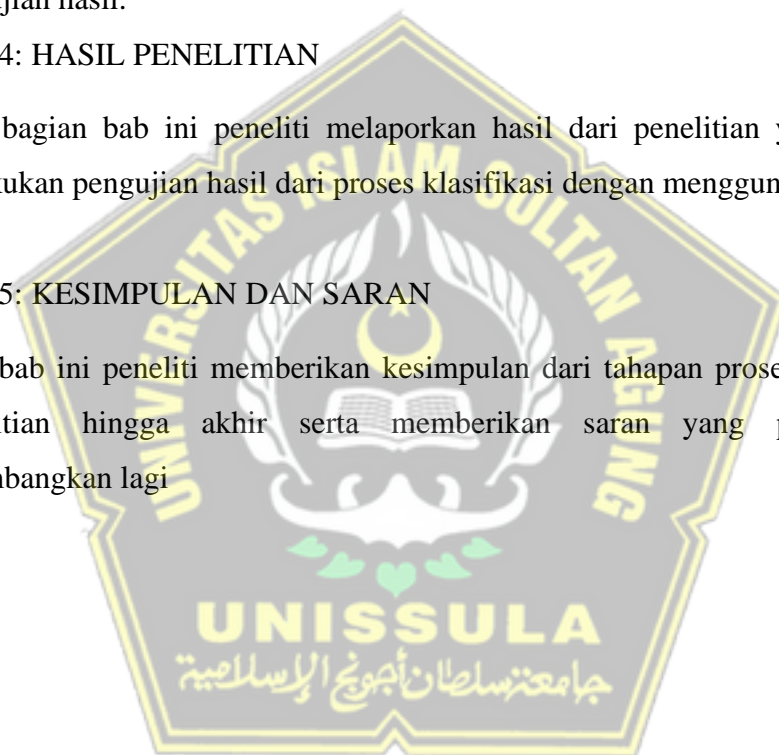
Pada bab ini merupakan pemaparan proses dari sebuah tahapan-tahapan penelitian, mulai dari pengambilan *dataset*, perancangan sistem, serta pengujian hasil.

BAB 4: HASIL PENELITIAN

Pada bagian bab ini peneliti melaporkan hasil dari penelitian yaitu dengan melakukan pengujian hasil dari proses klasifikasi dengan menggunakan *web*

BAB 5: KESIMPULAN DAN SARAN

Pada bab ini peneliti memberikan kesimpulan dari tahapan proses sejak awal penelitian hingga akhir serta memberikan saran yang perlu untuk dikembangkan lagi



BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Peneliti mengambil berbagai penelitian sebelumnya yang digunakan sebagai bahan pertimbangan serta referensi yang berkaitan dengan judul yang akan dibuat pada penelitian ini adalah sebagai berikut:

Dari penelitian sebelumnya yaitu sebuah penelitian yang melakukan perbandingan kinerja metode *Naive Bayes* dan *K-Nearest Neighbor* pada artikel berbahasa Indonesia, dimana pada penelitian tersebut peneliti membuat sebuah sistem yang mampu membagi kelompok artikel dengan akurat secara otomatis, peneliti menggunakan *Naive Bayes* Dan KNN sebagai metodenya dan pada penelitian tersebut peneliti akan membandingkan hasil dari masing-masing metode klasifikasi tersebut. Dan dari hasil penelitian tersebut menunjukkan bahwa hasil dari *Naive Bayes* menghasilkan nilai akurasi yang unggul sebesar 70 % dibandingkan dengan KNN yang menghasilkan akurasi lebih rendah yaitu sebesar 40 %.(Devita dkk., 2018).

Pada penelitian yang lain yaitu sebuah penelitian yang berjudul klasifikasi karya ilmiah (Tugas Akhir) mahasiswa menggunakan metode *Naive Bayes Classifier* (NBC). Pada penelitian tersebut peneliti membuat sebuah sistem yang mampu melakukan klasifikasi pada karya ilmiah dari mahasiswa dengan bidang ilmu Teknik Informatika, metode *naive bayes* dipilih pada penelitian ini dengan menggunakan lima kategori yang dibagi dari parameter latar belakang yaitu sistem pakar dan sistem informasi geografis, data mining, pengolahan citra dan sistem pengambilan keputusan. Kemudian jumlah dataset yang digunakan yaitu 170 data dengan rincian 150 untuk data *training* dan 20 untuk data *testing*. Dan hasil yang didapat dari penelitian tersebut naive bayes mendapat nilai rata-rata akurasi yang baik yaitu sebesar 86.68% dengan waktu pengujian rata-rata dibutuhkan 5,406 detik setiap pengujian.(Nuridin dkk., 2021)

Pada penelitian selanjutnya yaitu sebuah penelitian dengan judul klasifikasi

berita menggunakan metode *Naive Bayes Classifier*, pada penelitian tersebut peneliti membangun sistem yang mampu melakukan klasifikasi berita dengan mengelompokkannya sesuai kategori yang lebih spesifik, yang mana hal tersebut diharapkan dapat membantu bidang sosial Badan Pusat Statistik Provinsi Riau yang masih mengalami kendala dalam mencari dan mengelompokkan berbagai jenis berita mengenai Provinsi Riau. Pada penelitian ini menggunakan *Naive Bayes Classifier* sebagai metodenya, dengan menggunakan sampel data 510 berita kemudian membaginya menjadi tiga kategori yaitu tentang berita kemiskinan, kemudian kategori berita demokrasi dan terakhir kategori ketenagakerjaan. Kemudian pada tahap berikutnya dilakukan *preprocessing* data, kemudian dilakukan pembobotan kata dan setelah itu dilakukan penerapan metode. Dari penelitian tersebut didapatkan sebuah hasil yang menunjukkan bahwa metode *Naive Bayes* menghasilkan nilai akurasi yang baik yaitu didapatkan nilai akurasi sebesar 94% dengan pembagian data uji sebesar 10 % serta data latih 90%.(Habib dkk., 2022)

Pada penelitian berikutnya yaitu penelitian yang berjudul penerapan metode *Naive Bayes Classifier* pada sistem pengklasifikasian berita secara otomatis berbasis website (studi kasus: berita lokal dari media massa online Kalimantan Barat). Pada penelitian tersebut peneliti membuat sistem klasifikasi berita pada portal media massa berbasis online yang berada di daerah tersebut. Dimana sebelumnya dalam melakukan klasifikasi pada berita-berita pada media massa tersebut masih menggunakan cara manual. Pada kasus ini peneliti menggunakan klasifikasi *Naive Bayes* dengan sampel data yang berjumlah 18.794 berita, kemudian data dibagi menjadi data *training* sebanyak 936 data dan data *testing* sebanyak 405 data dengan perbandingan 70:30 untuk pengujian performa. Dari hasil uji tersebut mendapatkan hasil akurasi dengan nilai 98,90% dari hasil pengujian 9 kelas. 9 kelas tersebut meliputi tema politik dan pemerintahan sebanyak 4.821, kemudian tema berita pariwisata, olahraga dan event dengan jumlah 2.475 berita, kemudian tema pemilu dengan jumlah 2.254 berita, kemudian tema tentang kriminalitas serta hukum dengan jumlah 1.885 berita, kemudian tema tentang pendidikan, yang berjumlah 1.561 berita, kemudian

tema transportasi serta lalu-lintas yang berjumlah 1.400 berita, kemudian tema bencana alam yang berjumlah 1.179 berita, kemudian tentang tema kesehatan, memiliki data dengan jumlah 663 berita, kemudian yang terakhir tentang tema narkoba dengan jumlah 488 berita. (Rasila dkk., 2019).

Kemudian pada penelitian sebelumnya yang berjudul perbandingan *K-Nearest Neighbour* dan *Naive Bayes Classification* dalam membangun sistem rekomendasi berita olahraga online berbasis web, dimana pada penelitian tersebut peneliti melakukan uji perbandingan hasil dari pembuatan sistem klasifikasi rekomendasi berita olahraga online menggunakan metode KNN dan *Naive Bayes* pada penelitian tersebut menggunakan sebanyak 2500 judul berita olahraga sebagai dataset yang dibagi menjadi 9 kelas yaitu untuk sepak bola yang terdiri dari 4 kelas Negara Spanyol, Italia, Indonesia, Inggris, kemudian kelas esport, motogp, *F1*, bulu tangkis serta kelas basket. Dari pengujian dua algoritma tersebut menggunakan *training data* dengan jumlah 2252 data serta *testing data* dengan jumlah 248. Dan dari hasil penelitian tersebut *Naive Bayes* mendapatkan akurasi yang lebih unggul yaitu sebesar 95,161% dibanding KNN yang hanya menghasilkan akurasi sebesar 84,677%. (Seran dkk., 2021)

Pada penelitian yang lain dengan judul algoritma *Naive Bayes* dalam menentukan klasifikasi produk pada penjualan pulsa, pada penelitian tersebut peneliti membuat sistem yang dapat mengklasifikasikan produk yang laris dan tidak laris pada konter RA *Cell* dalam konter tersebut tersedia bermacam merek provider pulsa seperti Telkomsel, XL, Indosat dan 3, *Naive Bayes* menjadi metode yang digunakan untuk penelitian kali ini dengan menggunakan sampel data sebanyak 600 data, yang kemudian dilakukan pembagian data menjadi data latih sebanyak 480 serta data uji 120. Hasil yang didapat dari penelitian tersebut algoritma *Naive Bayes* mendapatkan produk terlaris yaitu pulsa Telkomsel dengan nilai akurasi 97,50%, kemudian Precision mendapat nilai 100%, dan selanjutnya *Recall* mendapat hasil nilai 93,48%. Dari hasil tersebut dapat disimpulkan bahwa metode *Naive Bayes* menjadi metode yang tepat dan baik untuk penelitian ini. (Wahyu Hadikristanto, 2020)

Pada penelitian yang lain yaitu sebuah penelitian dengan klasifikasi berita *hoax* menggunakan metode *naive bayes*, dalam penelitian tersebut penulis melakukan sebuah penelitian dikarenakan banyaknya berita *hoax* yang sering beredaran di sosial media internet, dan tidak bisa dipungkiri bahwasanya berita *hoax* muda sekali tersebar di internet terlebih sekarang di era digital yang mana teknologi informasi sangat berkembang pesat, dari permasalahan tersebut peneliti mencoba untuk membuat suatu *smart system* dengan kemampuan mengklasifikasikan berbagai berita tidak benar yang beredaran di sosial media, pada penelitian tersebut dilakukan beberapa tahapan antara lain yaitu *preprocessing*, pembobotan kata dengan menggunakan metode *naive bayes*, untuk pengukuran menggunakan metode *10-fold cross validation* dan mendapatkan hasil pada nilai *fold 6* memiliki tingkat akurasi yang paling tinggi sebesar 85,28 % yaitu dengan dokumen yang terklasifikasi sebanyak 307 secara relevan dan sebanyak 53 dokumen yang terklasifikasi tidak relevan atau menjadi sekitar 14.72%. dan berdasarkan dokumen berita *hoax* dan benar tersebut mendapat nilai rata-rata yaitu nilai *precision* sebesar 0,896 dan *recall* 0.853.(Mustofa & Mahfudh, 2019).

pada penelitian berikutnya yaitu sebuah penelitian dengan judul Penerapan Metode *Naive Bayes* Pada Klasifikasi Pelanggan. Pada penelitian tersebut penulis melakukan penelitian tentang bagaimana melakukan klasifikasi pelanggan potensial dan non potensial, hal itu dilakukan pemilik dalam memberikan bonus terhadap pelanggan, tujuannya yaitu sebagai strategi dalam meningkatkan kualitas kepada pelanggan yang potensial dalam membantu peningkatan penjualan, *Naive Bayes* menjadi metode yang dipilih pada penelitian ini. Aplikasi sistem dirancang menggunakan *PHP* dan *MySQL*, Dalam penelitian tersebut atribut yang digunakan untuk melakukan klasifikasi yaitu jumlah pembelian, interval waktu, dan lokasi. Hasil dari pengujian dari 25 sampel data mendapatkan hasil 23 sesuai dan 2 tidak sesuai, sehingga dari hasil tersebut berhasil mendapat prediksi pelanggan dengan jumlah nilai presisi 100%, *Recall* 91% dan akurasi 92%.(Putro dkk., 2020)

Pada penelitian yang lain yaitu sebuah penelitian yang berjudul

Perbandingan Kinerja Algoritma *Naive Bayes* dan *C.45* Pada Klasifikasi *Spam Email*, penulis melakukan penelitian tersebut dikarenakan banyaknya gangguan serangan balik *hacker*, *cracker* maupun virus, diantara platform yang sering mendapat gangguan ialah Email seperti misalnya banyaknya spam email yang masuk, tentunya hal itu sangat mengganggu bagi pengguna email, untuk itu peneliti berusaha membuat sebuah alat bantu berupa sistem yang dapat membantu dalam menentukan apakah pesan tersebut berupa spam atau non spam, dalam penelitian tersebut menggunakan metode *Naive Bayes* dan *C.45* yang mana nantinya tingkat keakuratannya akan dibandingkan dari dua metode tersebut mana yang paling mendapat hasil nilai tertinggi. Dan setelah dilakukan pengujian akhirnya mendapatkan hasil yaitu sebesar 96.70% untuk *Naive Bayes* yaitu dengan rincian data yang berjumlah 3385 data untuk prediksi Ham dan True Ham, kemudian 165 jumlah data untuk prediksi *Ham* dan *True Spam*, serta sejumlah 0 data untuk prediksi Spam dan True Spam. Sedangkan algoritma *C.45* mendapatkan nilai akurasi sebesar 96.68% yang mana sejumlah 3385 data untuk prediksi *Ham* dan *Tue Ham*, kemudian sejumlah 0 data untuk prediksi *Ham* dan *True Spam*, dan sejumlah 1447 data untuk prediksi *Spam* dan *Tue Spam*. Dan dari hasil perbandingan tersebut dapat disimpulkan *Naive Bayes* menghasilkan tingkat kinerja yang lebih unggul dibandingkan algoritma *C.45* pada penelitian ini. (Sulaeman dkk., 2022)

pada penelitian yang lain yaitu penelitian dengan judul Penerapan Algoritma *Naive Bayes* Pada Klasifikasi Konten Twitter dengan Indikasi Depresi, yang mana depresi merupakan bagian dari penyakit yang memiliki banyak pengaruh bagi penderitanya. Beberapa hal yang menjadi penyebab depresi banyak sekali seperti pekerjaan, kehidupan sosial dan lain-lain, sebanyak 6.1% dari total seluruh penduduk Indonesia yang berjumlah 267.7 juta jiwa diperkirakan mengalami tekanan depresi pada tahun 2018. Hal itu sangat mempengaruhi pandangan masyarakat tentang issue mental health yaitu dengan rendahnya tingkat kesadaran masyarakat tentang bahayanya penyakit kejiwaan, di era gempuran perkembangan teknologi yang sangat pesat terutama media sosial yang menjadi platform bagi masyarakat dalam menyalurkan ekspresinya melalui

berbagai konten yang mereka buat. Pada penelitian kali ini peneliti mengumpulkan data yang menjadi indikasi gangguan depresi dari platform media sosial Twitter dalam penelitian ini juga melibatkan psikiater untuk membuat pelabelan dataset mengenai apakah ini konten yang “Memiliki Indikasi Depresi” dan “Tidak memiliki indikasi” kemudian data tersebut dikembangkan menjadi prediktif dengan menggunakan metode *Complement Naïve Bayes* (CNB) dan *Multinomial Naïve Bayes* (MNB) untuk metode klasifikasi sedangkan metode ekstraksi fitur menggunakan metode *Term Frequency–Inverse Document Frequency* (TF–IDF). Hasil yang didapatkan dari percobaan yang telah dijalankan dengan menggabungkan MNB dan metode TF–IDF berhasil mendapat nilai *F-score* dengan skor 91.30% sedangkan penggabungan CNB dengan metode TF–IDF menghasilkan tingkat capaian performa dengan skor 91.98%.(Budiman dkk., 2021).

2.2 Dasar Teori

2.2.1 Garuda

Portal Garuda Indeks Publikasi Indonesia merupakan database akademik dan mesin pencari akademisi untuk peneliti di Indonesia, portal ini berdiri dari sebuah diskusi beberapa tahun terakhir antara *Institute of Advanced Engineering and Science* (IAES) dan pada tahun 2011 resmi diluncurkan, dimana mereka menghasilkan sebuah ide gagasan untuk mendirikan jurusan database akademik dan mesin pencari. Indeks Publikasi Indonesia (IPI) atau sebelumnya dikenal sebagai Portal Garuda Indeks Publikasi Indonesia (Portal Garuda IPI) kemudian pada tahun 2018 oleh Kementerian Riset, Teknologi dan Pendidikan Tinggi RI mengakuisisinya. Dimana saat ini Portal Garuda adalah sebuah platform yang bermanfaat dalam lingkungan akademik untuk mengakses dan menemukan artikel, dalam jurnal akademik Indonesia, repository, arsip dan berbagai koleksi karya ilmiah lainnya.

Agregator, pengindeks, dan layanan terdesentralisasi lainnya, seperti *ProQuest*, *EBSCOHost*, *IEEE Xplore*, *Ei Engineering Village*, *ScienceDirect*, *Scopus* dan *Web of Science*. memungkinkan peneliti menemukan artikel yang relevan berdasarkan pencarian mereka. Sebagian besar artikel yang

dikumpulkan diperlukan untuk mendapatkan *recall* yang tinggi dan dalam mendapatkan presisi yang tinggi maka diperlukan metode pengindeksan yang tinggi. Banyaknya artikel dan penerbit yang sulit ditemukan di internet dengan cepat dalam pencarian data menjadi masalah dalam pengumpulan data sehingga mesin pencari otomatis menjadi kebutuhan yang sangat penting sehingga dalam pencarian data tidak perlu menggunakan cara manual dengan mencari satu persatu. Untuk itu IPI dicancang untuk menjelajah, memantau mengabstraksi dan meningkatkan standar publikasi Indonesia. Pada saat ini jumlah artikel terindeks Garuda sebanyak 1.974.843 artikel dari 14.540 jurnal dan 183 *conferences*. Dalam hal interface pencarian IPI memiliki sejumlah pilihan pencarian. Ini diakses melalui Tab di bagian atas dari layar. Dengan menggunakan ini dimungkinkan untuk memilih antara "Pencarian Sederhana", "Pencarian Judul", "Pencarian Penulis", Opsi "Penelusuran Jurnal" dan "Penelusuran Penerbit".(Subroto dkk., 2014). Untuk penelitian kali ini penyusun akan menggunakan data judul artikel jurnal publikasi yang terdapat dalam portal Garuda dari data tersebut akan diproses melalui tahapan data mining

2.2.2 Data Mining

Data mining merupakan gabungan dari beberapa bidang ilmu seperti database, statistika, *machine learning* dan sebagainya, dari penggabungan tersebut didapatkan hasil pengetahuan baru yaitu berupa sebuah proses *data mining*, yang mana dengan data mining dapat digunakan dalam mencari dan menggali sebuah informasi yang sulit dan masih tersembunyi dalam sebuah *database* yang termasuk dalam tahapan proses *Knowledge Discovery in Database*.(Sukarna dkk., 2021)

Knowledge Discovery in Database (KDD) merupakan mekanisme yang dapat didefinisikan sebagai proses organisasi untuk mengidentifikasi dan menemukan pola secara andal dan berguna dari data mentah serta rumit yang terkumpul.(Simanjuntak dkk., 2021). Dalam *data mining* sebelum data diolah lebih lanjut maka terdapat beberapa proses diantaranya pengolahan data dalam

pengolahan data menggunakan tahapan *preprocessing*.

a. *Preprocessing*

Preprocessing adalah tahapan penting dalam *data mining* dan merupakan proses pertama yang terjadi sebelum dimulainya proses pelatihan dan pengujian dan mengubah data masukan menjadi data yang diformat dengan benar dan siap untuk diproses. *Preprocessing* mencakup berbagai operasi. Termasuk menggabungkan, mengubah, atau mengubah data untuk menghapus, menggabungkan, mengurangi, dan memperkirakan. Ini sering dilakukan dalam meminimalkan kesalahan atau kesalahan pada data mentah sebelum dilakukan proses ketahap berikutnya.(Ailmi dkk., 2020).

Langkah-langkah *preprocessing* yaitu :

1. *Cleaning* merupakan proses menghilangkan karakter tanda baca pada teks seperti angka, username hashtag, url, mention, dan simbol
2. *Tokenizing* merupakan proses Proses memecah rangkaian kalimat-kalimat tersebut menjadi kata
3. *Stopword* merupakan proses menghilangkan beberapa kata yang tidak relevan dan mengambil kata yang berguna contoh kata yang dibuang adalah “yang”, “dan”, “di”, dan “dari”.
4. *Stemming* merupakan proses mengubah kata yang memiliki imbuhan menjadi kata dasar caranya yaitu dengan membuang imbuhan yang terdapat dalam sebuah kata pada teks.
5. *Case Folding* bagian dari *preprocessing* yang berfungsi dalam merubah kata dari sebuah kalimat sehingga menjadi huruf kecil

Ketika data telah diolah dan dibersihkan menjadi data yang valid maka tahap selanjutnya penulis menentukan metode yang akan digunakan dalam mengolah data lebih lanjut,ada beberapa metode di dalam data mining, salah satunya adalah klasifikasi

b. Klasifikasi

Klasifikasi merupakan suatu metode yang digunakan untuk melakukan pengelompokan benda ataupun objek berdasarkan ciri-ciri yang dimiliki benda tersebut. Banyak cara dalam melakukan proses klasifikasi seperti cara manual atau menggunakan teknologi. Proses manual merupakan proses yang dilakukan sendiri oleh manusia tanpa menggunakan teknologi sedangkan klasifikasi yang berbasis teknologi adalah klasifikasi yang prosesnya menggunakan alat bantu seperti algoritma beberapa algoritma yang biasa digunakan adalah seperti *Naive Bayes*, *Fuzzy*, *Support Vector Machine* serta Jaringan Syaraf Tiruan. (Wibawa dkk., 2018).

Terdapat beberapa metode klasifikasi yang sering digunakan. Dalam penelitian ini penulis menggunakan metode algoritma *Naive Bayes Classification*

c. *Naive Bayes*

Naive bayes adalah bagian dari metode yang digunakan dalam proses klasifikasi. Probabilitas dan statistik adalah proses pengklasifikasian dari metode *Naive Bayes* yang memprediksi probabilitas keanggotaan *class*. *Thomas Bayes* merupakan ilmuwan Inggris yang mengemukakan *Teorema*, yakni dengan memprediksi suatu peluang dimasa yang akan datang berdasarkan pengalaman kinerja di masa lampau. Kombinasi antara *Teorema Bayes* dan *Naive* menjadikan setiap atribut bersifat bebas tidak bergantung dan berdiri sendiri. Metode tersebut juga bisa diterapkan secara efisiensi dalam pembelajaran yang terawasi (*supervised learning*). Manfaat klasifikasi adalah bahwa hanya sejumlah kecil data uji yang jarang diperlukan untuk memverifikasi parameter (sarana dan variasi variabel) yang dibutuhkan dalam klasifikasi. Karena asumsi bahwa setiap variabel adalah independen, hanya variasi dari kelas individu yang harus dipertimbangkan daripada semua rumus matematika kovarian. Dalam perjalanannya, *Naive Bayes* menganggap bahwa ada tidaknya ciri tertentu pada kelas tertentu tidak ada hubungannya dengan ada tidaknya ciri serupa pada kelas lain. Ketika proses klasifikasi, pendekatan

Bayes akan menghasilkan label kategori dengan probabilitas tertinggi Prinsip *Bayes* adalah sebuah teori.(Handoko & Neneng, 2021).

Rumus *Naive Bayes*

$$P(w_i|C) = \frac{\text{count}(w_i, C) + 1}{\text{count}(C) + |V|} \quad (1)$$

C = *Class*

D = Dokumen

W_i = kata ke i

(w_i,c) = jumlah kata w_i dalam C

Count (C) jumlah kata di class C

| V | =jumlah *vocabulary*

Untuk implementasi metode *naive bayes* peneliti membutuhkan *library PHP* dari *rubix ml* dimana *library PHP* tersebut telah mengakomodir berbagai jenis algoritma machine learning termasuk *Naive Bayes*

d. *Rubix ML*

Rubix ML adalah sebuah *library PHP* yang memfasilitasi dan menyediakan bagi penggunanya dalam belajar maupun membangun sebuah sistem mesin learning. *Rubix ML* sendiri memiliki banyak fasilitas dan *tools* yang disediakan bagi para penggunanya mulai dari membaca data, *training*, *preprocessing*, *testing*, *validation* serta *production* lebih dari 40 algoritma *supervised* maupun *unsupervised*. Untuk *Rubix ML* hanya dapat digunakan dengan minimal *PHP 7.2*. setelah melaksanakan implementasi metode *Naive Bayes* dengan *library Rubix ML* kemudian melakukan proses evaluasi dengan *confusion matrix* di *Rubix ML*

e. Evaluasi

Evaluasi merupakan proses yang digunakan dalam menetapkan nilai kegunaan dari model yang telah sukses dibangun pada tahap sebelumnya.(Sutoyo, 2018)

Evaluasi dilakukan untuk menentukan sebuah keputusan dari sebuah objek atau kegiatan penelitian yang telah dilakukan untuk memperbaiki hasil dari proses penelitian sebelumnya tentunya dengan bersandar pada data serta

sumber informasi yang telah kumpulkan. Evaluasi klasifikasi dengan metode *Naive Bayes* menggunakan empat bentuk evaluasi yaitu pertama *Accuracy*, kemudian *Precision*, selanjutnya *Recall* dan *F-1 score*, dan evaluasi tersebut digunakan dalam mengukur kinerja sistem untuk menghasilkan tingkat prediksi yang baik. Pada evaluasi tersebut menggunakan pengukuran yang dilakukan pada *confusion matrix* dimana pada rumus precession TP merupakan nantinya prediksi positif yang dijumlahkan semua dan benar positif. Dan TN atau benar negatif merupakan jumlah yang diprediksi salah. FP atau false negative saat diprediksi ternyata memiliki nilai positif.

- a. *precession* adalah rasio jumlah dokumen yang dianggap memenuhi kebutuhan seorang peneliti informasi. Akurasi menunjukkan seberapa akurat sistem memberikan informasi yang sebelumnya diminta oleh pengguna. Semakin tinggi nilai akurasi maka semakin tinggi pula kualitas metode yang digunakan dalam sistem. Rumus presisi:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

- b. *Recall* adalah persentase dari keseluruhan *file* yang dapat dipulihkan melalui proses penemuan dalam sistem.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

- c. *F-1 score* merupakan perbandingan antara rata-rata *precision* dan *recall* yang dibobotkan.

$$F-1score = 2 * \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

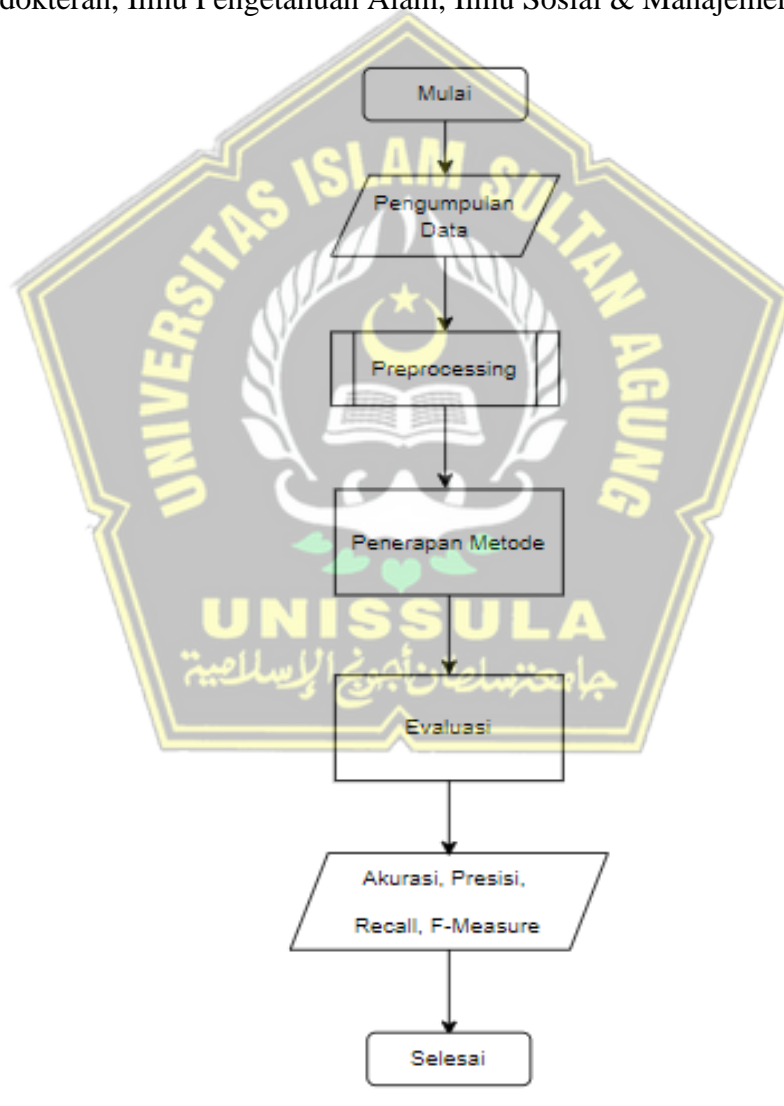
- d. *Accuracy* adalah rasio prediksi yang benar (negatif dan positif) disemua data. Akurasi memberikan jawaban untuk pertanyaan yang benar diprediksi dan salah prediksi

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (5)$$

BAB III METODE PENELITIAN

3.1 Tahapan Penelitian

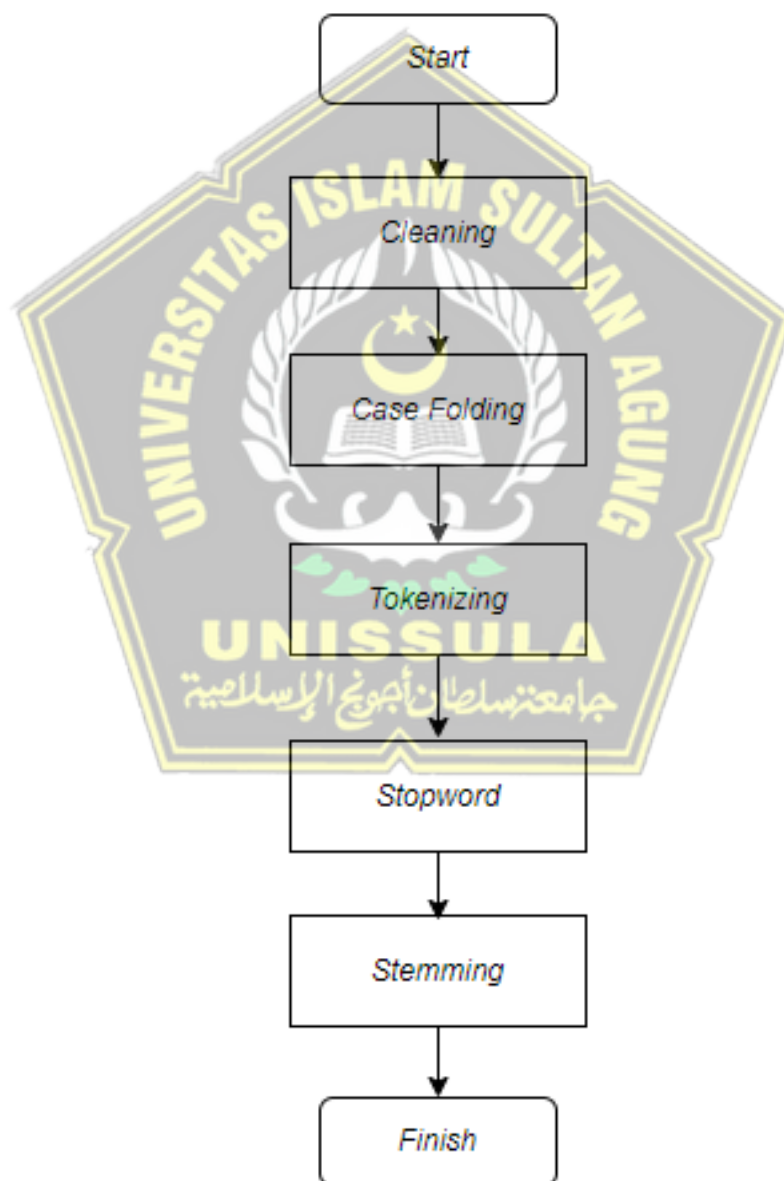
Adapun pada tahap penelitian ini peneliti menggunakan algoritma *Naive Bayes* dengan menggunakan data berupa judul artikel jurnal publikasi dari database Garuda yang mana data tersebut berupa judul artikel yang akan diklasifikasikan menjadi lima bidang ilmu yaitu seni & Humaniora, Ilmu Hayati & Kedokteran, Ilmu Pengetahuan Alam, Ilmu Sosial & Manajemen.



Gambar 3. 1 Skema tahapan proses penelitian

Gambar 3.1 merupakan skema tahapan proses penelitian dimana dari tahapan

penelitian ini yaitu tahap pertama dengan memulai pengambilan data garuda dari, setelah mendapatkan data tersebut kemudian melakukan proses validasi yang dilakukan oleh tim pengembang website garuda, setelah data tervalidasi kemudian melakukan preprocessing setelah melakukan *preprocessing* kemudian mengolah data menggunakan metode algoritma *Naive Bayes*, setelah data diolah kemudian akan dilakukan evaluasi guna mendapatkan hasil berupa akurasi, presisi dan *recall*.



Gambar 3. 2 Tahapan *preprocessing*

Gambar 3.2 merupakan tahapan *preprocessing* dimana dalam preprocessing memiliki lima tahapan yaitu *cleaning* yang merupakan proses proses menghilangkan karakter tanda baca pada teks, kemudian *case folding* yang berfungsi untuk merubah kata dalam kalimat sehingga menjadi huruf kecil, kemudian Proses memecah rangkaian kalimat-kalimat tersebut menjadi kata, kemudian *stopword* merupakan proses menghilangkan beberapa kata yang tidak relevan dan mengambil kata yang berguna contoh kata yang dibuang adalah “yang”, “dan”, “di”, dan “dari”. kemudian terakhir adalah *Stemming* merupakan proses mengubah kata yang memiliki imbuhan menjadi kata dasar caranya yaitu dengan membuang imbuhan yang terdapat dalam sebuah kata pada teks.

3.1.1 Koleksi data Garuda

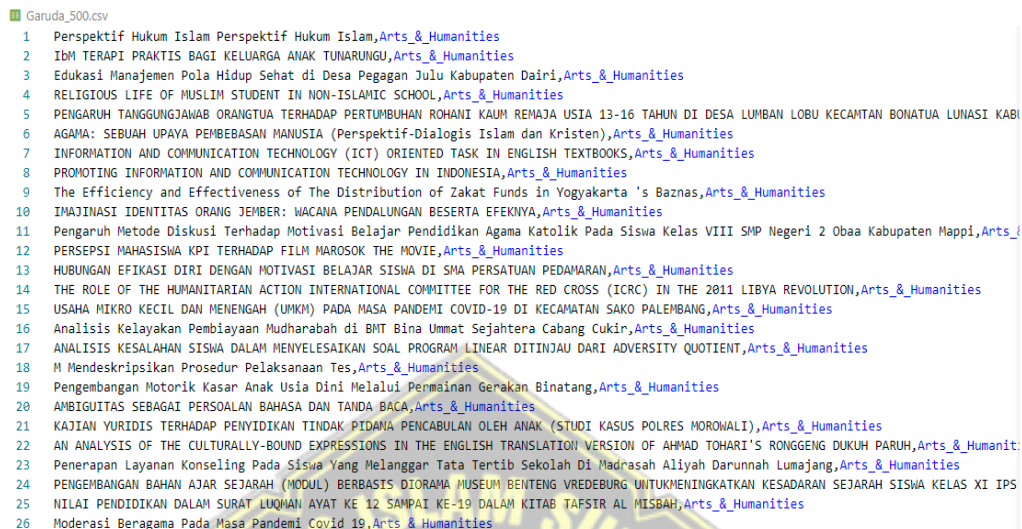
Pada penelitian ini penulis mengambil data yang bersumber dari website garuda, dengan mengambil 500 sampel data yang dibagi menjadi lima kategori bidang ilmu yaitu *arts & Humanities* sebanyak 100 sample, *Engineering & Technology* sebanyak 100 sample, *Life Science & Medicine* sebanyak 100 sampel, *Natural Sciences* sebanyak 100 sample, *Social Sciences & Management* sebanyak 100 sampel. Dalam sumber tersebut telah tervalidasi oleh pakar tim pengembang Garuda.

Contoh sampel data judul artikel jurnal dari garuda yang akan di proses

Tabel 3. 1 Contoh sampel data judul dari garuda

NO	DATA	BIDANG ILMU				
		<i>Arts & Humanities</i>	<i>Engineering & Technology</i>	<i>Life Science & Medicine</i>	<i>Natural Sciences</i>	<i>Social Sciences & Management</i>
1	Pendidikan karakter peduli lingkungan (studi kasus di sekolah menengah Assalihinayah Thailand)					
2	Transformasi Sampah Plastik Berbasis Destilasi Panel Surya					
3	Potensi Hasil dan Toleransi Galur-galur Inbrida Sorgum pada Tanah dengan Hara Fosfor Rendah					
4	Kerapatan dan Tutupan Kanopi Ekosistem Mangrove di Desa Wisata Pare Mas Lombok Timur					
5	Politik Adaptasi Perubahan Iklim dalam Pendekatan Multilevel <i>Governance</i> di Kota Semarang					

Sampel data 500 judul artikel jurnal garuda dari lima bidang ilmu yang akan digunakan dalam proses klasifikasi.



```

1 Perspektif Hukum Islam Perspektif Hukum Islam,Arts_& Humanities
2 IBM TERAPI PRAKTIS BAGI KELUARGA ANAK TUNARUNGU,Arts_& Humanities
3 Edukasi Manajemen Pola Hidup Sehat di Desa Pegagan Julu Kabupaten Dairi,Arts_& Humanities
4 RELIGIOUS LIFE OF MUSLIM STUDENT IN NON-ISLAMIC SCHOOL,Arts_& Humanities
5 PENGARUH TANGGUNGJAWAB ORANGTUA TERHADAP PERTUMBUHAN ROHANI KAUM REMAJA USIA 13-16 TAHUN DI DESA LUMBAN LOBU KECAMATAN BONATUA LUNASI KABI
6 AGAMA: SEBUAH UPAYA PEMBEBASAN MANUSIA (Perspektif-Dialogis Islam dan Kristen),Arts_& Humanities
7 INFORMATION AND COMMUNICATION TECHNOLOGY (ICT) ORIENTED TASK IN ENGLISH TEXTBOOKS,Arts_& Humanities
8 PROMOTING INFORMATION AND COMMUNICATION TECHNOLOGY IN INDONESIA,Arts_& Humanities
9 The Efficiency and Effectiveness of The Distribution of Zakat Funds in Yogyakarta 's Baznas,Arts_& Humanities
10 IMAJINASI IDENTITAS ORANG JEMBER: WACANA PENDALUNGAN BESERTA EFEKINYA,Arts_& Humanities
11 Pengaruh Metode Diskusi Terhadap Motivasi Belajar Pendidikan Agama Katolik Pada Siswa Kelas VIII SMP Negeri 2 Obaa Kabupaten Mappi,Arts_
12 PERSEPSI MAHASISWA KPI TERHADAP FILM MAROSOK THE MOVIE,Arts_& Humanities
13 HUBUNGAN EFIKASI DIRI DENGAN MOTIVASI BELAJAR SISWA DI SMA PERSATUAN PEDAMARAN,Arts_& Humanities
14 THE ROLE OF THE HUMANITARIAN ACTION INTERNATIONAL COMMITTEE FOR THE RED CROSS (ICRC) IN THE 2011 LIBYA REVOLUTION,Arts_& Humanities
15 USAHA MIKRO KECIL DAN MENENGAH (UMKH) PADA MASA PANDEMI COVID-19 DI KECAMATAN SAKO PALEMBANG,Arts_& Humanities
16 Analisis Kelayakan Pembiayaan Mudharabah di BMT Bina Ummat Sejahtera Cabang Cukir,Arts_& Humanities
17 ANALISIS KESALAHAN SISWA DALAM MENYELESAIKAN SOAL PROGRAM LINEAR DITINJAU DARI ADVERSITY QUOTIENT,Arts_& Humanities
18 M Mendeskripsikan Prosedur Pelaksanaan Tes,Arts_& Humanities
19 Pengembangan Motorik Kasar Anak Usia Dini Melalui Permainan Gerakan Binatang,Arts_& Humanities
20 AMBIGUITAS SEBAGAI PERSOALAN BAHASA DAN TANDA BACA,Arts_& Humanities
21 KAJIAN YURIDIS TERHADAP PEVIDIKAN TINDAK PIDANA PENCABULAN OLEH ANAK (STUDI KASUS POLRES MOROWALI),Arts_& Humanities
22 AN ANALYSIS OF THE CULTURALLY-BOUND EXPRESSIONS IN THE ENGLISH TRANSLATION VERSION OF AHMAD TOHARI'S RONGGENG DUKUH PARUH,Arts_& Humanit
23 Penerapan Layanan Konseling Pada Siswa Yang Melanggar Tata Tertib Sekolah Di Madrasah Aliyah Darunnah Lumajang,Arts_& Humanities
24 PENGEMBANGAN BAHAN AJAR SEJARAH (MODUL) BERBASIS DIORAMA MUSEUM BENTENG VREDEBURG UNTUKMENINGKATKAN KESADARAN SEJARAH SISWA KELAS XI IPS
25 NILAI PENDIDIKAN DALAM SURAT LUQMAN AYAT KE 12 SAMPAI KE-19 DALAM KITAB TAFSIR AL MISBAH,Arts_& Humanities
26 Moderasi Beragama Pada Masa Pandemi Covid 19,Arts_& Humanities

```

Gambar 3. 3 Data sampel 500 judul artikel jurnal garuda

Pada gambar 3.3 merupakan sampel *dataset* judul artikel jurnal publikasi terindeks garuda dengan jumlah total 500 judul artikel dengan kategori lima bidang ilmu bidang ilmu tersebut meliputi *arts & Humanities* sebanyak 100 sampel, *Engineering & Technology* sebanyak 100 sample, *Life Science & Medicine* sebanyak 100 sampel, *Natural Sciences* sebanyak 100 sample, *Social Sciences & Management* sebanyak 100 sampel dimana datum tersebut telah divalidasi oleh pengembang garuda, kemudian data tersebut dibuat dengan format CSV untuk dimasukan kedalam sistem pemrosesan

3.1.2 Preprocessing

Pada tahapan preprocessing dataset yang digunakan merupakan data yang diambil dari portal garuda untuk proses preprocessing tersebut yaitu dengan menyeleksi artikel ilmiah yang terdapat didalam portal garuda yang kemudian dibagi menjadi lima bidang ilmu yaitu *arts & Humanities*, *Engineering & Technology*, *Life Science & Medicine*, *Natural Sciences*, *Social Sciences & Management*. Pada preprocessing beberapa tahapan proses yaitu:

```
// 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
  ])
);
```

Gambar 3. 4 Penerapan *preprocessing* pada kode program

Pada gambar 3.4 merupakan gambar yang menampilkan kode program dalam melakukan *preprocessing*

a) *Cleaning*

Contoh:

Tabel 3. 2 Contoh proses data *input* dan *output* pada *cleaning*

Data Input	Data Output
Pendidikan karakter peduli lingkungan (studi kasus di sekolah menengah Assalihinayah Thailand)	Pendidikan karakter peduli lingkungan studi kasus di sekolah menengah Assalihinayah Thailand

Adapun penerapan *cleaning* dalam kode program sebagai berikut:

```
new RegexFilter([
  RegexFilter::EXTRA_WHITESPACE,
  RegexFilter::EXTRA_WORDS
]);
```

Gambar 3. 5 Kode Program *Cleaning*

Pada gambar 3.5 merupakan bentuk kode program *text cleaning preprocessing* yang digunakan dalam memproses data seperti menghilangkan karakter tanda baca pada teks seperti angka, *username hashtag*, *url*, *mention*, dan simbol.

b) *Case folding*

Contoh:

Tabel 3. 3 Contoh Proses data *input* dan *output* pada *case folding*

Data <i>Input</i>	Data <i>Output</i>
Transformasi Sampah Plastik Berbasis Destilasi Panel Surya	transformasi sampah plastik berbasis destilasi panel surya

Adapun penerapan *case folding* dalam kode program sebagai berikut:

```

    }), //text cleaning
    new MultibyteTextNormalizer(), //case folding - lower case

```

Gambar 3. 6 kode Program *Case Folding*

Pada gambar 3.6 merupakan bentuk kode program *case folding* yang berfungsi untuk merubah kata yang terdapat pada kalimat sehingga menjadi huruf kecil

c) *Tokenizing*

Contoh:

Tabel 3. 4 Contoh proses data *input* dan *output* pada *tokenizing*

Data <i>Input</i>	Data <i>Output</i>
Potensi Hasil dan Toleransi Galur-galur Inbrida Sorgum pada Tanah dengan Hara Fosfor Rendah	Potensi, Hasil, dan, Toleransi, Galur-galur, Inbrida, Sorgum, pada, Tanah, dengan, Hara, Fosfor, Rendah,

Adapun penerapan *tokenizing* pada kode program sebagai berikut:

```

    new WordCountVectorizer(), //tokenizer

```

Gambar 3. 7 Kode Program *Tokenizing*

Pada gambar 3.7 merupakan bentuk kode program *Tokenizing* yang berfungsi untuk memecah rangkaian kalimat-kalimat tersebut menjadi kata

d) *Stopword*

Contoh:

Tabel 3. 5 Contoh proses data *input* dan *output* pada *stopword*

Data Input	Data Output
Kerapatan dan Tutupan Kanopi Ekosistem Mangrove di Desa Wisata Pare Mas Lombok Timur	Kerapatan Tutupan Kanopi Ekosistem Mangrove Desa Wisata Pare Mas Lombok Timur

Adapun penerapan *stopword* pada kode program sebagai berikut:

```
new StopWordFilter($stopwords_en), //stopwords
```

Gambar 3. 8 Kode Program *Stopwords*

Pada gambar 3.8 merupakan bentuk kode program *stopword* yang berfungsi menghapus kata-kata yang tidak relevan dan mengambil kata yang penting

e) *Stemming*

Contoh:

Tabel 3. 6 contoh proses data *input* dan *output* pada *stemming*

Data Input	Data Output
Politik Adaptasi Perubahan Iklim dalam Pendekatan Multilevel Governance di Kota Semarang	Politik Adaptasi bah Iklim dalam Dekat Multilevel Governance di Kota Semarang

Adapun penerapan *stemming* pada kode program sebagai berikut:

```
$dataset->apply(new LambdaFunction($stemminggris, 'stemmer'));
```

Gambar 3. 9 Kode Program *Stemming*

Pada gambar 3.9 merupakan bentuk kode program *stemming* yang berfungsi menghapus kata-kata yang tidak relevan dan mengambil kata yang penting

3.1.3 Penerapan Metode

Naive Bayes adalah satu diantara jenis-jenis metode klasifikasi data yang ada sekarang ini. *Naive Bayes* merupakan klasifikasi statistik, proses tersebut dimaksudkan untuk memprediksi probabilitas dari kelas yang ada. *Naive Bayes* menjadi metode klasifikasi yang sangat sederhana dengan memperkirakan klasifikasi atribut. Dalam menggunakan metode *Naive Bayes* Nilai Probabilitas dicari terlebih dahulu dan likelihood maksimum dari setiap atribut untuk masing-masing kelas

Adapun contoh perhitungannya yaitu:

Tabel 3. 7 Tabel pembagian contoh dataset yang dibagi menjadi data *training* dan *testing*

		Teks Judul	Class
<i>Training</i>	D1	Menuju paradigma pendidikan islam transformatif	<i>Arts And Humanities</i>
	D2	Percampuran deposito investasi dan pewarisan uang asuransi dalam perspektif hukum islam	
	D3	Sisi gender dalam pemikiran islam liberal nurcholish madjid	
	D4	Studi sistem pengisian cepat baterai kendaraan listrik berbasis papan pengendali openevse	<i>Engineering And Technology</i>
	D5	Redesain pasar garung wonosobo dengan pendekatan arsitektur modern	
	D6	Residu pestisida di sentra produksi padi di jawa tengah	<i>Life Sciences And Medicine</i>
	D7	Analisis nilai tambah kopi srikandi	
	D8	Indeks bahaya erosi pada lahan reklamasi	<i>Natural Sciences</i>
	D9	Pengaruh antioksidan terhadap degradasi plastik polipropilen	
	D10	Pembelajaran jarak jauh taruna sekolah tinggi kedinasan	<i>Social Sciences And Management</i>
	D11	Faktor volume jam perencanaan (faktor k)	
<i>Testing</i>	D12	Studi pemikiran pendidikan islam modern	?

Tabel 3. 8 Pemecahan kata-kata pada teks dimasing-masing kategori kelas

<i>Class</i>	Pecahan Kata	Jumlah Kata Per Class
<i>Arts And Humanities</i>	[menuju] [paradigma] [pendidikan] [islam 3] [transformatif] [percampuran] [deposito] [investasi] [dan] [pewarisan] [uang] [asuransi] [dalam 2] [perspektif] [hukum] [sisi] [gender] [pemikiran] [liberal] [nurcholish] [madjid]	21
<i>Engineering And Technology</i>	[studi] [sistem] [pengisian] [cepat] [baterai] [kendaraan] [listrik] [berbasis] [papan] [pengendali] [openevse] [redesain] [pasar] [garung] [wonosobo] [dengan] [pendekatan] [arsitektur] [modern]	19
<i>Life Sciences And Medicine</i>	[residu] [pestisida] [di 2] [sentra] [produksi] [padi] [jawa- tengah] [analisis] [nilai] [tambah] [kopi] [srikandi]	12
<i>Natural Sciences</i>	[indeks] [bahaya] [erosi] [pada] [lahan] [reklamasi] [pengaruh] [antioksidan] [terhadap] [degradasi] [plastik] [polipropilen]	12
<i>Social Sciences And Management</i>	[pembelajaran] [jarak] [jauh] [taruna] [sekolah] [tinggi] [kedinasan] [faktor] [volume] [jam] [perencanaan] [faktor k]	12
Jumlah Total Vocabulary		76

Dalam perhitungannya merujuk pada persamaan rumus di bawah ini

$$P(w_i|C) = \frac{\text{count}(w_i, C) + 1}{\text{count}(C) + |V|} \quad (6)$$

Tabel 3. 9 Contoh data *testing* yang akan dicari kategori kelasnya

<i>Data Testing</i>
Studi pemikiran pendidikan islam modern
[Studi] [pemikiran] [pendidikan] [islam] [modern]

1. *Class Arts And Humanities*

Tabel 3. 10 Proses perhitungan untuk *class arts and humanities*

Studi	$(0 + 1)/(21 + 76)$	0.0103092784
Pemikiran	$(1 + 1)/(21 + 76)$	0.0206185567
Pendidikan	$(1 + 1)/(21 + 76)$	0.0206185567
Islam	$(3 + 1)/(21 + 76)$	0.0412371134
Modern	$(0 + 1)/(21 + 76)$	0.0103092784
Jumlah		1,86320789e-9

2. *Class Engineering And Technology*

Tabel 3. 11 Proses perhitungan untuk *class engineering and technology*

Studi	$(1 + 1)/(19 + 76)$	0.0210526316
Pemikiran	$(0 + 1)/(19 + 76)$	0.0105263158
Pendidikan	$(0 + 1)/(19 + 76)$	0.0105263158
Islam	$(0 + 1)/(19 + 76)$	0.0105263158
Modern	$(1 + 1)/(19 + 76)$	0.0210526316
Jumlah		5,16942177e-10

3. *Class Life Sciences And Medicine*

Tabel 3. 12 Proses perhitungan untuk *class life sciences and medicine*

Studi	$(0 + 1)/(12 + 76)$	0.0113636364
Pemikiran	$(0 + 1)/(12 + 76)$	0.0113636364
Pendidikan	$(0 + 1)/(12 + 76)$	0.0113636364
Islam	$(0 + 1)/(12 + 76)$	0.0113636364
Modern	$(0 + 1)/(12 + 76)$	0.0113636364
Jumlah		1,89490153e-10

4. *Class Natural Sciences*

Tabel 3. 13 Proses perhitungan untuk *class natural sciences*

Studi	$(0 + 1)/(12 + 76)$	0.0113636364
Pemikiran	$(0 + 1)/(12 + 76)$	0.0113636364
Pendidikan	$(0 + 1)/(12 + 76)$	0.0113636364
Islam	$(0 + 1)/(12 + 76)$	0.0113636364
Modern	$(0 + 1)/(12 + 76)$	0.0113636364
Jumlah		1,89490153e-10

5. Class Social Sciences And Management

Tabel 3. 14 Proses perhitungan untuk *class social sciences and management*

Studi	$(0 + 1)/(12 + 76)$	0.0113636364
Pemikiran	$(0 + 1)/(12 + 76)$	0.0113636364
Pendidikan	$(0 + 1)/(12 + 76)$	0.0113636364
Islam	$(0 + 1)/(12 + 76)$	0.0113636364
Modern	$(0 + 1)/(12 + 76)$	0.0113636364
Jumlah		1,89490153e-10

Tabel 3. 15 Hasil dari perhitungan dari masing-masing kelas dengan mengalikan hasil pada setiap pecahan kata

<i>Class Arts And Humanities</i>	0.0103092784* 0.0206185567* 0.0206185567* 0.0412371134* 0.0103092784	1,86320789e-9
<i>Class Engineering And Technology</i>	0.0210526316* 0.0105263158* 0.0105263158* 0.0105263158* 0.0210526316	5,16942177e-10
<i>Class Life Sciences And Medicine</i>	0.0113636364* 0.0113636364* 0.0113636364* 0.0113636364* 0.0113636364	1,89490153e-10
<i>Class Natural Sciences</i>	0.0113636364* 0.0113636364* 0.0113636364* 0.0113636364* 0.0113636364	1,89490153e-10
<i>Class Social Sciences And Management</i>	0.0113636364* 0.0113636364* 0.0113636364* 0.0113636364* 0.0113636364	1,89490153e-10

Adapun penerapan metode *Naive Bayes* pada kode program sebagai berikut:

```

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

```

Gambar 3. 10 Proses penerapan metode di dalam *coding*

Pada gambar 3.10 merupakan gambar yang menampilkan kode program dalam melakukan proses penerapan metode *Naive Bayes*

3.1.4 Evaluasi

Pada penelitian ini evaluasi klasifikasi dengan metode *Naive Bayes* menggunakan empat jenis evaluasi yaitu *Precision* adalah rasio prediksi positif yang benar dibandingkan dengan hasil prediksi positif secara keseluruhan, *Recall* adalah rasio prediksi *true positive* dibandingkan dengan jumlah total data *true positive* pada keseluruhan data yang benar positif, *F-1 score* adalah salah satu perhitungan evaluasi dalam pencarian informasi yang menggabungkan tingkat *recall* dan tingkat presisi, dan *Accuracy* adalah rasio prediksi yang benar (positif dan negatif) terhadap total data. evaluasi tersebut dimaksudkan untuk mengukur kinerja sistem untuk menghasilkan tingkat prediksi yang baik. Pada evaluasi tersebut menggunakan pengukuran yang dilakukan pada *confusion matrix*, *confusion matrix* merupakan salah satu bentuk data mining berupa prediksi tipe klasifikasi yang juga merupakan metode pembelajaran terawasi, dengan label yang dapat digunakan sebagai acuan dalam mengukur kinerja suatu model, yang merepresentasikan bilangan aktif maupun ekstrinsik. Seiring dengan nilai yang diharapkan, *confusion matrix* adalah bagian dari keluarga pembelajaran mesin yang memeriksa data yang sudah tersedia dan mengklasifikasikannya sebagai data baru untuk membuat variabel keluaran (nominal atau pengurutan) kelas.

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

Gambar 3. 11 Proses evaluasi dengan *Confusion Matrix* di dalam coding

Pada gambar 3.11 merupakan tampilan kode program perhitungan *confusion matrix* yang meliputi akurasi, presisi, *recall* dan *f1 score*

3.2 Metodologi Perancangan Alur Sistem

3.2.1 Analisa Kebutuhan Fungsi Sistem

Pada bagian analisa kebutuhan sistem akan dianalisis mengenai apa saja yang harus bisa dilakukan sistem untuk melakukan proses klasifikasi data judul artikel sesuai lima bidang ilmu, untuk itu dalam sistem ini perlu adanya beberapa fungsi yang harus diterapkan pada sistem tersebut diantaranya:

1. Fungsi input judul

Merupakan bagian dari fungsi yang sangat perlu dimiliki oleh sistem ini, dimana fungsi tersebut nantinya akan menjadi proses yang harus dilewati oleh user dalam mendapatkan hasil akurasi yaitu dengan menginputkan judul artikel ke dalam kolom input judul yang telah tersedia.

2. Fungsi menu prediksi

Merupakan fungsi yang akan menjadi inti dari sistem ini yaitu sebuah bagian proses yang nantinya user akan melakukan input judul artikel yang sebelumnya telah dilakukan pelabelan dengan kategori lima bidang ilmu, kemudian sistem akan melakukan proses prediksi mengenai tingkat keakuratan dan kesesuaian dari judul tersebut dengan salah satu dari kategori lima bidang ilmu.

3. Fungsi Menampilkan Prediksi

Merupakan fungsi yang akan menampilkan hasil prediksi yang telah diproses sistem sebelumnya dimana ketika judul telah diproses oleh sistem maka akan menghasilkan hasil tampilan berupa output yang berisi judul artikel yang telah diinputkan sebelumnya beserta persentase akurasi dari lima bidang ilmu tersebut

4. *Download dataset*

Merupakan yang berfungsi untuk mendownload dataset yang ada pada sistem, data tersebut berupa judul artikel jurnal yang telah memiliki label dengan bidang ilmunya dalam bentuk tabel, user dapat mendownload dataset tersebut dengan mengklik tombol unduh *dataset* maka dataset tersebut secara otomatis akan terdownload.

3.2.2 Analisa Kebutuhan Sistem

Untuk tahap ini peneliti akan menganalisa beberapa komponen peralatan apa saja yang dapat mendukung dalam mengembangkan sistem ini, berikut adalah beberapa komponen yang diperlukan:

1. *Rubix ML*

Rubix ML merupakan *library PHP* yang digunakan dalam membuat sistem maupun aplikasi yang berbasis penggunaan mesin learning dalam *Rubix ML* terdapat banyak fasilitas yang sangat mendukung dalam pembuatan aplikasi sistem dengan berbagai macam *tools* seperti *training, testing, preprocessing, validation* serta *production* dengan lebih dari 40 jenis algoritma baik *supervised* maupun *unsupervised*.

2. *Visual studio Code*

Visual Studio Code adalah aplikasi code editor yang mampu dijalankan disemua desktop dan operasi sistem seperti windows, linux, mac, aplikasi ini dikembangkan oleh perusahaan microsoft, aplikasi ini menyediakan berbagai jenis bahasa pemrograman telah tersedia di dalam aplikasi ini antara lain Java, *JavaScript, Go, C++*, dan juga dengan kemampuannya yang sangat *powerfull* dalam mengedit berbagai macam kode program seperti *php* dan *python*, walaupun aplikasi ini terbilang sangat hebat namun *Visual Studio* mempunyai lintas *platform* kode *editor* yang ringan, dan mudah digunakan oleh semua kalangan dalam mengembangkan *web*.

3. *Bootstrap*

Bootstrap merupakan *CSS framework* yang berbasis *HTML, Javascript* dan *CSS* yang dikembangkan untuk membangun sebuah *website* yang *responsive*, *bootstrap* juga sangat populer bagi kalangan *developer* karena memudahkan mereka dalam mengembangkan *website*.

3.2.3 Analisa alur sistem

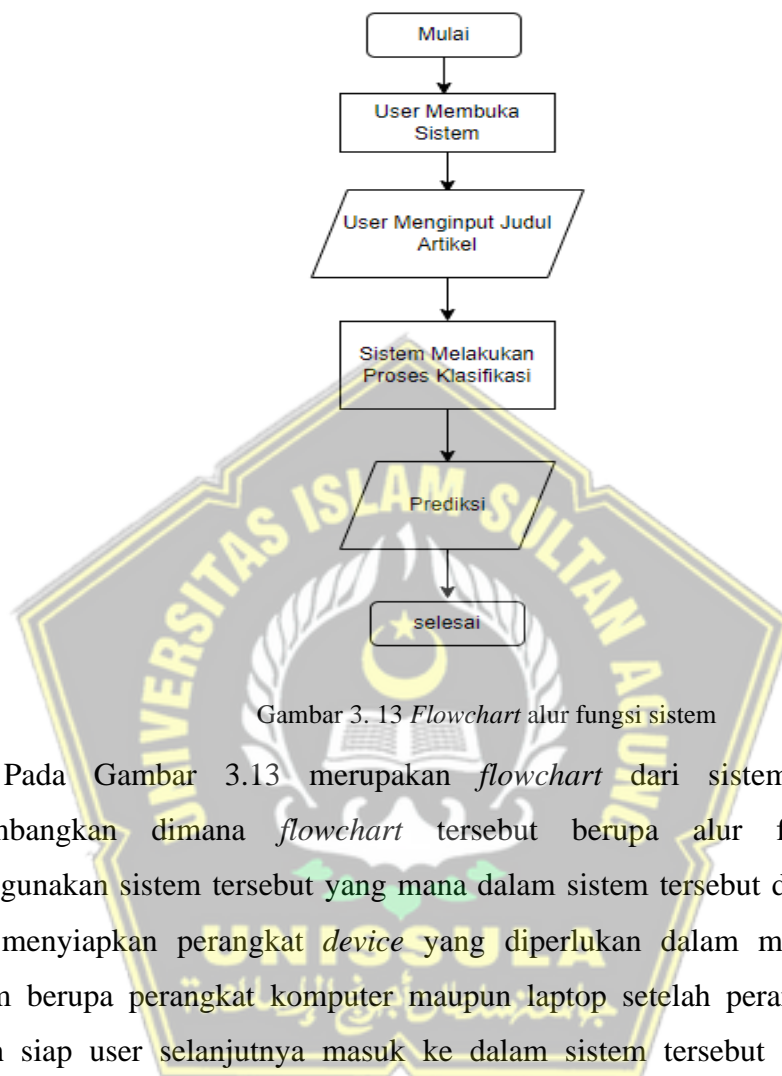
Pada Analisis alur sistem, akan dibuat sebuah *flowchart* yang menunjukkan alur perancangan dan sekaligus alur kerja dari sistem ini, dimana *flowchart* dari alur perancangan ini dapat dilihat pada gambar



Gambar 3. 12 *Flowchart* perancangan alur sistem

Pada Gambar 3.12 adalah gambar *flowchart* perancangan alur sistem dimana pada gambar tersebut menunjukkan beberapa tahapan yaitu yang dimulai dari proses pengumpulan data garuda kemudian mengolah data tersebut sehingga menjadi data yang valid dan divalidasi oleh pengembang dari *website* garuda tersebut, setelah itu dilakukan perancangan sistem dan melakukan implementasi

dalam membangun sistem tersebut dengan proses coding serta membuat user *interface* sistem tersebut kemudian dilakukan pengujian sistem.



Gambar 3. 13 *Flowchart* alur fungsi sistem

Pada Gambar 3.13 merupakan *flowchart* dari sistem yang akan dikembangkan dimana *flowchart* tersebut berupa alur fungsi dalam menggunakan sistem tersebut yang mana dalam sistem tersebut diawali dengan user menyiapkan perangkat *device* yang diperlukan dalam mengakses *web* sistem berupa perangkat komputer maupun laptop setelah perangkat tersebut sudah siap user selanjutnya masuk ke dalam sistem tersebut setelah sistem terbuka maka sistem tersebut akan menampilkan sebuah halaman dengan tiga menu bar berupa menu *Home*, *Dataset*, dan *Predict* untuk menjalankan sistem tersebut *user* perlu masuk pada menu *predict* kemudian user memasukan data berupa judul artikel yang dapat diambil dari menu *dataset* ke dalam kolom input yang terdapat di menu *predict* selanjutnya user mengklik tombol *predict* untuk sistem dalam melakukan proses prediksi dengan metode *Naive Bayes* setelah beberapa saat maka sistem akan mengeluarkan *output* hasil prediksi berupa judul artikel beserta akurasi di lima bidang ilmu. Setelah mendapatkan hasil prediksi maka proses tersebut selesai.

3.3 Perancangan *Interface*

Perancangan *interface* pada sistem yang dibuat dalam penelitian kali ini adalah berupa desain *mockup* yang nantinya akan diimplementasikan dalam pembuatan sistem tersebut. Berikut ini adalah rancangan *interface* dari program Klasifikasi Bidang Ilmu Pada Publikasi Terindeks Garuda Menggunakan Metode *Naive Bayes* berbasis *web*:

1. Tampilan *Menu Bar*

Merupakan rancangan tampilan menu-menu yang berada di samping kiri dari isi halaman:



Gambar 3. 14 Rancangan tampilan menu bar

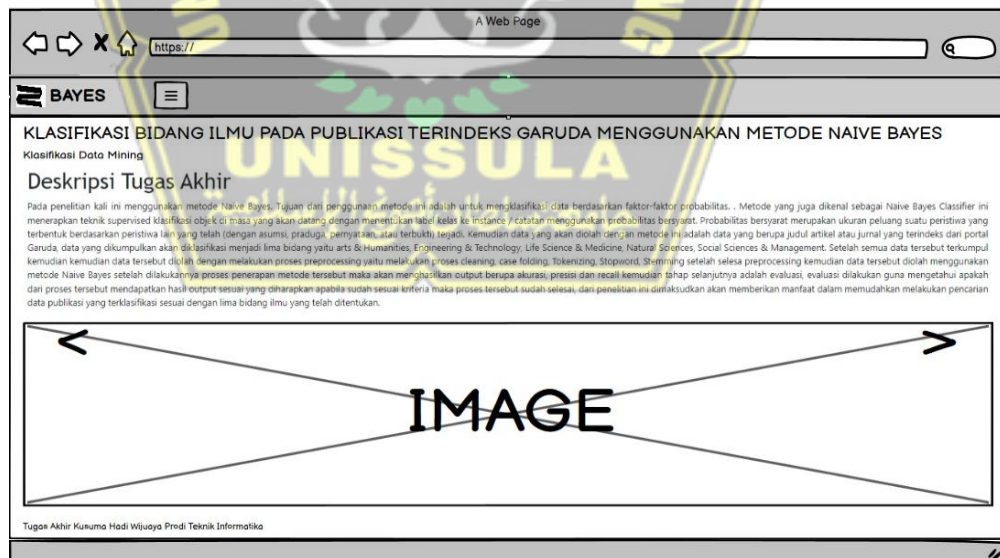
Pada gambar 3.14 merupakan gambar rancangan bentuk menu bar yang akan diimplementasikan pada *user interface* sistem yang akan dibuat dimana pada tampilan tersebut terdapat 3 menu utama yaitu menu *home*, *dataset* dan *predict*, tiga menu tersebut nantinya akan terletak pada samping kiri dari isi halaman

2. Halaman *Home*

halaman *home* adalah halaman yang menampilkan judul tugas akhir beserta deskripsi tugas akhir.



Gambar 3. 15 Halaman *Home*



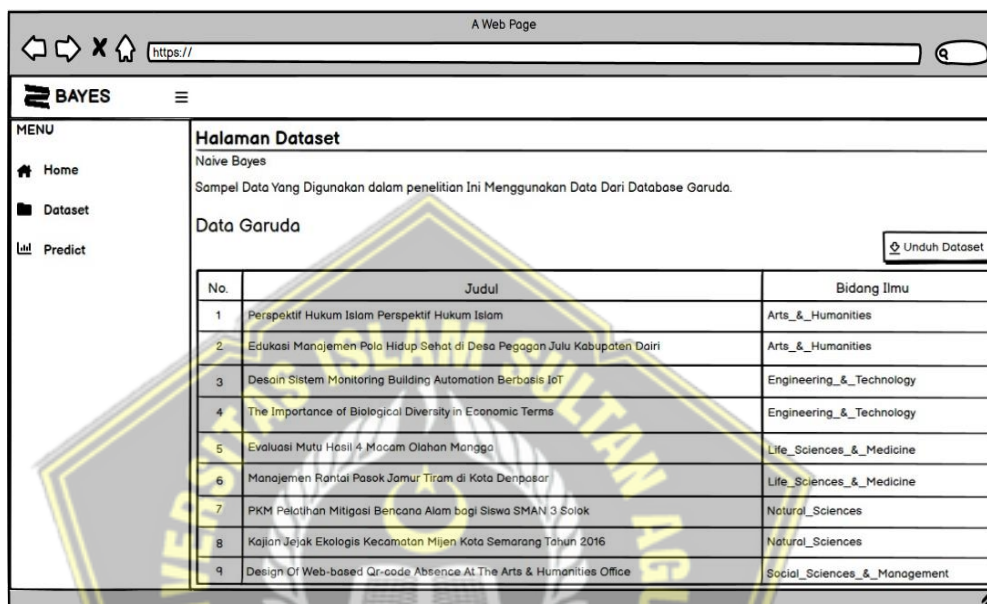
Gambar 3. 16 Halaman *Home full screen*

Pada Gambar 3.15 dan 3.16 Adalah tampilan dari halaman *home* yang mana tampilan tersebut berisi judul tugas akhir dan deskripsi tugas akhir dan beberapa gambar *slide show* yang berkaitan dengan tema dari

lima bidang ilmu.

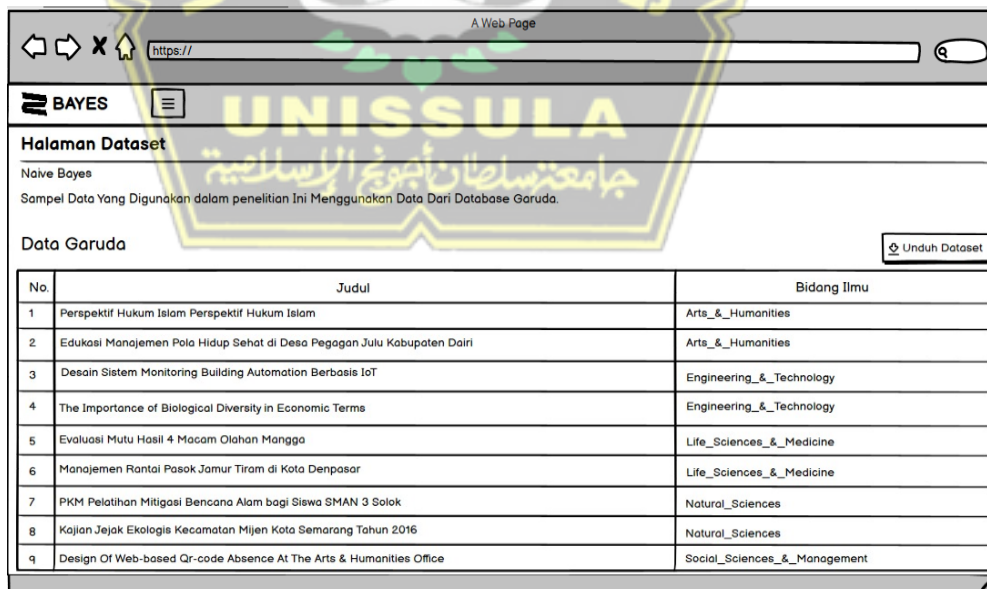
3. Halaman *Dataset*

Halaman *dataset* adalah bagian halaman yang berfungsi untuk menampilkan *dataset*.



No.	Judul	Bidang Ilmu
1	Perspektif Hukum Islam Perspektif Hukum Islam	Arts_&_Humanities
2	Edukasi Manajemen Pola Hidup Sehat di Desa Pegagan Julu Kabupaten Dairi	Arts_&_Humanities
3	Desain Sistem Monitoring Building Automation Berbasis IoT	Engineering_&_Technology
4	The Importance of Biological Diversity in Economic Terms	Engineering_&_Technology
5	Evaluasi Mutu Hasil 4 Macam Olahan Mangga	Life_Sciences_&_Medicine
6	Manajemen Rantai Pasok Jamur Tiram di Kota Denpasar	Life_Sciences_&_Medicine
7	PKM Pelatihan Mitigasi Bencana Alam bagi Siswa SMAN 3 Solok	Natural_Sciences
8	Kajian Jejak Ekologis Kecamatan Mijen Kota Semarang Tahun 2016	Natural_Sciences
9	Design Of Web-based Qr-code Absence At The Arts & Humanities Office	Social_Sciences_&_Management

Gambar 3. 17 Halaman *Dataset*



No.	Judul	Bidang Ilmu
1	Perspektif Hukum Islam Perspektif Hukum Islam	Arts_&_Humanities
2	Edukasi Manajemen Pola Hidup Sehat di Desa Pegagan Julu Kabupaten Dairi	Arts_&_Humanities
3	Desain Sistem Monitoring Building Automation Berbasis IoT	Engineering_&_Technology
4	The Importance of Biological Diversity in Economic Terms	Engineering_&_Technology
5	Evaluasi Mutu Hasil 4 Macam Olahan Mangga	Life_Sciences_&_Medicine
6	Manajemen Rantai Pasok Jamur Tiram di Kota Denpasar	Life_Sciences_&_Medicine
7	PKM Pelatihan Mitigasi Bencana Alam bagi Siswa SMAN 3 Solok	Natural_Sciences
8	Kajian Jejak Ekologis Kecamatan Mijen Kota Semarang Tahun 2016	Natural_Sciences
9	Design Of Web-based Qr-code Absence At The Arts & Humanities Office	Social_Sciences_&_Management

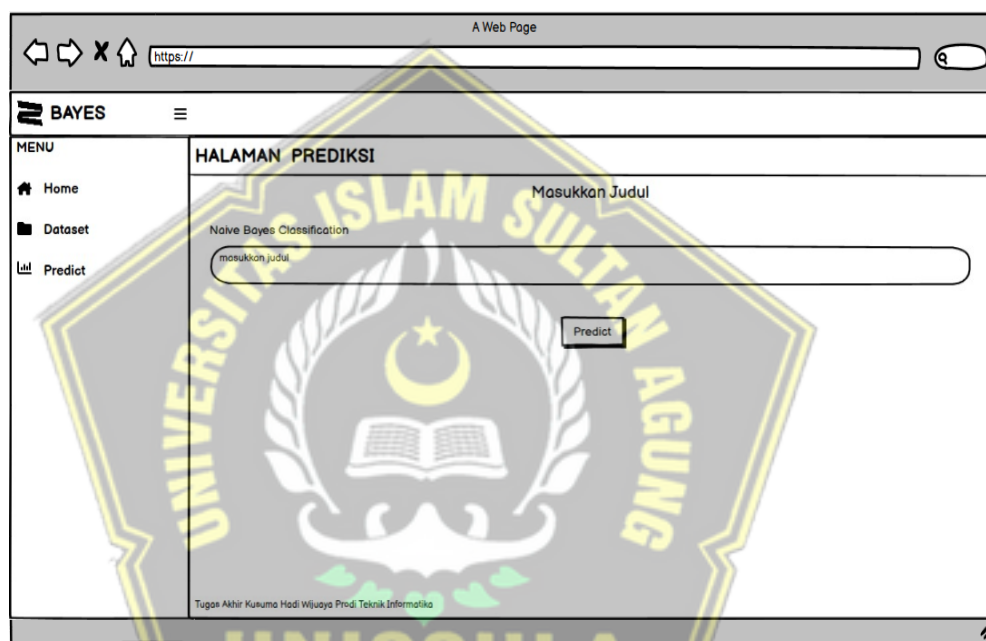
Gambar 3. 18 halaman *Dataset Full Screen*

Pada Gambar 3.17 dan 3.18 Merupakan tampilan halaman *dataset*, dimana halaman tersebut nantinya akan menampilkan data yang berupa

judul artikel dan bidang ilmunya dalam bentuk tabel, kemudian pada halaman tersebut juga akan dilengkapi fitur *download dataset* sehingga nanti user dapat mengunduh data tersebut.

4. Halaman *Predict*

Halaman *predict* merupakan halaman inti dari sistem yang akan dibangun kali ini dimana halaman tersebut nantinya yang akan digunakan dalam melakukan pengujian akurasi dari setiap judul yang akan diuji.

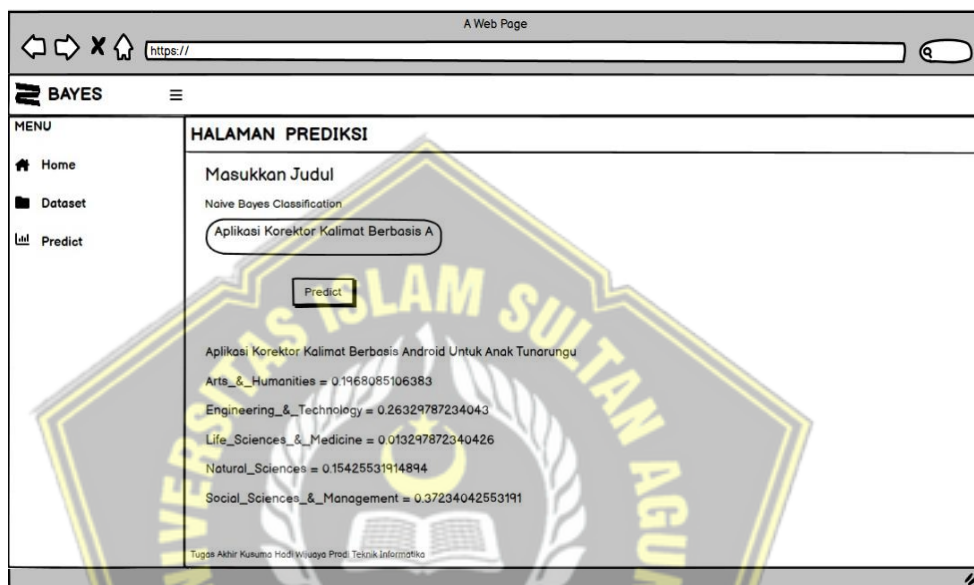


Gambar 3. 19 Halaman *Predict*

Pada Gambar 3.10 adalah rancangan gambar halaman prediksi yang akan diimplementasikan dalam pembuatan *web* klasifikasi judul artikel, pada halaman tersebut berisi kolom untuk menginput judul artikel serta *button* untuk aksi dalam memproses klasifikasi tersebut.

5. *Output* dari halaman *predict*

Rancangan tampilan *output* halaman prediksi merupakan tampilan yang sangat penting dari sistem ini yang mana nantinya dari tampilan *output* tersebut *user* dapat mengetahui hasil dari prediksi akurasi judul artikel.



Gambar 3. 20 Halaman *output predict*

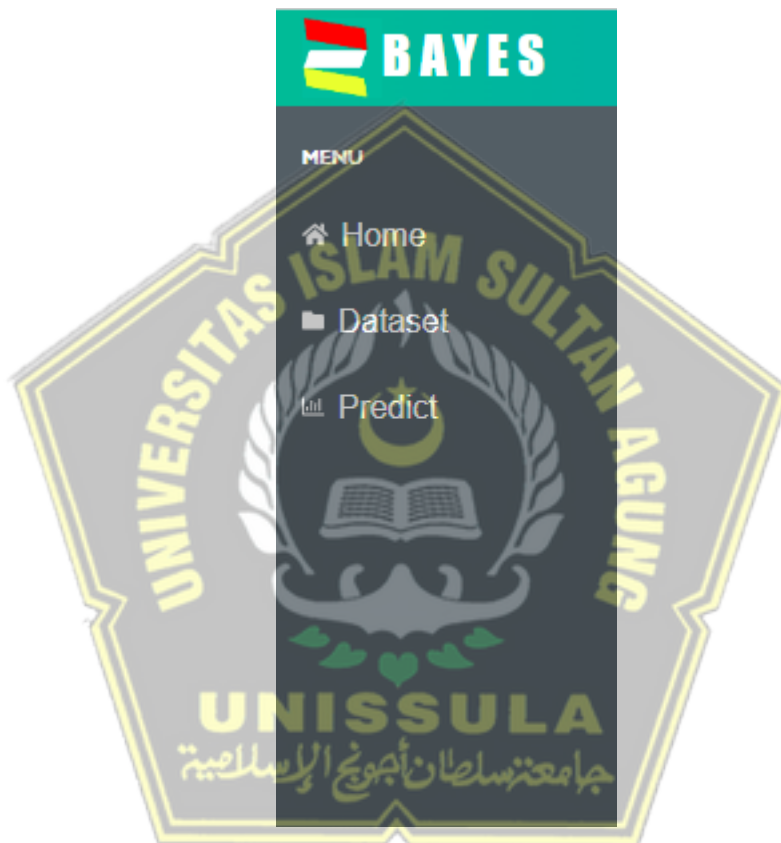
Pada Gambar 3.11 merupakan rancangan gambar *output* yang nantinya akan diaplikasikan dalam pembuatan sistem klasifikasi judul artikel, rancangan tampilan *output* tersebut nantinya akan menampilkan judul artikel yang telah diinput oleh *user* dan diproses oleh sistem beserta bidang ilmunya.

BAB IV HASIL DAN ANALISIS PENELITIAN

4.1 Implementasi *User Interface* (UI)

a. Tampilan *Menu Bar*

Berikut merupakan hasil tampilan *menu bar* pada sistem klasifikasi artikel jurnal yang telah diimplementasikan.



Gambar 4. 1 Tampilan Menu Bar

Pada gambar 4.1 adalah gambar hasil implementasi untuk tampilan menu bar yang telah dibuat pada sistem klasifikasi judul artikel jurnal publikasi garuda, pada *menu bar* tersebut terdiri dari tiga menu yaitu menu *Home*, *dataset* dan *predict* menu tersebut terletak di sebelah kiri dari isi halaman.

b. Halaman *Home*

Berikut merupakan tampilan halaman home yang berisi deskripsi tugas akhir

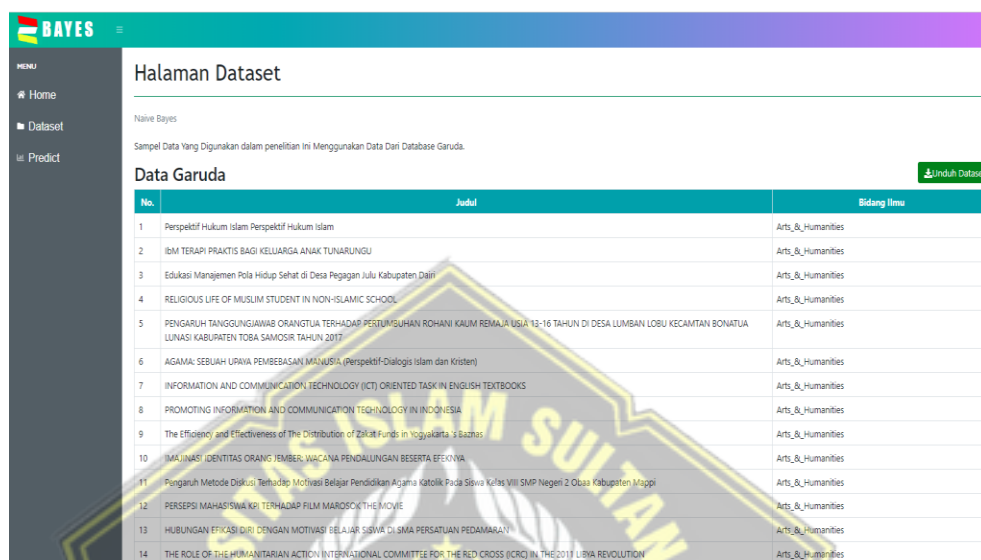


Gambar 4. 2 Tampilan halaman awal sistem klasifikasi

Pada gambar 4.1 merupakan gambar tampilan halaman home dimana tampilan tersebut berisi judul tugas akhir dan deskripsi tugas akhir yang memuat apa saja yang dilakukan dalam penyusunan tugas akhir serta apa saja komponen yang diperlukan dalam membangun sistem pada tugas akhir tersebut, pada halaman home juga menampilkan gambar-gambar yang berkaitan dengan lima bidang ilmu dalam pembuatan tugas akhir tersebut yaitu *arts & Humanities, Engineering & Technology, Life Science & Medicine, Natural Sciences, Social Sciences & Management*.

c. Halaman dataset

Dibawah ini merupakan tampilan halaman dataset



Halaman Dataset

Naive Bayes

Sampel Data Yang Digunakan dalam penelitian Ini Menggunakan Data Dari Database Garuda.

Data Garuda [Unduh Dataset](#)

No.	Judul	Bidang Ilmu
1	Perspektif Hukum Islam Perspektif Hukum Islam	Arts_& Humanities
2	IBM TERAPI PRAKTIS BAGI KELUARGA ANAK TUNARUNGU	Arts_& Humanities
3	Edukasi Manajemen Pola Hidup Sehat di Desa Pegagan Julu Kabupaten Dairi	Arts_& Humanities
4	RELIGIOUS LIFE OF MUSLIM STUDENT IN NON-ISLAMIC SCHOOL	Arts_& Humanities
5	PENGARUH TANGGUNGJAWAB ORANGTUA TERHADAP PERTUMBUHAN ROHANI KAJUM REMAJA USIA 13-16 TAHUN DI DESA LUMBAN LOBU KECAMATAN BONATUA LUNASI KABUPATEN TOBA SAMOSIR TAHUN 2017	Arts_& Humanities
6	AGAMA: SEBUAH UPAYA PEMBEBASAN MANUSIA (Perspektif-Dialogis Islam dan Kristen)	Arts_& Humanities
7	INFORMATION AND COMMUNICATION TECHNOLOGY (ICT) ORIENTED TASK IN ENGLISH TEXTBOOKS	Arts_& Humanities
8	PROMOTING INFORMATION AND COMMUNICATION TECHNOLOGY IN INDONESIA	Arts_& Humanities
9	The Efficiency and Effectiveness of The Distribution of Zakat Funds in Yogyakarta 's Baznas	Arts_& Humanities
10	IMAJINASI IDENTITAS ORANG JEMBER: WACANA PENDALINGAN BESERTA EFEKNYA	Arts_& Humanities
11	Pengaruh Metode Diskusi Terhadap Motivasi Belajar Pendidikan Agama Katolik Pada Siswa Kelas VIII SMP Negeri 2 Obaa Kabupaten Maggi	Arts_& Humanities
12	PERSEPSI MAHASISWA KPI TERHADAP FILM MAROSOK 'THE MOVIE	Arts_& Humanities
13	HUBUNGAN EFikasi DIRI DENGAN MOTIVASI BELAJAR SISWA DI SMA PERSATUAN PEDAMARAN	Arts_& Humanities
14	THE ROLE OF THE HUMANITARIAN ACTION INTERNATIONAL COMMITTEE FOR THE RED CROSS (ICRC) IN THE 2011 LIBYA REVOLUTION	Arts_& Humanities

Gambar 4. 3 Tampilan halaman *dataset*



Halaman Dataset

Naive Bayes

Sampel Data Yang Digunakan dalam penelitian Ini Menggunakan Data Dari Database Garuda.

Data Garuda [Unduh Dataset](#)

No.	Judul	Bidang Ilmu
1	Perspektif Hukum Islam Perspektif Hukum Islam	Arts_& Humanities
2	IBM TERAPI PRAKTIS BAGI KELUARGA ANAK TUNARUNGU	Arts_& Humanities
3	Edukasi Manajemen Pola Hidup Sehat di Desa Pegagan Julu Kabupaten Dairi	Arts_& Humanities
4	RELIGIOUS LIFE OF MUSLIM STUDENT IN NON-ISLAMIC SCHOOL	Arts_& Humanities
5	PENGARUH TANGGUNGJAWAB ORANGTUA TERHADAP PERTUMBUHAN ROHANI KAJUM REMAJA USIA 13-16 TAHUN DI DESA LUMBAN LOBU KECAMATAN BONATUA LUNASI KABUPATEN TOBA SAMOSIR TAHUN 2017	Arts_& Humanities
6	AGAMA: SEBUAH UPAYA PEMBEBASAN MANUSIA (Perspektif-Dialogis Islam dan Kristen)	Arts_& Humanities
7	INFORMATION AND COMMUNICATION TECHNOLOGY (ICT) ORIENTED TASK IN ENGLISH TEXTBOOKS	Arts_& Humanities
8	PROMOTING INFORMATION AND COMMUNICATION TECHNOLOGY IN INDONESIA	Arts_& Humanities
9	The Efficiency and Effectiveness of The Distribution of Zakat Funds in Yogyakarta 's Baznas	Arts_& Humanities
10	IMAJINASI IDENTITAS ORANG JEMBER: WACANA PENDALINGAN BESERTA EFEKNYA	Arts_& Humanities

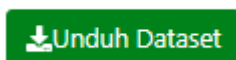
Gambar 4. 4 Tampilan halaman *dataset full screen*

Pada gambar 4.3 dan 4.4 adalah tampilan halaman *dataset* dimana pada halaman tersebut menampilkan data garuda yang telah diberi label sesuai dengan bidang ilmunya, data tersebut berjumlah 500 masing dari bidang ilmu terdapat 100 data judul artikel jurnal publikasi.pada halaman

dataset juga terdapat menu download yang berfungsi untuk mendownload dataset tersebut.

d. *Download button*

Dibawah ini merupakan tombol *download* yang tersedia pada halaman *dataset*



Gambar 4. 5 *Download button*

Pada gambar 4.3 merupakan menu *download button* dimana tombol tersebut berfungsi untuk mendownload *dataset* yang terdapat pada halaman tersebut, dimana *user* cukup hanya mengklik tombol tersebut kemudian dataset secara otomatis akan terdownload ke dalam *device user*

e. *Proses download dataset*

Dibawah ini merupakan tampilan proses ketika *user* mendownload *dataset*

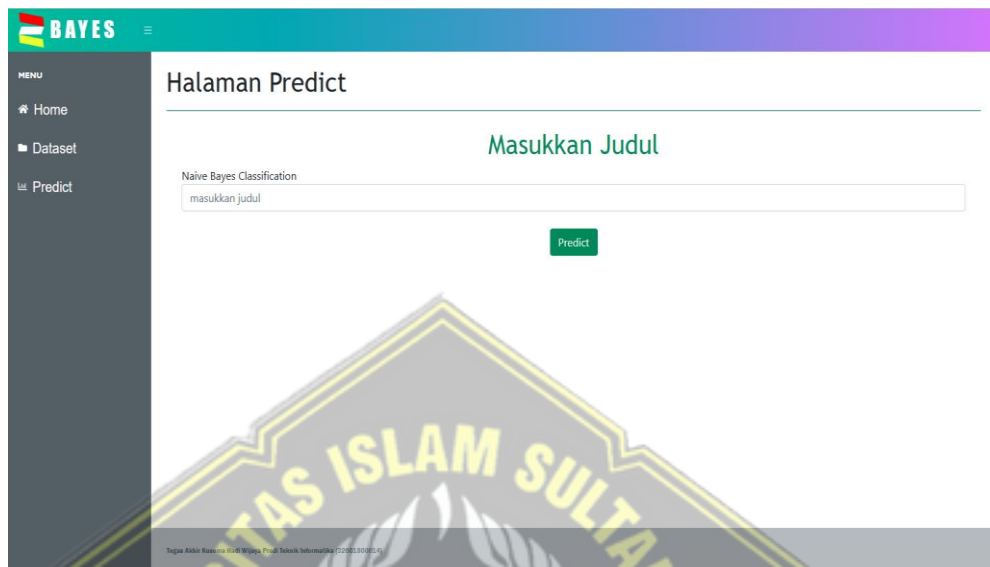
No.	Judul	Bidang Ilmu
1	Perspektif Hukum Islam Perspektif Hukum Islam	Arts_& Humanities
2	IBI TERAPI PRAKTIS BAGI KELLARGA ANAK TUNARUNGU	Arts_& Humanities
3	Edukasi Manajemen Pola Hidup Sehat di Desa Pegagan Julu Kabupaten Dairi	Arts_& Humanities
4	RELIGIOUS LIFE OF MUSLIM STUDENT IN NON-ISLAMIC SCHOOL	Arts_& Humanities
5	PENGARUH TANGGUNGJAWAB ORANGTUA TERHADAP PERTUMBUHAN ROHANI KALAM REMAJA USIA 18-16 TAHUN DI DESA LUMBAN LOBU KECAMATAN BONATUA LUNASI KABUPATEN TORA SAMOSIR TAHUN 2017	Arts_& Humanities
6	AGAMA: SEBUAH UPIRA PEMBEBASAN MANUSIA (Perspektif-Dialogis Islam dan Kristen)	Arts_& Humanities
7	INFORMATION AND COMMUNICATION TECHNOLOGY (ICT) ORIENTED TASK IN ENGLISH TEXTBOOKS	Arts_& Humanities
8	PROMOTING INFORMATION AND COMMUNICATION TECHNOLOGY IN INDONESIA	Arts_& Humanities
9	The Efficiency and Effectiveness of The Distribution of Zakat Funds in Yogyakarta 's Bantras	Arts_& Humanities
10	IMAINASI IDENTITAS ORANG JEMBER: WAGANA PENDOLLINGAN BESERTA EFENYA	Arts_& Humanities
11	Pengaruh Metode Diskusi Terhadap Motivasi Belajar Pendidikan Agama Katolik Pada Siswa Kelas VIII SMP Negeri 2 Obaa Kabupaten Magpi	Arts_& Humanities
12	PERSEPSI MAHASISWA KPI TERHADAP FILM MAROKOK THE MOVIE	Arts_& Humanities
13	HUBUNGAN EFikasi DIRI DENGAN MOTIVASI BELAJAR SISWA DI SMA PERSATUAN PEDAMARAN	Arts_& Humanities
14	THE ROLE OF THE HUMANITARIAN ACTION INTERNATIONAL COMMITTEE FOR THE RED CROSS (ICRC) IN THE 2011 LIBYA REVOLUTION	Arts_& Humanities
	SAHA MIKRO KECIL DAN MENENGAH (UMKM) PADA MASA PANDEMI COVID-19 DI KECAMATAN SAKO PALEMBANG	Arts_& Humanities

Gambar 4. 6 Proses *dataset* yang telah terdownload

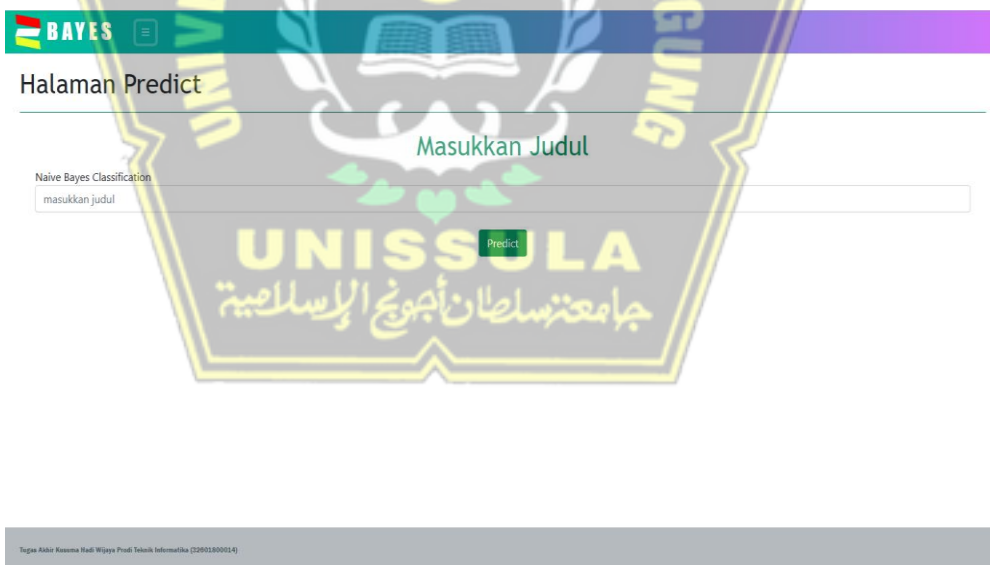
Pada gambar 4.4 merupakan tampilan ketika user mendownload *dataset* tersebut dimana *user* setelah mengklik tombol unduh tersebut data akan terunduh ke dalam *folder download* di perangkat komputer pengguna.

f. Tampilan halaman *predict*

Berikut merupakan tampilan halaman prediksi



Gambar 4. 7 Tampilan halaman *Predict*



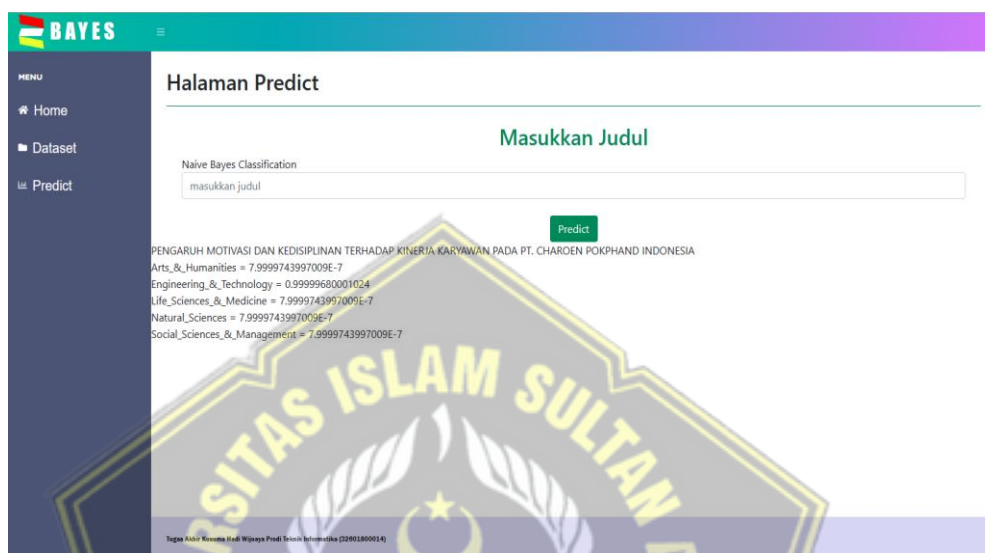
Gambar 4. 8 Tampilan halaman *Predict full screen*

Pada gambar 4.7 dan 4.8 merupakan tampilan pada halaman *predict* dimana halaman tersebut 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

g. *Output* dari hasil prediksi

Berikut adalah gambar *output* halaman prediksi



Gambar 4. 9 Tampilan *output* hasil prediksi

Pada gambar 4.6 merupakan tampilan *output* hasil prediksi dimana tampilan *output* tersebut berupa judul artikel beserta bidang lima ilmunya, dalam *output* tersebut terdapat 5 bidang ilmu dengan masing-masing memiliki nilai akurasi dimana akurasi tertinggi menunjukkan bahwa bidang ilmu tersebut sesuai dengan bidang ilmu dari judul artikel jurnal yang telah diinputkan dan pada gambar diatas user memasukkan judul dari bidang ilmu *Engineering & Technology* dan ketika sistem memproses prediksi menghasilkan bahwa dari lima bidang ilmu yang tampil hanya *Engineering & Technology* yang memiliki nilai prediksi tertinggi. Dengan tampilan tersebut memeungkinkan *user* akan lebih mudah mengetahui kategori bidang ilmu dari judul artikel yang *user* cari.

4.2 Pengujian Sistem

Tahap selanjutnya merupakan pengujian sistem, dimana pengujian yang digunakan pada penelitian ini memakai pengujian *black box testing*, *black box testing* merupakan teknik pengujian sebuah sistem yang bertujuan untuk mengetes apakah sistem berjalan normal sesuai fungsi yang direncanakan, dalam *black box testing* memiliki beberapa macam jenis diantaranya, *function testing*, *non-functional testing*, dan *regression testing*. Untuk pengujian kali ini menggunakan jenis *function testing* yang mana akan dilakukan pengujian di setiap fungsi yang terdapat pada sistem.

Tabel 4. 1 Pengujian *Black Box*

Skenario Pengujian	Kasus Pengujian	Hasil Pengujian	Kesimpulan
Tampil <i>dataset</i>	Halaman menampilkan <i>dataset</i>	Sesuai	Normal
<i>Input</i> judul	Menginput judul artikel jurnal publikasi ke dalam kolom <i>predict</i>	Sesuai	Normal
Test hasil prediksi	Mengecek <i>output</i> hasil prediksi klasifikasi judul artikel jurnal publikasi	Sesuai	Normal
<i>Download dataset</i>	Melakukan <i>download dataset</i> yang terdapat pada halaman <i>dataset</i>	Sesuai	Normal

Pada Tabel 1 merupakan hasil pengujian sistem dimana dalam pengujian tersebut ada empat pengujian fitur dalam sistem yaitu fitur untuk menampilkan dataset, fitur untuk menginput judul, fitur untuk melihat hasil prediksi, dan yang terakhir berupa fitur untuk mendownload dataset dan dari pengujian black box

tersebut fitur berjalan sesuai dengan yang diharapkan.

Hasil yang didapat dari pengujian sistem yang dilakukan menunjukkan bahwa sistem klasifikasi judul artikel publikasi terindeks berbasis *web* ini, sudah dapat menjalankan setiap fungsi yang dimiliki dengan baik. Setidaknya ada 4 fungsi utama pada sistem yang berhasil berjalan dengan normal dan semestinya

4.3 Analisa Hasil Validasi Penerapan Algoritma *Naive Bayes*

Tahap ini merupakan bagian dari tahapan untuk mengetahui sejauh mana tingkat akurasi sistem yang telah dibuat

Tabel 4. 2 Jumlah Data

Nama Data	Jumlah (%)	Jumlah angka
Data <i>Training</i>	80%	400
Data <i>Testing</i>	20%	100

Total Data : 500

Tabel 4. 3 Rincian Data

<i>class</i>	<i>Arts & Humanities</i>	<i>Engineering & Technology</i>	<i>Life Science & Medicine</i>	<i>Natural Sciences</i>	<i>Social Sciences & Management</i>
Data <i>Training</i>	80	80	80	80	80
Data <i>Testing</i>	20	20	20	20	20

Dari data tersebut nantinya akan menghasilkan akurasi berupa *confusion matrix*, dari *confusion matrix* ini akan dilakukan pencarian nilai akurasi, presisi, *recall* dan *F 1 score* . Dari hasil *training* yang telah di uji coba menghasilkan *confusion matrix* dengan nilai *accuracy, precision, recall* dan *F-1 Score* sebagai berikut:

Tabel 4. 4 *Confusion Matrix*

	<i>Actually Positive(1)</i>	<i>Actually Negatif(1)</i>
<i>Predicted Positive(1)</i>	20 (TP)	80 (FP)
<i>Actually Negatif(1)</i>	80 (FN)	80 (TN)

```
"true positives": 20,
"true negatives": 80,
"false positives": 80,
"false negatives": 80,
```

Gambar 4. 10 *Output hasil Confusion Matrix*

Gambar 4.10 merupakan *output confusion matrix* dari hasil percobaan yang dilakukan mendapatkan hasil *true poisitives* 20, *true negatives* 80, *false positive* 80, dan *false negatives* 80.

Dengan hasil yang didapatkan dari *confusion matrix* maka akan didapatkan hasil *accuracy, precision, recall, f1 score*

Tabel 4. 5 Hasil Pengukuran *Accuracy, Recall, Precision, F1 score*

Accuracy	Precision	Recall	F1 score
0.44	0.04	0.2	0.06

```
Training model.... Model disimpan.... OK {
  "breakdown": {
    "overall": {
      "accuracy": 0.44000000000000006,
      "balanced accuracy": 0.5,
      "f1 score": 0.06666666666666668,
      "precision": 0.04,
      "recall": 0.2,
```

Gambar 4. 11 Pengukuran *Accuracy, Recall, Precision, F1 score*

Gambar 4.11 menunjukkan bahwa hasil pengujian dengan sampel data 500 judul artikel dengan *Naive Bayes* mendapatkan hasil nilai akurasi sebesar 0.44 dengan presisi, *recall*, *f-1 score* masing-masing mendapat nilai 0.06, 0.04, dan 0.2, dengan demikian menunjukkan bahwa metode *Naive Bayes* masih kurang maksimal untuk melakukan klasifikasi judul artikel jurnal publikasi garuda dimana metode tersebut hanya menghasilkan akurasi 0.44 dimana hasil tersebut masih tergolong kategori rendah. Terdapat beberapa faktor yang juga dapat menjadi penyebab nilai akurasi yang rendah pertama adalah data, dimana data yang peneliti gunakan terbilang data yang masih sangat kotor dan perlu banyak pembersihan yang mana pembersihan yang lebih dalam akan mengubah 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 beberapa tema judul yang tidak sesuai bidang ilmunya.

4.4 Pembahasan hasil akurasi

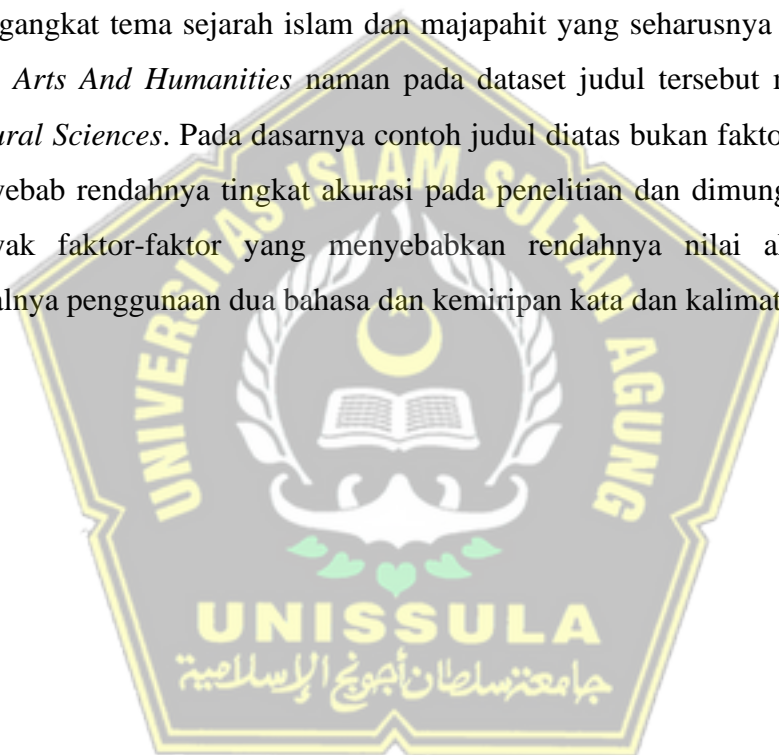
Pada hasil akurasi menunjukkan hasil yang masih belum sesuai yang diharapkan ,hasil tersebut masih tergolong sangat rendah dimana hasil akurasinya hanya sebesar 0.4400 dengan presisi, *recall*, *f-1 score* masing-masing mendapat nilai 0.066, 0.04, dan 0.2, hasil tersebut tergolong masih rendah, terdapat banyak faktor yang menjadi pemicu rendahnya hasil akurasi pertama kondisi data dimana data tersebut belum akurat dan masih terdapat beberapa kesamaan kata atau kalimat dari satu bidang ilmu ke bidang ilmu lainnya, kemudian terdapat beberapa tema judul yang tidak sesuai dengan bidang ilmunya sehingga menyebabkan tingkat akurasi rendah.

Tabel 4. 6 Judul yang memiliki tema yang tidak sesuai bidangnya.

No.	Judul Artikel	Bidang ilmu yang terlabel	Bidang ilmu yang seharusnya
1.	Perancangan Kampanye Etika Komunikasi pada Media Sosial	<i>Engineering & Technology</i>	<i>Arts And Humanities</i>
2.	Deteksi Dini Kesehatan Dan Edukasi Ergonomi Pada Karyawan <i>Elling Custom Made Bra</i>	<i>Engineering & Technology</i>	<i>Life Sciences & Medicine</i>
3.	<i>The Role of Christian Religious Arts & Humanities as a Strategy in Dating Radicalism of Arts & Humanities in Indonesia</i>	<i>Engineering & Technology</i>	<i>Arts And Humanities</i>
4.	Pengaruh Islam Sebagai Salah Satu Penyebab Mundurnya Kerajaan Majapahit	<i>Natural Sciences</i>	<i>Arts And Humanities</i>

Pada Tabel 2.7 merupakan tabel yang berisi contoh judul yang mempengaruhi nilai akurasi, pada tabel tersebut berisi judul-judul yang memiliki tema tidak sesuai bidangnya seperti tabel satu dimana pada judul tersebut bernama “Perancangan Kampanye Etika Komunikasi pada Media Sosial” yang mana judul tersebut telah memiliki label bidang ilmu *Engineering & Technology* tetapi pada judul tersebut mengangkat tema *Arts And Humanities* dan seharusnya juga masuk bidang tema *Arts And Humanities*, kemudian pada judul 2 yang berjudul “Deteksi Dini Kesehatan Dan Edukasi Ergonomi Pada Karyawan *Elling Custom Made Bra*” dimana judul tersebut

mengangkat tema *Life Sciences & Medicine* tetapi pada *dataset* telah terabeli *Engineering & Technology* kemudian pada judul 3 yang berjudul “*The Role of Christian Religious Arts & Humanities as a Strategy in Dating Radicalism of Arts & Humanities in Indonesia*” judul tersebut mengangkat tema keagamaan tetapi telah memiliki label pada bidang ilmu *Engineering & Technology* dimana seharusnya judul tersebut masuk pada bidang ilmu *Arts And Humanities*, kemudian untuk judul 4 yang berjudul “Pengaruh Islam Sebagai Salah Satu Penyebab Mundurnya Kerajaan Majapahit” pada judul tersebut mengangkat tema sejarah islam dan majapahit yang seharusnya masuk bidang *ilmu Arts And Humanities* naman pada *dataset* judul tersebut memiliki label *Natural Sciences*. Pada dasarnya contoh judul diatas bukan faktor satu-satunya penyebab rendahnya tingkat akurasi pada penelitian dan dimungkinkan masih banyak faktor-faktor yang menyebabkan rendahnya nilai akurasi seperti misalnya penggunaan dua bahasa dan kemiripan kata dan kalimat.



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Hasil yang didapatkan dalam penelitian ini menunjukkan bahwa Klasifikasi Bidang Ilmu Pada Publikasi Terindeks Garuda Menggunakan Metode Naive Bayes belum cukup maksimal dalam mendapatkan tingkat akurasi yang diharapkan, dari 500 dataset yang telah dilakukan proses training yaitu dengan data sampel 400 dan data testing 100 hanya mendapat hasil akurasi sebesar 0.44 sedangkan presisi, recall, dan F1 Score masing-masing mendapat nilai 0.04, 0.2, dan 0.06 hasil tersebut tergolong masih rendah, banyak faktor yang menjadi pemicu rendahnya tingkat akurasi dari penelitian ini pertama yaitu data yang digunakan masih kurang tepat dikarenakan memang dari pihak pengembang data-data tersebut tidak bisa diubah oleh pihak luar termasuk penulis dari penelitian ini, yang kedua ialah masih terdapat banyak data yang memiliki tingkat kemiripan baik perkataan maupun kalimat di beberapa bidang ilmu yang berbeda kemudian banyak tema judul artikel jurnal yang tidak sesuai dengan bidang ilmunya, jadi hal ini menjadi kesulitan tersendiri dalam mendapatkan nilai akurasi yang tinggi, dilihat dari nilai hasil akurasi yang hanya sebesar 0.44 menjadikan metode naive bayes kurang akurat dalam melakukan klasifikasi ini

5.2 Saran

Saran untuk perangkat Sistem:

1. Untuk penelitian selanjutnya, perlu adanya tambahan algoritma lain untuk menunjang hasil keakuratan suatu hasil yang lebih baik.
2. Dikarenakan nilai akurasi yang dihasilkan masih tergolong rendah, untuk pengembangan berikutnya bisa dilakukan dengan metode yang berbeda selain *Naive Bayes*
3. Menggunakan data sampel yang lebih akurat.
4. Desain rancangan tampilan sistem bisah lebih dikembangkan lagi
5. Untuk perangkat sistem fitur-fitur di dalamnya bisa dikembangkan lagi

DAFTAR PUSTAKA

- Ailmi, N., Saharuna, Z., & Tungadi, E. (2020). Metode Klasifikasi Pada Aplikasi Pendukung Keputusan Untuk Pemilihan Unit Kegiatan Mahasiswa. *Prosiding Seminar Nasional Teknik Elektro dan Informatika (SNTEI) 2020*, 142–147.
- Budiman, A., Young, J. C., & Suryadibrata, A. (2021). Implementasi algoritma naïve bayes untuk klasifikasi konten twitter dengan indikasi depresi. *Jurnal Informatika: Jurnal Pengembangan IT*, 6(1), 133–138. <http://jurnal.kopertipindonesia.or.id/index.php/kopertip/article/view/130>
- 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>
- Habib, S. M., Haerani, E., Gusti, S. K., & Ramadhani, S. (2022). Klasifikasi Berita Menggunakan Metode Naïve Bayes Classifier. *Jurnal Nasional Komputasi dan Teknologi Informasi*, 5(3), 248–258. <https://doi.org/10.33005/scan.v16i3.2870>
- Handoko, M. R., & Neneng. (2021). Sistem Pakar Diagnosa Penyakit Ispa Menggunakan Metode Naive Bayes Classifier Berbasis Web. *Jurnal Teknologi dan Sistem Informasi*, 2(3), 50–58. <https://doi.org/10.22303/csrid.10.3.2018.127-138>
- Mustofa, H., & Mahfudh, A. A. (2019). Klasifikasi berita hoax dengan menggunakan metode naive bayes. *Walisongo Journal of Information Technology*, 1(1), 1–12. <https://doi.org/10.21580/wjit.2019.1.1.3915>
- Nurdin, N., Suhendri, M., Afrilia, Y., & Rizal, R. (2021). Klasifikasi Karya Ilmiah (Tugas Akhir) Mahasiswa Menggunakan Metode Naive Bayes Classifier (NBC). *Sistemasi*, 10(2), 268. <https://doi.org/10.32520/stmsi.v10i2.1193>
- Putri, H., Purnamasari, A. I., Dikananda, A. R., Nurdiawan, O., & Anwar, S. (2021). Penerima Manfaat Bantuan Non Tunai Kartu Keluarga

- Sejahtera Menggunakan Metode NAÏVE BAYES dan KNN. *Building of Informatics, Technology and Science (BITS)*, 3(3), 331–337. <https://doi.org/10.47065/bits.v3i3.1093>
- Putro, H. F., Vuldari, R. T., & Saptomo, W. L. (2020). Penerapan metode naive bayes untuk klasifikasi pelanggan. *Jurnal Teknologi Informasi dan Komunikasi (TIKomSiN)*, 8(2). <https://doi.org/10.30646/tikomsin.v8i2.500>
- Rasila, I., Ristian, U., Rekayasa Sistem Komputer, J., & MIPA Universitas Tanjungpura Jl Hadari Nawawi, F. H. (2019). Implementasi Metode Naive Bayes Classifier Pada Sistem Pengklasifikasi Berita Otomatis Berbasis Website (Studi Kasus: Berita Lokal Dari Mediamassa Online Kalimantan Barat). *Coding : Jurnal Komputer dan Aplikasi*, 07(2), 49–60.
- Seran, D. R. M., Setyawan, Y., & Bekti, R. D. (2021). ... Metode K-Nearest Neighbour Dan Naive Bayes Classification Untuk Membentuk Sistem Rekomendasi Berita Olahraga Online *Prosiding Snast, 2013*, 128–137. <https://journal.akprind.ac.id/index.php/prosidingsnast/article/view/3347>
%0Ahttps://journal.akprind.ac.id/index.php/prosidingsnast/article/download/3347/2416
- Setiawan, A., Santoso, L. W., & Adipranata, R. (2020). Klasifikasi Artikel Berita Bahasa Indonesia Dengan Naive Bayes Classifier. *Jurnal Infra*, 8(1), 146–151.
- Simanjuntak, K. F. L., Barus, A. C. B., & Anita. (2021). Implementasi Metode Decision Tree dan Algoritma C4.5 Untuk Klasifikasi Kepribadian Masyarakat. *JOISIE Journal Of Information System And Informatics Engineering*, 5(1), 51–59.
- Subroto, I. M. I., Sutikno, T., & Stiawan, D. (2014). The architecture of indonesian publication index: A major indonesian academic database. *Telkomnika (Telecommunication Computing Electronics and Control)*, 12(1), 1–5. <https://doi.org/10.12928/TELKOMNIKA.v12i1.1790>

- Sukarna, R., Ansori, Y., & Syaechurodji. (2021). Implementasi Data Mining Menggunakan Metode Naive Bayes Dengan Feature Selection Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu. *Jurnal Ilmiah Sains dan Teknologi*, 5(1), 9–22. <https://doi.org/10.47080/saintek.v6i1.1467>
- Sulaeman, Suarna, N., Ajiz, A., Bahtiar, A., & Fathurrohman. (2022). Perbandingan kinerja algoritma naïve bayes dan C.45 dalam klasifikasi spam email. *KOPERTIP : Jurnal Ilmiah Manajemen Informatika dan Komputer*, 6(1), 8–14. <https://doi.org/10.32485/kopertip.v6i1.130>
- Sutoyo, I. (2018). Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik. *Jurnal Pilar Nusa Mandiri*, 14(2), 217. <https://doi.org/10.33480/pilar.v14i2.926>
- Wahyu Hadikristanto, A. (2020). Penerapan Algoritma Naive Bayes Untuk Menentukan Klasifikasi Produk Terlaris Pada Penjualan Pulsa. *Paper Knowledge . Toward a Media History of Documents*, 3(Juni), 49–58.
- Wibawa, A. P., Purnama, M. G. A., Akbar, M. F., & Dwiyanto, F. A. (2018). Metode-metode Klasifikasi. *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi*, 3(1), 134–138.

