

**KLASIFIKASI BIDANG ILMU PADA PUBLIKASI TERINDEKS
WEB OF SCIENCE MENGGUNAKAN METODE NAÏVE BAYES**

LAPORAN TUGAS AKHIR

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



DISUSUN OLEH

ABDULLOH MUNIR

NIM 32601800001

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM SULTAN AGUNG SEMARANG
DESEMBER 2022**

FINAL PROJECT

**CLASSIFICATION OF WEB OF SCIENCE INDEXED
PUBLICATIONS USING THE NAÏVE BAYES METHOD**

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



Arranged By:

**ABDULLOH MUNIR
NIM 32601800001**

**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 Web Of Science Menggunakan Metode Naïve Bayes**” ini disusun oleh:

Nama : Abdulloh Munir

NIM : 32601800001

Program Studi : Teknik Informatika

Telah disahkan oleh dosen pembimbing pada :

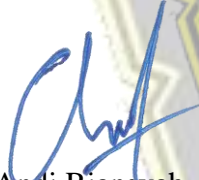
Hari : Senin


Tanggal : 13 Januari 2023

Mengesahkan,

Pembimbing I

Pembimbing II


Andi Riansyah, ST.M.Kom
NIDN. 0609108802


Sam Farisa C.H, 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 Web Of Science Menggunakan Metode Naïve Bayes**” ini telah dipertahankan di depan dosen penguji Tugas Akhir pada :

Hari : Senin

Tanggal : 09 Januari 2023



Penguji I

Penguji II

Ir. Sri Mulyono, M.Eng
NIDN. 0626066601

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

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Abdulloh Munir

NIM : 32601800001

Judul Tugas Akhir : Klasifikasi Bidang Ilmu pada Publikasi Terindeks Web Of Science 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, Desember 2022

Yang menyatakan,


Abdulloh Munir



SURAT PERSETUJUAN UNGGAH KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Abdulloh Munir
NIM : 32601800001
Program Studi : Teknik Informatika
Fakultas : Teknologi Industri

Dengan ini menyerahkan karya ilmiah berupa Tugas Akhir dengan judul :

Klasifikasi Bidang Ilmu Pada Publikasi Terindeks Web Of Science Menggunakan Metode Naïve Bayes dan menyetujuinya menjadi hak milik Universitas Islam Sultan Agung serta memberikan Hak Bebas Royalti Non-eksklusif untuk disimpan, dialihmediakan, dikelola dalam pangkalan data, dan dipublikasikannya di internet atau media lain untuk kepentingan akademis selama tetap mencantumkan nama penulis sebagai pemilik Hak Cipta.

Pernyataan ini saya buat dengan sungguh-sungguh. Apabila dikemudian hari terbukti ada pelanggaran Hak Cipta/Plagiarisme dalam karya ilmiah ini, maka segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa melibatkan pihak Universitas Islam Sultan Agung.

Semarang, Desember 2022

Yang menyatakan,



Abdulloh Munir

KATA PENGANTAR

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

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

1. Rektor UNISSULA Bapak Prof. Dr. H. Gunarto, SH., M.Hum yang mengizinkan penulis menimba ilmu di kampus ini.
2. Dekan Fakultas Teknologi Industri Ibu Dr. Novi Marlyana, ST., MT.
3. Dosen pembimbing I penulis Imam 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 telah memberikan banyak nasehat dan saran.
5. Orang tua penulis yang telah mengizinkan untuk menyelesaikan laporan ini.
6. Dan kepada semua pihak yang tidak dapat saya satu persatu.

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

Semarang, Desember 2022



Abdulloh Munir

DAFTAR ISI

LAPORAN TUGAS AKHIR.....	i
<i>FINAL PROJECT</i>	ii
LEMBAR PENGESAHAN PEMBIMBING.....	iv
LEMBAR PENGESAHAN PENGUJI.....	ivv
SURAT PERNYATAAN KEASLIAN TUGAS AKHIR.....	v
SURAT PERSETUJUAN PUBLIKASI KARYA ILMIAH.....	viv
KATA PENGANTAR.....	viiv
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL.....	xii
ABSTRAK.....	xii
BAB I.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah.....	2
1.3 Pembatasan Masalah.....	2
1.4 Tujuan.....	2
1.5 Manfaat.....	3
1.6 Sistematika Penulisan.....	3
BAB II.....	4
TINJAUAN PUSTAKA DAN DASAR TEORI.....	4
2.1 Tinjauan Pustaka.....	4
2.2 Dasar Teori.....	8
2.2.1 Klasifikasi.....	8
2.2.2 Publikasi.....	8
2.2.3 Data Mining.....	9
2.2.4 Preprocessing.....	9
2.2.5 PHP.....	11
2.2.6 SINTA.....	11
2.2.7 <i>Web Of Science</i>	11

2.2.8	<i>Naïve Bayes</i>	11
2.2.9	5 Bidang Ilmu	12
2.2.10	Evaluasi	13
BAB III	15
METODE PENELITIAN	15
3.1	Metode Penelitian	15
3.2	Preprocessing	16
3.3	Penerapan Metode <i>Naïve Bayes</i>	20
3.4	Evaluasi	26
3.5	Metode Perancangan Alur Sistem	26
3.5.1	Analisa Kebutuhan	26
3.5.2	Implementasi Sistem	27
3.5.3	Analisis Alur Sistem	28
3.6	Perancangan Antarmuka	30
BAB IV	33
HASIL DAN ANALISIS PENELITIAN	33
4.1	User Interface dan Penggunaan Sistem	33
4.2	Analisa dan Pengujian	36
4.3	Analisis Akurasi	37
4.4	Pembahasan Hal Yang Mempengaruhi Hasil Akurasi	40
BAB V	44
KESIMPULAN DAN SARAN	44
5.1	Kesimpulan	44
5.2	Saran	44
DAFTAR PUSTAKA	45
LAMPIRAN	49

DAFTAR GAMBAR

Gambar 3. 1 Flowchart Metode Penelitian	15
Gambar 3. 2 Penerapan kode program preprocessing.....	19
Gambar 3. 3 Persamaan naive bayes.....	23
Gambar 3. 4 Penerapan metode naive bayes.....	25
Gambar 3. 5 Evaluasi (confusion matrix, akurasi, presisi dan recall).....	26
Gambar 3. 6 Alur perancangan sistem	29
Gambar 3. 7 Flowchart Sistem.....	30
Gambar 3. 8 Halaman utama.....	31
Gambar 3. 9 Halaman utama.....	31
Gambar 3. 10 Halaman menu dataset	32
Gambar 3. 11 Halaman predict	32
Gambar 4. 1 Halaman Home.....	33
Gambar 4. 2 Tampilan halaman dataset.....	34
Gambar 4. 3 Tampilan fungsi unduh dataset	35
Gambar 4. 4 Tampilan halaman predict sebelum input judul	35
Gambar 4. 5 Hasil Tampilan output hasil predict.....	36
Gambar 4. 6 Output hasil klasifikasi 5000 data	38
Gambar 4. 7 Output hasil (TP, TN, FP, FN).....	39

DAFTAR TABEL

Tabel 2. 1 Confusion Matrix	14
Tabel 3. 1 Proses data cleaning	17
Tabel 3. 2 Proses case folding.....	17
Tabel 3. 3 Proses tokenizing	18
Tabel 3. 4 Proses stop word	18
Tabel 3. 5 Proses stemming	19
Tabel 3. 6 Rincian data testing dan data training	20
Tabel 3. 7 Pemecahan data training dan data testing menjadi perkata(token)	21
Tabel 3. 8 Pengkategorian perkata(token) ke bidang ilmu yang sesuai.....	22
Tabel 3. 9 Perhitungan dari bidang ilmu 1 (Art and Humanities).....	23
Tabel 3. 10 Perhitungan dari bidang ilmu 2 (Engineering and Technology).....	23
Tabel 3. 11 Perhitungan dari bidang ilmu 3 (Life Science and Medicine)	24
Tabel 3. 12 Perhitungan dari bidang ilmu 4 (Natural Science).....	24
Tabel 3. 13 Perhitungan dari bidang ilmu 5 (Social Science and Management) ..	24
Tabel 4. 1 Pengujian sistem	36
Tabel 4. 2 Jumlah Data.....	37
Tabel 4. 3 Rincian Data.....	38
Tabel 4. 4 Hasil accuracy, precision, recall dan F1 score	38
Tabel 4. 5 Hasil confusion matrix	39
Tabel 4. 6 Judul dengan topik "Covid-19".....	40
Tabel 4. 7 Judul dengan topik "Diagnosis".....	42
Tabel 4. 8 Judul dengan topik "Business".....	43

ABSTRAK

Seiring berjalannya waktu, Keberadaan internet sebagai alat yang berdampak besar bagi perkembangan ilmu pengetahuan tidak dapat dipungkiri. Salah satu contoh perihal yang diutamakan yaitu penerbitan jurnal ilmiah. Kecepatan dalam mengakses di dunia maya menjadikan jurnal ilmiah berkembang sangat cepat, pesat dan banyak. Adanya permasalahan yang ditimbulkan dari maraknya jurnal ilmiah seperti menemukan jurnal ilmiah sesuai dengan bidang ilmu. Dari permasalahan tersebut perlu adanya sebuah metode pengklasifikasian yang bisa diterapkan dalam pencarian jurnal ilmiah tentunya sesuai akan isi topik yang dituju, salah satunya adalah metode data mining. Data Mining adalah suatu proses menemukan pola dalam sejumlah data besar dengan tujuan untuk melakukan klasifikasi, estimasi, prediksi, asosiasi dan klaster. Salah satu metode klasifikasi dalam data mining adalah metode naïve bayes. *Naïve Bayes* merupakan sebuah metode atas dasar probabilitas dan statistika dengan menerapkan teorema bayes, memprediksi peluang hasil yang akan didapatkan dengan memperhitungkan berdasarkan data pengalaman di masa sebelumnya. Tujuan pada penelitian ini yaitu dengan menggunakan metode naïve bayes dalam pengklasifikasian jurnal ke dalam 5 bidang ilmu dapat merumuskan masalah berupa memudahkan dalam pengelompokan jurnal ilmiah sesuai akan topik. Hasil akhir pada penelitian ini yaitu Naïve bayes menghasilkan output kesesuaian judul dari artikel dalam SINTA yang sudah terindeks Web of Science dalam 5 bidang ilmu tersebut. Dengan hasil dari pengujian sistem nilai akurasi, recall, presisi sebesar : akurasi 44,2% recall 24% dan presisi 20,2%.

Kata kunci: Klasifikasi, 5 Bidang Ilmu, *Web of Science*, *Naïve Bayes*, *SINTA*.

ABSTRACT

Along with time, the existence of the internet as a tool that has a big impact on the development of science cannot be denied. One example of priority is the publication of scientific journals. Speed in accessing the virtual world makes scientific journals grow very quickly, rapidly and in large numbers. There are problems arising from the rise of scientific journals such as finding scientific journals in accordance with the field of science. From these problems it is necessary to have a classification method that can be applied in the search for scientific journals, of course according to the contents of the intended topic, one of which is the data mining method. Data Mining is a process of finding patterns in a large amount of data with the aim of classifying, estimating, predicting, associating and clustering. One of the classification methods in data mining is naïve Bayes method. Naïve Bayes is a method based on probability and statistics by applying Bayes' theorem, predicting the probability of the results to be obtained by calculating based on past experience data. The purpose of this study is to use the Naïve Bayes method in classifying journals into 5 fields of science to formulate a problem in the form of facilitating the grouping of scientific journals according to the topic. The final result of this study is that Naïve Bayes produces output that matches the title of articles in SINTA that have been indexed by the Web of Science in the 5 fields of science. With the results of system testing the accuracy, recall, precision values are: 44.2% accuracy, 24% recall and 20.2% precision.

Key words: *Classification, 5 Fields of Science, Web of Science, Naïve Bayes, SINTA.*

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di masa era modern digital saat ini, tidak akan asing lagi dengan adanya tentang publikasi, yaitu suatu proses mengenalkan, memberitahukan hasil suatu karya kepada publik lewat berbagai media offline maupun online (digital), contoh sederhana seperti hasil karya jurnal, karya ilmiah, penelitian ataupun sebagainya.

Publikasi merupakan suatu tahapan paling akhir dalam sebuah penulisan karya ilmiah.(Studi & Kimia, 2018). Dengan adanya publikasi seorang penulis dalam menyimpan dan mempublikasikan karya yang telah dikerjakan. Hal ini merupakan titik tolak untuk mengetahui sejauh mana perkembangan ilmu-ilmu dalam segala bidang di seluruh dunia. Dalam dunia digital banyak sekali para penulis yang mempublikasikan hasil karyanya dari berbagai banyaknya macam dan jenis yang tersebar di media internet. Dengan adanya klasifikasi dapat mempermudah dalam mencari suatu informasi dan sejenisnya, karena sudah dalam bentuk pengelompokan sesuai dengan jenis-jenis nya tersendiri.

Pengklafisikasian merupakan suatu proses pengelompokan atau pemilahan menurut jenis kesamaan atau perbedaan. Contoh metode yang dapat diterapkan yaitu menggunakan metode *Naïve Bayes*. Klasifikasi adalah suatu proses pengelompokan atau mengumpulkan benda/entitas yang sama serta memisahkan benda/entitas yang tidak sama, berfokus untuk mendeskripsikan jenis kelas dari data yang belum mempunyai kategori jenis kelas data tersebut, dalam proses pengklasifikasian hal utama yang dilewati adalah proses tahap training dan tahap testing, dalam kedua tahapan tersebut akan dipergunakan dataset yang sudah diketahui jenis kategori kelas objeknya. (Bimo dkk., 2020)

Banyak sekali layanan terindeks yang terdapat di media internet yang sudah terkenal dan terpercaya hingga saat ini dan banyak digunakan para penulis di dunia. Beberapa layanan pengindeksan yang terdapat di internet dan sudah terkenal di dunia dan banyak digunakan para penulis di seluruh dunia salah satunya adalah *Web of Science*. *Web of Science* adalah karya ilmiah terkemuka di dunia, pencarian

kutipan dan platform informasi analisis, ini digunakan sebagai alat penelitian mendukung beragam tugas ilmiah di berbagai domain pengetahuan serta himpunan data untuk studi intensif data skala besar. (Li dkk., 2018)

Banyak sekali metode data mining yang digunakan untuk mengambil atau memutuskan suatu masalah, salah satunya adalah metode *Naïve Bayes*. *Naïve Bayes* adalah metode klasifikasi probabilistik sederhana. Metode ini mencari tingkat nilai kemungkinan atau probabilitas dengan cara penjumlahan pasangan nilai sinyal dari kumpulan data yang didapatkan dan diolah. (Devita dkk., 2018)

Dari uraian diatas maka penelitian ini mengusulkan sebuah judul “Klasifikasi bidang ilmu pada publikasi terindeks *Web of Science* menggunakan metode *naïve bayes*” yang bertujuan untuk mengklasifikasi kategori jenis bidang keilmuan yang dimiliki jurnal tersebut hanya dilihat dari judul saja tanpa adanya membaca secara keseluruhan, dan jurnal-jurnal tersebut bisa masuk lebih dari satu ataupun lebih masuk kategori bidang ilmu. sehingga perlu dibuat sebuah sistem pengelompokan atau klasifikasi menurut jenis kategori yang semestinya untuk menentukan bidang keilmuan yang sesuai dengan jurnal jurnal, dan mempermudah dalam mencari ataupun memahami isi jurnal.

1.2 Perumusan Masalah

Rumusan masalah dalam penelitian ini adalah bagaimana metode *Naïve Bayes* dapat diterapkan pada publikasi terindeks *Web of Science*.

1.3 Pembatasan Masalah

Batasan-batasan masalah dalam penelitian ini yaitu :

1. Data didapatkan dari artikel terindeks *Web of Science* yang tersimpan pada database sinta.
2. Klasifikasi pada penelitian ini hanya dibatasi pada 5 bidang ilmu menggunakan metode *Naïve Bayes*.
3. Data yang digunakan dalam penelitian ini hanya berdasarkan judul publikasi.

1.4 Tujuan

Tujuan dari penelitian ini adalah merancang sebuah sistem yang mampu mengklasifikasikan data WoS (*Web of Science*) yang terindeks SINTA sesuai

dengan 5 bidang ilmu menggunakan *Naïve Bayes*.

1.5 Manfaat

Manfaat yang diharapkan dari pembuatan sistem ini adalah memudahkan dalam mencari ataupun memahami keseluruhan isi jurnal tanpa harus membaca total isi jurnal.

1.6 Sistematika Penulisan

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

BAB I PENDAHULUAN

Dalam bab I penulis mendeskripsikan latar belakang, pemilihan judul, rumusan masalah, batasan masalah, tujuan penelitian, metodologi penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA DAN DASAR TEORI

Dalam bab II berisi tentang artikel penelitian-penelitian sebelumnya serta berisi dasar-dasar teori yang berguna sebagai penunjang permasalahan yang dibahas dalam penelitian ini, juga membantu penulis dalam memahami dan mendalami konsep kerja sistem pada *website* untuk mengklasifikasikan judul dengan bahasa pemrograman dan *framework* yang sudah dipilih.

BAB III METODE PENELITIAN

Dalam bab III menyajikan proses tahap penelitian dimulai dengan perancangan sistem klasifikasi judul dengan metode *naïve bayes*, alur kerja sistem, desain aplikasi, dan pengujian aplikasi.

BAB IV HASIL PENELITIAN

Dalam bab IV, menjelaskan gambaran hasil desain, hasil penelitian berupa hasil pengujian sistem klasifikasi salah satu sampel judul, hasil akurasi, analisis hasil akurasi dan pembahasan hal yang mempengaruhi hasil akurasi.

BAB V KESIMPULAN DAN SARAN

Dalam bab V, penulis mendeskripsikan isi dari kesimpulan yaitu berupa proses penelitian dari awal sampai dengan proses paling akhir dan juga berisi saran yang akan dikembangkan sebagai pelengkap untuk menyempurnakan tugas akhir ini.

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Dalam sebuah penelitian perlu adanya bahan tinjauan Pustaka guna menjadi sebagai bahan acuan pertimbangan dan referensi, adapun beberapa penelitian sebelumnya yang menjadi bahan pertimbangan dan referensi yang terhubung akan judul penelitian ini, diantaranya yaitu sebagai berikut :

Dalam penelitian sebelumnya oleh Nurdin pada tahun 2021 tentang pengklasifikasian karya tulis ilmiah berupa tugas besar akhir dari mahasiswa dengan menerapkan algoritma *Naïve Bayes Classifier*. Pada penelitian tersebut menggunakan data dari rincian 170 artikel karya ilmiah yang terbagi 150 data latih, 20 data uji. Hasil dalam penelitian tersebut menunjukkan algoritma Naive Bayes Classifier adalah sebuah algoritma yang mudah. Dapat diterapkan dalam pengklasifikasian artikel ilmiah menghasilkan akurasi rata-rata 86.68%. Waktu proses rata-rata untuk setiap tes adalah 5,7406 detik/tes. (Nurdin dkk., 2021)

Dalam penelitian sebelumnya oleh Devita dan kawan-kawan pada tahun 2018 tentang mencari nilai perbandingan dari kinerja metode Naïve Bayes dan K-Nearest Neighbor dalam mengklasifikasikan data berupa artikel berbahasa Indonesia. Pada penelitian tersebut, metode Naive Bayesian dipilih dengan alasan dapat mencapai akurasi yang sangat akurat dengan data training yang lebih sedikit. Metode K-nearest neighbor dipilih karena kuat terhadap data yang berisik. Hasil perbandingan kinerja kedua metode menentukan metode apa yang tepat untuk mengklasifikasi dokumen. Dan hasil yang diperoleh yaitu Naive Bayes metode memberikan nilai akurasi yang tinggi dari metode (KNN)K-Nearest Neighbor yaitu hingga 70% untuk naive bayes metode dan 40% untuk (KNN)K-Nearest Neighbor metode. (Devita dkk., 2018)

Dalam penelitian sebelumnya oleh Fitria dan Azis pada tahun 2018 tentang analisa kinerja dari sistem berupa pengklasifikasian skripsi dengan menerapkan metode algoritma Naïve Bayes Classifier. Hasil pengujian penelitian tersebut, pengujian data menerapkan metode naive bayes berupa judul skripsi menghasilkan akurasi 50%, sedikit banyaknya data yang diproses sangat mempengaruhi hasil

akhir dalam pengujian. (Fitria & Azis, 2018)

Dalam penelitian sebelumnya oleh Erdwika Putri pada tahun 2018 tentang pengelompokan data dari Jurnal scimago dan country rank dengan metode algoritma C4.5. Pada penelitian tersebut jurnal internasional yang telah terindeks di jurnal scimago dan country rank of scopus, klasifikasi menggunakan pohon keputusan atau decision tree menerapkan algoritma C4.5, alasan penggunaan metode tersebut supaya orang biasa dan pengguna dapat dengan mudah memahami algoritma C4.5, karena dalam algoritma tersebut bisa menangani data kategori ataupun data numerik. (Merangin dkk., 2018)

Dalam penelitian sebelumnya oleh Hardianti dan kawan-kawan tahun 2018 tentang penggunaan metode naïve bayes dalam mengklasifikasikan judul-judul jurnal. Pada penelitian tersebut, Bayesian diterapkan dalam pengklasifikasian judul- judul jurnal dengan memperhitungkan akurasi data. Selain itu, menghapus kata berhenti menghilangkan kata-kata yang tidak masuk akal, menghasilkan pengelompokan kata dasar dari judul jurnal, dan juga indeks terbalik, yang merupakan matriks antara istilah dan tanggal. Berdasarkan pengujian penelitian tersebut, hasil keakuratan data sebanyak 50%. (Hardianti dkk., 2018)

Dalam penelitian sebelumnya oleh Nuraeni dan kawan-kawan pada tahun 2021 tentang perbandingan dari metode algoritma naïve bayes classifier (Nbc) dan metode algoritma decision tree dalam menganalisa sebuah sistem pengklasifikasian judul skripsi. Pada penelitian tersebut menggunakan algoritma metode naïve bayes dan algoritma decision tree dan menghasilkan perbandingan indeks akurasi persamaan dari algoritma naïve bayes sebesar 55% pada metode 30 pohon keputusan, Kesimpulan pohon keputusan mempunyai nilai akurasi persentase lebih tinggi. Perbandingan nilai tingkat akurasi dari hasil pengujian, algoritma pohon keputusan menghasilkan tingkat akurasi lebih tinggi dari algoritma naïve bayes yaitu dengan perbandingan akurasi sebesar 90% berbanding 89,58. (Nuraeni dkk., 2021)

Dalam penelitian sebelumnya oleh Muhabatin dan kawan-kawan pada tahun 2021 tentang Pengklasifikasian Berita Hoax Dengan Metode Naïve Bayes Berbasis PSO. Proses klasifikasi dilakukan dalam tahap pre-processing, setelah itu kata-kata

diberi bobot dan diklasifikasikan dengan Naive Bayes. Setelah proses uji selesai, didapatkan bahwa metode naïve bayes menghasilkan akurasi 73%, dan metode naïve bayes PSO menghasilkan akurasi 91%.(Muhabatin dkk., 2021)

Dalam penelitian sebelumnya oleh Sholih dan kawan-kawan pada tahun 2021 tentang pengklasifikasian berupa judul-judul dari berita online dengan studi kasus daerah seputar Banjarmasin menerapkan metode naïve bayes. Dalam penelitian tersebut, data headline berita diuji dengan metode Naive Bayes. Metode tersebut bisa digunakan dalam mengklasifikasikan berita-berita online tersebut. Dengan memakai rincian data 400 juga 4 kelas, ditemukan metode Naive Bayes dari uji yang dibangun dalam Rapidminer mencapai akurasi sebesar 78,75%. Dengan skor recall 80,56, akurasi 78,75%.(Sholih dkk., 2021)

Dalam penelitian sebelumnya oleh Sabrani dan kawan-kawan tentang Metode Multinomial Naïve Bayes dalam mengklasifikasikan artikel online seputar gempa yang ada di Indonesia. Penelitian tersebut menguji keefektifan metode proba multinomial naïve bayes mengklasifikasikan artikel-artikel online dengan topik seputar gempa bumi yang ada di Indonesia dengan menerapkan teknologi TF-IDF. Proses uji meliputi dua fitur, fitur pertama unigram dan kedua berupa fitur bigram, dan juga melakukan uji kombinasi kedua fitur tersebut. Selain itu, juga diuji dengan menghapus penghapusan kata awal dari langkah pra-pemrosesan. F-Measure tertinggi yang dicapai sebesar 95,20%. Metode polynomial Naïve Bayesian dapat diterapkan pada kasus klasifikasi artikel dimana penelitian mencapai hasil akurasi tinggi yaitu berupa nilai f-measure sebesar 95,20 dengan standar deviasi 1,58% menguji 5 ulangan dengan 5 kali persilangan hasil validasi.(Sabrani dkk., 2020)

Dalam penelitian sebelumnya oleh Widaningsih dan kawan-kawan tentang Klasifikasi Jurnal Ilmu Komputer Berdasarkan Pembagian Web Of Science Dengan Menggunakan Text Mining. Pada penelitian tersebut pengelompokkan karya ilmiah dalam bidang ilmu komputer bahasa Inggris dilakukan menurut pengkategorian penelitian yang berdasarkan Web of Science antara lain yaitu kecerdasan buatan, sibernatika, perangkat keras dan arsitektur, sistem informasi, aplikasi interdisipliner, rekayasa perangkat lunak, teori dan metode serta pencitraan dan fotografi. Dari masing-masing kategori diambil sebanyak 50 data, dengan rincian

40 data training dan 360 data testing, dengan jumlah sebanyak 400. Pengklasifikasian dengan tahapan text mining dengan dua algoritma metode, berupa metode pertama (NBC) Naive Bayes Classifier dan metode kedua (SPM) Support Vector Machine. Tahapan penambangan teks menerapkan perangkat lunak rapidminer versi 8.0, dengan kinerjanya adalah Recall, Accuracy, F-Measure dan Precision. Berbeda dengan Support Vector Machine metode, hasil nilai dari naive Bayes classifier lebih efisien yaitu dengan nilai sebesar 64,90%, akurasi sebesar 69,23%, F-Measure 66,99n dengan akurasi sebesar 64,42%.(Widaningsih dkk., 2018)

Dalam penelitian sebelumnya oleh Wibowo dan kawan-kawan pada tahun 2022 tentang metode algoritma k-nearest neighbor, neural network dan naïve bayes, untuk pengklasifikasian topik untuk tugas akhir. Pada penelitian tersebut jenis klasifikasi paling efisien untuk klasifikasi mata kuliah atau judul skripsi ditentukan dengan membandingkan model klasifikasi. Pada penelitian tersebut mengembangkan tiga jenis model klasifikasi yaitu menggunakan metode algoritma k-nearest neighbor, neural network dan naïve bayes. Hasil evaluasi menerapkan metode fusion matrix untuk menentukan presisi, akurasi, recall dan fscore. Hasil percobaan menghasilkan dengan pengklasifikasian menerapkan metode neural network mendapatkan akurasi fading tinggi sebesar 94%, sedangkan akurasi fading rendah dengan model pengklasifikasian naive bayes sebesar 79%.(Wibowo dkk., 2022)

Dalam penelitian sebelumnya oleh Feri irfanto dan kawan-kawan pada tahun 2021 tentang Pengelompokkan Dokumen-dokumen Skripsi menerapkan metode Text Mining dengan studi kasus FTI(Fakultas Teknologi Informasi). Pada penelitian tersebut menerapkan metode Naive Bayes Classifier, sebuah metode pengklasifikasian yang mencari nilai proba atau nilai kemungkinan dengan proses menghitung nilai frekuensi dengan penggabungan nilai didalam data set. algoritma tersebut bertujuan mengelompokkan dari data uji yang sesuai dengan atribut data latih. Data mentah berupa data mahasiswa-mahasiswa dari (FTI) Fakultas Teknologi Informasi yang telah lulus studi. Pengkategorian ada 5 yaitu, Pengolahan Citra, Keamanan Sistem dan Jaringan, SPK, RPL dan Data Mining. Dalam tahap

pengujian pengelompokan dokumen skripsi menggunakan naïve bayes classifier, ada 5 tahapan yaitu pertama input data latih, kedua preprocess, ketiga menghitung nilai (kemunculan kata), keempat proses perhitungan nilai proba kata-kata dari data latih, dan proses terakhir adalah perhitungan nilai proba maksimal dari semua kategori yang ada. Rincian data yaitu 49, 15 data uji dan 34 data latih. hasil perhitungan 1.5 data uji didapatkan nilai sebesar 14 data telah terkategori dengan tepat, 1 sisanya terkategori salah. Didapatkan hasil akurasi dari sistem tersebut sebesar 93%.(Studi dkk., 2021)

Ditinjau dari beberapa uraian tinjauan pustaka diatas, maka dalam penelitian ini ingin melakukan pengklasifian jurnal yang telah terindeks *Web of Science* kedalam 5 bidang ilmu yaitu (Arts & Humanities, engineering & Technology, Life Sciences & Medicine, Natural Sciences, dan Social Sciences & Management) menggunakan metode *Naïve Bayes*, dimana dengan pengklasifikasian berupa menerapkan metode naïve bayes supaya mudah untuk dipahami serta kemampuannya untuk merincikan dan memangkas proses dalam pengambilan keputusan yang kompleks menjadi lebih simple, sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan dan merumuskan masalah berupa memudahkan dalam pengelompokan jurnal ilmiah sesuai akan topik.

2.2 Dasar Teori

2.2.1 Klasifikasi

Dalam data mining, proses pembersihan data perlu adanya pengelompokan-pengelompokan data atau biasa disebut dengan klasifikasi. Klasifikasi dilakukan agar data dapat tertata disusun secara rapi menurut jenis ciri -cirinya masing-masing. Klasifikasi merupakan suatu proses pengelompokan data berdasarkan ciri – ciri persamaan maupun perbedaan dan berdasarkan keterikatan data terhadap data sampel. (Oktanisa & Supianto, 2018)

2.2.2 Publikasi

Publikasi adalah kegiatan menyebarkan informasi, konten, produk, dan pengetahuan untuk menarik perhatian dan minat masyarakat. Publikasi juga dapat diartikan suatu bentuk kegiatan penyebaran informasi, konten, produk dan

pengetahuan untuk menarik perhatian dan minat masyarakat. Publikasi merupakan suatu langkah yang harus dilewati peneliti agar karyanya dari hasil riset yang telah dilakukan, bisa diketahui dan tersebar luas menyeluruh dan juga telah dipublikasikan, yang nantinya suatu saat dapat berkontribusi dan dipergunakan dalam pengembangan ilmu. (Studi & Kimia, 2018)

2.2.3 Data Mining

Data mining merupakan tahapan berupa penggalian informasi, digunakan untuk pengklasifikasian maupun prediksi, dimana dikemas dalam bentuk menganalisis data, bisa dipergunakan dalam pengambilan sampel, yang biasa memakai ilmu kecerdasan buatan, statistika, matematika, dan juga teknik pembelajaran mesin untuk memproses, menyebarkan, mengidentifikasi, dan juga menyimpan informasi-informasi penting yang dibutuhkan. (Utomo & Mesran, 2020)

2.2.4 Preprocessing

Preprocessing adalah suatu proses berisi tahapan menghilangkan problem atau permasalahan di dalam data yang mungkin dapat mengganggu hasil daripada proses data, sehingga data dapat digunakan dalam kondisi yang ideal. (Najjichah dkk., 2019). Preprocessing didefinisikan sebagai suatu proses pengolahan data mentah atau data asli sebelum data tersebut nantinya akan digunakan, dan data dalam keadaan data bersih siap dipakai ke tahap selanjutnya. Adapun beberapa tujuan dari proses preprocessing yaitu selain membersihkan data, juga memperkecil atau memperbesar data, memperjelas fitur data, mengkonversi data asli ke bentuk data yang sesuai dibutuhkan, juga guna menghilangkan noise. (Ma'rifah dkk., 2020)

Preprocessing merupakan sebuah tahapan dalam data mining yang diperlukan untuk memaksimalkan kinerja dalam algoritma pengklasifikasian. Pada umumnya tahapan dalam preprocessing, antara lain yaitu sebagai berikut : data cleaning (pembersihan data), case folding (pengubahan), tokenizing (memecah), stopwords (eliminasi) dan stemming (mengubah).

a. Data Cleaning

Data Cleaning adalah sebuah tahap berupa pembersihan isi dalam data dari kumpulan tanda baca, simbol ataupun icon seperti (!, @, #, \$, ^, &, *, dan lain

sebagainya). Juga merupakan suatu tahapan proses dimana jika ada karakter ataupun tanda baca yang tidak diperlukan maka akan dihapus. Ini dapat berguna untuk meminimalisir kesalahan dalam sebuah data. Seperti contoh karakter-karakter bisa bersihkan yaitu URL link, tagar (#), (@), (#), (\$) dan tanda-tanda baca lainnya.(Lubis, 2020)

b. Case Folding

Case folding biasanya didefinisikan berupa perubahan semua huruf, mulai dari huruf kapital besar ke bentuk huruf kecil. Seperti contoh “DEFINISI” dalam kata tersebut mengandung huruf besar kapital lalu akan diubah ke dalam huruf kecil yaitu menjadi “definisi”. Proses case folding berupa meratakan seluruh isi teks menjadi lowercase juga menyingkirkan karakter-karakter yang bersifat non-word seperti contoh berupa symbol-simbol, karakter tanda baca dan juga angka-angka dan yang disisakan dan dibutuhkan hanya berupa teks alfabet yang memuat huruf a sampai huruf z. (Rahutomo dkk., 2019)

c. Tokenizing

Tokenizing, merupakan proses memecah dokumen atau data menjadi satuan kata-per kata. tahap pemotongan kalimat menjadi lebih kecil atau perkata token. Adanya tokenizing digunakan dalam proses stopwords removing yang berbasis dengan kamus berjalan, dimana di setiap ada pengulangan pada setiap kata dalam data.(Prasetya, 2022)

d. Stopword

Stopwords merupakan kata-kata yang muncul dengan rating persentase nilai tinggi di dalam dokumen akan tetapi tidak memuat nilai atau makna informasi yang tepat. Seperti contoh stopwords misalnya dalam bahasa inggris yaitu kata or, the, and, a, of, in, as, is, this, that. Dengan adanya penghapusan kata-kata stopwords memiliki efek yang sangat signifikan untuk meningkatkan hasil nilai akurasi dalam classifier.(Wibisono dkk., 2020)

e. Stemming

Stemming yaitu sebuah tahap perubahan kata dari hasil proses filtering (penyaringan) ke dalam bentuk dasar awal, yaitu lewat mengeliminasi kata

tambahan yang ada dalam kata tersebut, misal membuat menjadi buat, menjual menjadi jual, menjaga menjadi jaga, dan lain-lain.(Dwi Normawati, 2021)

2.2.5 PHP

PHP atau (*Hypertext Preprocessor*) merupakan sebuah bahasa pemrograman sangat terkenal, biasa digunakan dalam pembuatan atau pengembangan aplikasi website. Hingga saat ini, php terus berkembang dengan versi terbaru saat ini yaitu versi 8.1.6. PHP bisa menjadi salah satu pilihan bahasa pemrograman yang cocok bagi pemula yang ingin belajar di dunia pemrograman, karena memiliki referensi yang banyak, mudah diakses, *open source* atau gratis dan juga mudah diaplikasikan dengan macam-macam sistem operasi, contoh Windows, Mac Os, Linux. Dalam pembuatan sistem website ini menggunakan PHP versi 8.1.6.(Fadila dkk., 2019)

2.2.6 SINTA

SINTA (Indeks Sains dan Teknologi) adalah portal yang memuat ukuran kinerja ilmiah dan teknologi, termasuk kinerja peneliti/penulis/penulis, kinerja jurnal, kinerja lembaga iptek, dan lainnya. SINTA adalah pusat pengindeksan kutipan dan pakar berbasis web terbesar di Indonesia, mengukur kinerja peneliti serta lembaga yang berdasarkan dengan publikasi yang dihasilkan, kinerja jurnal berdasarkan jumlah artikel serta kutipan yang dihasilkan. Menyediakan akses pengukuran yang cepat, mudah dan komprehensif. (Fadhilaturrahmi dkk., 2020)

2.2.7 Web Of Science

Web of Science (WoS) merupakan sebuah layanan pengindeksan kutipan karya ilmiah yang berbasis langganan dimana pada awalnya layanan pengindeksan ini dikelola oleh Institute for Scientific Information dan saat ini dikelola oleh Clarivate Analytics dimana menyediakan pencarian kutipan yang komprehensif. Dalam layanan ini menyediakan akses ke berbagai database yang bertindak sebagai referensi penelitian, yang memungkinkan penelitian mendalam di bidang akademik atau disiplin ilmu tertentu. (Yaman dkk., 2019)

2.2.8 Naïve Bayes

a. Pengertian *Naïve Bayes*

Dalam penelitian pengklasifikasian judul publikasi terindeks Web of Science ini, metode yang digunakan adalah metode *Naïve Bayes*. *Naïve Bayes* adalah

algoritma yang umum digunakan untuk pengambilan keputusan, *Naive Bayes* menerapkan teknik pembelajaran mesin yang menggunakan probabilitas dan perhitungan statistik untuk memprediksi tingkat probabilitas atau kemungkinan di masa depan dengan berdasarkan dari pengalaman masa lalu. (Nuraeni dkk., 2021)

b. Persamaan *Naive Bayes Classifier*

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

Ket :

C : Class (kelas)
 W_i : kata yang sama ke (i)
 (W_i,C) : jumlah total kata (w_i) yang ada di dalam C
 count(C) : jumlah total kata yang ada di kelas C
 |V| : jumlah total (Vocabulary) atau total semua kata

2.2.9 5 Bidang Ilmu

Ada beberapa referensi dalam membagi bidang ilmu publikasi diantaranya adalah QS RANKING yang merupakan publikasi tahunan peringkat universitas yang dilakukan oleh *Quacquarelli Symonds* (QS), organisasi yang melakukan pemeringkatan perguruan tinggi di seluruh dunia berdasarkan 5 bidang ilmu yaitu :

a. Arts & Humanities

Ilmu humaniora cenderung “memanusiakan” manusia, berurusan dengan aspek perilaku manusia yang bertujuan menjadikan manusia lebih manusiawi, termasuk nilai-nilai kemanusiaan, studi agama, filsafat, seni, sejarah, dan linguistic.

b. Engineering & Technology

Ilmu rekayasa teknologi adalah penerapan ilmu pengetahuan dan teknologi yang digunakan untuk memecahkan masalah di dunia nyata manusia sehari-hari yang berkaitan dengan penerapan prinsip-prinsip dasar teknik dan rekayasa.

c. Life Sciences & Medicine

Ilmu hayat mencakup semua bidang ilmu terhubung satu dengan yang lain meliputi studi ilmu ilmiah yang membahas organisme hidup seperti contoh mikroba, hewan, manusia, tumbuhan dan lain sebagainya. Ilmu kehidupan

berkontribusi pada peningkatan kualitas hidup dan standar pokok dalam kehidupan. Ilmu kehidupan diterapkan dan digunakan di dalam bermacam-macam ilmu, seperti contoh ilmu kesehatan, ilmu pertanian, ilmu kedokteran, ilmu industri, ilmu farmasi, ilmu makanan dan lain sebagainya.

d. Natural Sciences

Ilmu-ilmu alam adalah ilmu-ilmu yang secara sistematis mengetahui alam dengan mempelajari apa yang terjadi di alam sekitar kehidupan manusia. Sains adalah cabang ilmu pengetahuan yang didasarkan pada klasifikasi pengamatan dan data, dikelompokkan dan disahkan dengan hukum kuantitatif dan menyangkut dengan penerapan penalaran matematis analisis data dalam fenomena alam.

e. Social Sciences & Management

Ilmu sosial mempertimbangkan berbagai masalah sosial, terutama yang diungkapkan oleh publik dalam berbagai interpretasi (fakta, konsep, dan teori) dari berbagai bidang keahlian ilmiah di bidang ilmu-ilmu sosial. Sosiologi, antropologi, psikologi sosial, dll. Ilmu manajemen adalah bidang sosial yang mempelajari serangkaian proses yang berjalan secara simultan untuk mencapai tujuan organisasi dan individu dengan meliputi tahapan perencanaan, pengorganisasian, pengarahan, pengendalian, dan pengawasan.

2.2.10 Evaluasi

Evaluasi merupakan tahapan proses mengidentifikasi berupa perhitungan kembali mencari nilai terbaik dari suatu pengujian, dengan tujuan dalam proses kegiatan yang telah dilakukan sesuai dengan rencana yang ditargetkan atau tidak. Salah satunya adalah dengan metode *Confusion Matrix*.

Confusion Matrix merupakan suatu metode yang digunakan dalam perhitungan mencari nilai akurasi dalam data mining. Adapun berikut adalah gambaran tabel dari *confusion matrix*.

Tabel 2. 1 Confusion Matrix

		Kelas Hasil Prediksi	
		Tidak (+)	Ya (-)
Kelas Asli	Tidak (+)	TP (<i>True Positive</i>)	FN (<i>False Negative</i>)
	Ya (-)	FP (<i>False Positive</i>)	TN (<i>True Negative</i>)

Tahapan evaluasi *confusion matrix* berupa nilai dari akurasi (*accuracy*), presisi (*precision*), dan pemanggilan kembali (*recall*). *Accuracy* di dalam pengklasifikasian yaitu berupa nilai persentase keakuratan dari data yang diuji, sedangkan *precision* merupakan penjabaran porsi dari nilai prediksi berupa positif dan positif benar dalam data mentah. *Recall* merupakan pembagian pembahasan akan positif benar, diprediksikan dengan positif dengan tepat. Dengan persamaan :

$$a. \text{ Accuracy (Akurasi)} = (TP + TN) / (TP + FP + FN + TN) \quad (3)$$

$$b. \text{ Precision (Presisi)} = TP / (TP + FP) \quad (4)$$

$$c. \text{ Recall (Pemanggilan kembali)} = TP / (TP + FN) \quad (5)$$

$$d. \text{ F1 score} = (\text{precision} \times \text{recall}) / (\text{precision} + \text{recall}) \quad (6)$$

Keterangan :

1. *True Positive* (TP) yaitu nilai pengujian positif yang dikategorikan menjadi positif
2. *True Negative* (TN) yaitu jumlah pengujian negative yang dikategorikan menjadi negatif
3. *False Positive* (FP) yaitu jumlah pengujian negative yang dikategorikan menjadi positif
4. *False Negative* (FN) yaitu jumlah pengujian positif yang dikategorikan menjadi negative

BAB III METODE PENELITIAN

3.1 Metode Penelitian

Dalam tahapan penelitian, menerapkan metode algoritma naïve bayes. Metode naïve bayes tersebut akan melakukan pengklasifikasian dari koleksi data WoS didalam SINTA ke dalam 5 bidang ilmu. Adapun langkah tahap yang dilakukan selama melakukan penelitian ini, antara lain:



Gambar 3. 1 Flowchart Metode Penelitian

Tahapan dalam metode penelitian tugas akhir ini ditunjukkan pada gambar 3.1 mencakup beberapa aktivitas yaitu Pengumpulan Data, Pre-processing, Penerapan Metode, dan Evaluasi. Dengan penjelasan sebagai berikut :

1. Pengumpulan data

Pengumpulan data merupakan tahap mengumpulkan data berupa judul-judul penelitian yang nantinya akan dipergunakan dalam proses pengklasifikasian *Naïve Bayes*. Data yang telah dicari berdasarkan dengan tujuan prosedur akan digunakan. Koleksi data pada penelitian ini diambil sesuai dengan 5 bidang ilmu yaitu Arts &

Humanities, engineering & Technology, Life Sciences & Medicine, Natural Sciences, dan Social Sciences & Management. Jumlah masing-masing data yang akan dipakai untuk eksperimen dalam penelitian ini sejumlah masing-masing 1000 judul publikasi yang mencakup 5 bidang ilmu.

2. Pre-processing

Pre-processing merupakan tahap dimana terjadinya proses seleksi data agar data yang diambil lebih terstruktur dalam setiap dokumen mengubah data yang sudah dikumpulkan lewat berbagai sumber agar menjadi sebuah data bersih dan dapat digunakan mengolah data berkelanjutan. Tahap preprocessing dalam penelitian ini yaitu pengkodean data pada judul artikel di SINTA yang terindeks Web Of Science.

3. Penerapan metode

Penerapan metode dalam penelitian ini yaitu menerapkan metode algoritma *naïve bayes*. Penerapan metode *naïve bayes* tidak perlu menggunakan data pelatihan (training) dalam jumlah besar, akan tetapi dalam jumlah data kecil sudah mampu bekerja untuk menentukan estimasi berapa besar perhitungan yang akan diperlukan dalam proses pengklasifikasian nantinya. Dalam penerapan metode *naïve bayes* akan menggunakan perhitungan *naïve bayes classifier*.

4. Evaluasi

Evaluasi merupakan sebuah tahapan dalam penelitian dimana prosesnya berisi pengidentifikasian atau penafsiran penilaian yang bertujuan mengukur perhitungan dan penilaian apakah semua tahapan proses dalam penelitian yang telah dikerjakan tepat sesuai perencanaan dan berhasil dengan tujuan yang diharapkan atau sebaliknya tidak. Salah satunya adalah menggunakan *Confusion Matrix*. *Confusion Matrix* merupakan sebuah proses tahapan yang dipakai dalam perhitungan mencari nilai utama yaitu akurasi dalam data mining.

3.2 Preprocessing

Dalam tahapan preprocessing pada penelitian ini adalah berisi pembersihan data, Adapun proses dalam pembersihan data memuat 5 tahapan, yaitu :

- a. Data Cleaning, yaitu proses untuk menghapus tanda baca, angka, simbol-simbol

(!, :, @, #, \$, %, ^, dll) dan juga karakter khusus pada teks atau dokumen. Seperti contoh terdapat dalam tabel 3.1

Tabel 3. 1 Proses data cleaning

No.	Data Input	Data Output
1.	How to write systematically in works and scientific journals that are correct and structured?	How to write systematically in works and scientific journals that are correct and structured
2.	What are the correct and sequential steps in preparing the final project report?	What are the correct and sequential steps in preparing the final project report

- b. Case folding, yaitu proses untuk mengganti semua huruf di dalam data ke huruf kecil (lowercase). Contoh seperti terlihat dalam tabel 3.2

Tabel 3. 2 Proses case folding

No.	Data Input	Data Output
1.	How to write systematically in works and scientific journals that are correct and structured	how to write systematically in works and scientific journals that are correct and structured
2.	What are the correct and sequential steps in preparing the final project report	what are the correct and sequential steps in preparing the final project report

- c. Tokenizing, yaitu proses pemotongan teks kalimat atau paragraf menjadi bagian-bagian yang lebih kecil, yang disebut token. Contoh seperti terlihat dalam tabel

3.3

Tabel 3. 3 Proses tokenizing

No.	Data Input	Data Output
1.	how to write systematically in works and scientific journals that are correct and structured	how / to / write / systematically / in / works / and / scientific / journals / that / are / correct / and / structured
2.	what are the correct and sequential steps in preparing the final project report	what / are / the / correct / and / sequential / steps / in / preparing / the / final / project / report

- d. Stop Word, yaitu proses memilah dan diambil hanya kata - kata penting saja.. dalam pemrosesan filter atau pemilahan bisa dengan menerapkan algoritma *stoplist* yaitu (mengeliminasi kata-kata kurang penting ataupun tidak penting) dan algoritma *wordlist* yaitu (penyimpanan kata-kata penting). Tahap pemilihan kata yang bernilai dari hasil token dengan menyimpan kata yang bernilai dan membuang kata yang tidak bernilai. Contoh terdapat dalam tabel 3.4

Tabel 3. 4 Proses stop word

No.	Data Input	Data Output
1.	how to write systematically in works and scientific journals that are correct and structured	how write systematically works scientific journals correct structured
2.	what correct sequential steps preparing final project report	what correct sequential steps preparing final project report

- e. Stemming, *stemming* merupakan proses menghilangkan infleksi kata imbuhan (afiks) baik yang terdapat dalam awalan (prefiks), sisipan (infiks), akhiran

(sufiks), maupun kombinasi dari awalan dan akhiran (konfiks) dan perubahan kata ke dalam bentuk kata dasar awal, namun dalam kata dasar awal tersebut tidak mengartikan makna sesuai atau sama dari kata akar (*root word*). Contoh terlihat dalam tabel 3.5

Tabel 3. 5 Proses stemming

No.	Data Input	Data Output
1.	how write systematically works scientific journals correct structured	how write systematically works scientific journals correct structured
2.	what correct sequential steps preparing final project report	what correct sequential steps preparing final project report

Penerapan kode program preprocessing terdapat dalam gambar 3.2 pada gambar tersebut memuat kode program fungsi proses di dalam tahap *preprocessing*, yaitu tertulis dengan syntax `Multibyte Text Normalizer` yang memuat fungsi `case folding` dan `lowercase`, juga pada syntax `StopWord Filter` yang memuat Fungsi `stopword`.

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

Gambar 3. 2 Penerapan kode program preprocessing

3.3 Penerapan Metode Naïve Bayes

Dalam penerapan metode Naïve Bayes ada penerapan metode melalui pencarian nilai secara manual lewat rumus persamaan Naïve Bayes dan melalui kode program. Untuk penerapan metode melalui rumus persamaan Naïve Bayes yaitu sebagai berikut :

Langkah pertama yaitu pengambilan data, data yang diambil berupa judul dengan topik tentang “Covid-19”. Rincian data berupa data training dan data testing. Data training diambil sampel sebanyak 5 judul yang mewakili bidang ilmu masing masing, dan data testing uji sebanyak 1. Untuk data training terlihat dalam tabel 3.6

Tabel 3. 6 Rincian data testing dan data training

Data		Teks	Class
Training	D1	Keberagamaan, kenyamanan, dan kemarahan siswa Indonesia terhadap Tuhan selama pandemi COVID-19	Art and Humanities
	D2	Contextual Based E-learning (CBE): Model Baru Pembelajaran Online di Dinas Kesehatan Masyarakat untuk Pembelajaran Selama Pandemi Covid-19	Engineering and Technology
	D3	Stroke iskemik akut sebagai komplikasi pada covid-19 dengan sindrom gangguan pernapasan akut di unit perawatan intensif	Life Science and Medicine
	D4	Karakteristik Masyarakat Indonesia Dalam Memanfaatkan Herbal untuk Pencegahan Covid di Masa Pandemi Covid-19	Natural Science
	D5	Reaksi pasar terhadap pandemi Covid-19: Studi peristiwa pada saham yang terdaftar di indeks LQ45	Social Science and Management
Testing	D6	Covid-19 dalam teologi islam dan dampaknya terhadap bidang sosial keagamaan di Indonesia	

Langkah kedua yaitu pemecahan judul dari data training dan testing menjadi per kata atau *token*, seperti terdapat dalam tabel 3.7

Tabel 3. 7 Pemecahan data training dan data testing menjadi perkata(token)

No	Bidang Ilmu	Perkata (<i>token</i>)
D1	Art and Humanities	Keberagaman / kenyamanan / dan / kemarahan / siswa / Indonesia / terhadap / Tuhan / selama / pandemi / COVID-19
D2	Engineering and Technology	Contextual / Based / E learning(CBE) / model / baru / pembelajaran / online / di / Dinas / Kesehatan / Masyarakat / untuk / Pembelajaran / Selama / Pandemi / Covid-19
D3	Life Science and Medicine	Stroke / iskemik / akut / sebagai / komplikasi / pada / Covid-19 / dengan / sindrom / gangguan / pernapasan / akut / di / unit / perawatan / intensif
D4	Natural Science	Karakteristik / masyarakat / Indonesia / dalam / memanfaatkan / Herbal / Untuk / Pencegahan / Covid / Di / Masa / Pandemi / Covid-19
D5	Social Science and Management	Reaksi / pasar / terhadap / pandemi / Covid-19 / Studi / peristiwa / pada / saham / yang / terdaftar / di / indeks / LQ45
D6	?	Covid-19 / dalam / teologi / islam / dan / dampaknya / terhadap / bidang / sosial / keagamaan / di / Indonesia

Setelah pemecahan kalimat menjadi per kata (token) kemudian masuk ke tahapan pengklasifikasian dari perkata kedalam kategori bidang ilmu dengan lambang pengkategorian (A = Art and Humanities, B = Engineering and Technology, C = Life Science and Medicine, D = Natural Science, E = Social Science and Management) terlihat dalam tabel 3.8

Tabel 3. 8 Pengkategorian perkata(token) ke bidang ilmu yang sesuai

No	Vocabulary	Kategori bidang ilmu	Class sama	Jumlah
1	Keberagaman	A		A = 4
2	Kenyamanan	A		
3	kemarahan	A		
4	siswa		Indonesia	
5	terhadap			
6	Tuhan	A		
7	selama			
8	pandemi		Covid-19	
9	Contextual	B		B = 6
10	Based	B		
11	E-learning	B		
12	Model	B		
13	Baru			
14	Pembelajaran	B		
15	Online	B		
16	Dinas	E	Selama	
17	Kesehatan	C	Pandemi	C = 11
18	Masyarakat		Covid-19	
19	Stroke	C		
20	iskemik	C		
21	akut	C		
22	sebagai			
23	komplikasi	C	Covid-19	
24	sindrom	C		
25	gangguan	C		
26	pernapasan	C		
27	akut			
28	unit	C		
29	perawatan	C		
30	intensif	C		
31	Karakteristik	D		D = 6
32	Masyarakat		Indonesia	
33	Memfaatkan	D		
34	Herbal	D		
35	Pencegahan	D	Covid-19	
36	Masa	D	Pandemi	
37	Reaksi	D		
38	pasar	E	terhadap	E = 8
39	Studi	E	pandemi	
40	peristiwa	E	Covid-19:	
41	saham	E		
42	terdaftar	E		

43	indeks	E		
44	LQ45	E		
45				
46				Jumlah = 35

Setelah pengkategorian didapatkan hasil nilai dari : Kategori bidang ilmu A = 4, B = 6, C = 11, D = 6, E = 8 dan jumlah vocabulary (total kata/token) = 35. Selanjutnya masuk ke tahap penerapan rumus persamaan Naïve Bayes dengan fungsi sebagai berikut terlihat dalam gambar 3.3

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

Gambar 3. 3 Persamaan naive bayes

Dari persamaan fungsi tersebut, langkah selanjutnya yaitu perhitungan dengan data testing dengan acuan perkata(token) dari data training, dengan rincian hasil perhitungan terlihat dalam tabel 3.9, tabel 3.10, tabel 3.11, tabel 3.12 dan tabel 3.13.

Bidang ilmu 1 (Art and Humanities)

Tabel 3. 9 Perhitungan dari bidang ilmu 1 (Art and Humanities)

Covid 19	(1 + 1) : (4 + 35)	0,0512820513
dalam	(0 + 1) : (4 + 35)	0,0256410256
Teologi	(0 + 1) : (4 + 35)	0,0256410256
Islam	(0 + 1) : (4 + 35)	0,0256410256
Dampak	(0 + 1) : (4 + 35)	0,0256410256
Terhadap	(0 + 1) : (4 + 35)	0,0256410256
Bidang	(0 + 1) : (4 + 35)	0,0256410256
sosial	(0 + 1) : (4 + 35)	0,0256410256
Keagamaan	(0 + 1) : (4 + 35)	0,0256410256
Indonesia	(1 + 1) : (4 + 35)	0,0512820513
Jumlah		= 4,91376E-16

Bidang Ilmu 2 (Engineering and Technology)

Tabel 3. 10 Perhitungan dari bidang ilmu 2 (Engineering and Technology)

Covid 19	(1 + 1) : (6 + 35)	0,0487804878
dalam	(0 + 1) : (6 + 35)	0,0243902439
Teologi	(0 + 1) : (6 + 35)	0,0243902439
Islam	(0 + 1) : (6 + 35)	0,0243902439
Dampak	(0 + 1) : (6 + 35)	0,0243902439
Terhadap	(0 + 1) : (6 + 35)	0,0243902439

Bidang	$(0 + 1) : (6 + 35)$	0,0243902439
sosial	$(0 + 1) : (6 + 35)$	0,0243902439
Keagamaan	$(0 + 1) : (6 + 35)$	0,0243902439
Indonesia	$(0 + 1) : (6 + 35)$	0,0243902439
Jumlah		= 3,63419E-18

Bidang Ilmu 3 (Life Science and Medicine)

Tabel 3. 11 Perhitungan dari bidang ilmu 3 (Life Science and Medicine)

Covid 19	$(1 + 1) : (11 + 35)$	0,0434782609
dalam	$(0 + 1) : (11 + 35)$	0,0217391304
Teologi	$(0 + 1) : (11 + 35)$	0,0217391304
Islam	$(0 + 1) : (11 + 35)$	0,0217391304
Dampak	$(0 + 1) : (11 + 35)$	0,0217391304
Terhadap	$(0 + 1) : (11 + 35)$	0,0217391304
Bidang	$(0 + 1) : (11 + 35)$	0,0217391304
sosial	$(0 + 1) : (11 + 35)$	0,0217391304
Keagamaan	$(0 + 1) : (11 + 35)$	0,0217391304
Indonesia	$(0 + 1) : (11 + 35)$	0,0217391304
Jumlah		= 4,71467E-17

Bidang Ilmu 4 (Natural Science)

Tabel 3. 12 Perhitungan dari bidang ilmu 4 (Natural Science)

Covid 19	$(1 + 1) : (6 + 35)$	0,0487804878
dalam	$(0 + 1) : (6 + 35)$	0,0243902439
Teologi	$(0 + 1) : (6 + 35)$	0,0243902439
Islam	$(0 + 1) : (6 + 35)$	0,0243902439
Dampak	$(0 + 1) : (6 + 35)$	0,0243902439
Terhadap	$(0 + 1) : (6 + 35)$	0,0243902439
Bidang	$(0 + 1) : (6 + 35)$	0,0243902439
sosial	$(0 + 1) : (6 + 35)$	0,0243902439
Keagamaan	$(0 + 1) : (6 + 35)$	0,0243902439
Indonesia	$(1 + 1) : (6 + 35)$	0,0487804878
Jumlah		= 2,98004E-16

Bidang Ilmu 5 (Social Science and Management)

Tabel 3. 13 Perhitungan dari bidang ilmu 5 (Social Science and Management)

Covid 19	$(1 + 1) : (8 + 35)$	0,0465116279
dalam	$(0 + 1) : (8 + 35)$	0,023255814
Teologi	$(0 + 1) : (8 + 35)$	0,023255814
Islam	$(0 + 1) : (8 + 35)$	0,023255814
Dampak	$(0 + 1) : (8 + 35)$	0,023255814
Terhadap	$(1 + 1) : (8 + 35)$	0,0465116279
Bidang	$(0 + 1) : (8 + 35)$	0,023255814
sosial	$(0 + 1) : (8 + 35)$	0,023255814

Keagamaan	$(0 + 1) : (8 + 35)$	0,023255814
Indonesia	$(0 + 1) : (8 + 35)$	0,023255814
Jumlah		= 1,85087E-16

Dari hasil perhitungan ke 5 bidang ilmu, didapatkan masing-masing hasil dengan rincian Bidang ilmu 1 (Art and Humanities = 4,91376E-16), Bidang Ilmu 2 (Engineering and Technology = 3,63419E-18), Bidang Ilmu 3 (Life Science and Medicine = 4,71467E-17), Bidang Ilmu 4 (Natural Science = 2,98004E-16) dan Bidang Ilmu 5 (Social Science and Management = 1,85087E-16). Dari kelima bidang hasil perhitungan yang paling besar dan mendekati adalah bidang ilmu 1 yaitu (Art and Humanities sebesar = 4,91376E-16), dengan begitu maka data training masuk kedalam kategori bidang keilmuan yang pertama yaitu “Art and Humanities”.

Untuk penerapan metode melalui kode program yaitu sebagai berikut. Penerapan kode program metode naïve bayes dapat dilihat dalam gambar 3.3 yang memuat kode program fungsi proses yang ada pada tahap penerapan metode *naïve bayes*.

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

Gambar 3. 4 Penerapan metode naive bayes

3.4 Evaluasi

Penerapan kode program evaluasi yang dapat dilihat dalam gambar 3.4 yang memuat kode program fungsi proses pencarian nilai confusion matrix, akurasi, presisi dan recall.

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

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

echo $results;
```

Gambar 3. 5 Evaluasi (confusion matrix, akurasi, presisi dan recall)

3.5 Metode Perancangan Alur Sistem

3.5.1 Analisa Kebutuhan

Analisis kebutuhan merupakan sebuah tahapan dimana sistem ini dianalisa tentang fungsi apa saja yang harus bisa dilakukan sistem ini dalam melakukan proses fungsional program, mulai dari input, proses klasifikasi, perhitungan prediksi, menampilkan output berupa nilai akurasi, sampai dengan unduh data set pada sistem ini. Berikut adalah penjelasan mengenai fungsi-fungsi tersebut :

A. Input Judul

Input Judul merupakan tahap proses pertama dalam sistem ini nantinya, dimana sistem ini membutuhkan inputan berupa sebuah judul yang selanjutnya akan digunakan untuk melakukan ke proses selanjutnya yaitu klasifikasi. Sebelumnya sistem ini sudah menyimpan database berupa dataset publikasi judul.

B. Proses klasifikasi

Klasifikasi merupakan proses dimana sistem ini akan mengolah judul dari hasil inputan judul pada tahap sebelumnya, menerapkan proses pengklasifikasian dengan algoritma metode Naïve Bayes.

C. Melakukan prediksi

Dalam proses prediksi, sistem melakukan perhitungan prediksi dari sebuah judul yang telah diinputkan dan akan divalidasi oleh sistem dengan database dataset, sistem ini akan melakukan pencarian nilai akurasi terdekat, dimana dalam mencari probabilitas terdekat sistem ini akan menggunakan metode Naïve Bayes.

D. Menampilkan hasil prediksi

Output hasil prediksi, menampilkan hasil dari tahap sebelumnya yaitu perhitungan prediksi nilai akurasi, fungsi ini menampilkan hasil prediksi berupa nilai akurasi, bertujuan untuk menampilkan hasil dari pemrosesan uji yang telah berjalan, dalam fungsi ini akan menampilkan berapa hasil tingkat persentase nilai akurasi terdekat dari setiap kelas yang dimiliki 5 bidang ilmu yang diprediksi.

E. Unduh data set

Pada fungsi unduh data set adalah proses dimana user dapat mengunduh data yang ada pada program tersebut, yaitu berupa judul publikasi terindeks Web of Science sejumlah 5000 data dengan hasil unduhan format excel.

3.5.2 Implementasi Sistem

Tahap Analisis Sistem, tahap dimana menganalisa *tools* yang akan dipergunakan sebagai penunjang dalam pembangunan sistem ini, dan berikut adalah beberapa *tools* yang digunakan dalam sistem ini:

A. Visual Studio Code

Visual Studio Code sering disebut VScode merupakan aplikasi pengeditan open source dikembangkan oleh Microsoft. VS code memberikan akses kemudahan dalam penulisan kode, yaitu dengan mendukung bermacam-macam jenis bahasa coding pemrograman, seperti contoh bahasa pemrograman Java, Python, C++, C#, PHP dan lain sebagainya. Kode Visual Studio terhubung dengan platform Github. Contoh fitur lainnya yaitu dapat menambahkan plugin, di mana pengembang dapat menambahkan fungsionalitas yang belum disertakan dalam Visual Studio Code. Dalam pembangunan sistem, Bahasa pemrograman yang dipergunakan yaitu PHP. Visual Studio Code terdapat fitur yaitu Extensions Marketplace yaitu fitur untuk menginstal *tools* pendukung dari Bahasa yang digunakan, adapun *tools* pendukung yang digunakan seperti Html Css support dan CSV rainbow.

B. PHP

Pada penelitian ini menggunakan bahasa pemrograman PHP. PHP adalah bahasa dalam pemrograman berjalan pada sisi server, artinya dapat menangani perintah tugas seperti mengumpulkan data, mengolah file didalam server, memodifikasi database, mengupdate data dan lain-lain. PHP dapat dipakai dalam pembuatan website, dengan menyisipkan PHP dalam kode HTML dan CSS, website yang dibuat bisa berjalan secara dinamis.

C. RubixML

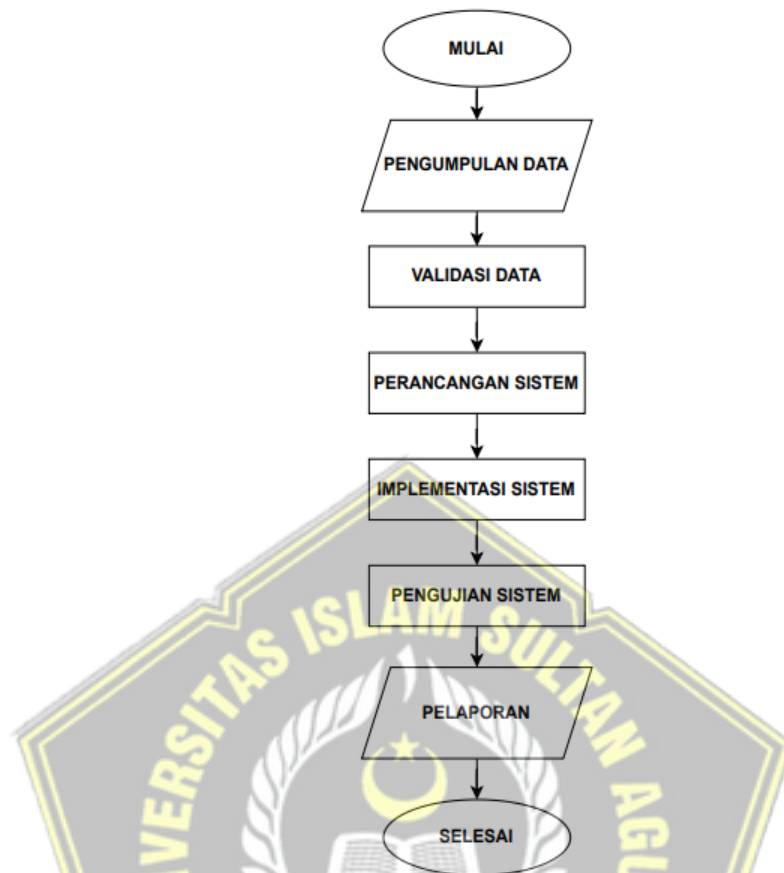
Pustaka library yang digunakan dalam penelitian ini adalah RubixML. Rubix ML adalah perpustakaan sumber terbuka gratis untuk bahasa PHP, juga menjadi sebuah sistem machine learning proses penukaran dan penyimpanan data yang didalamnya menggunakan Bahasa pemrograman PHP.

D. CSV

Format penyimpanan data pada penelitian ini adalah dalam bentuk CSV (Comma Separated Values) dimana bentuk penyimpanan data dalam file dengan nilai-nilai dipisahkan dengan tanda koma. Data berupa judul dan jenis bidang ilmu dipisah dengan tanda koma. Seperti contoh sample : Primary localized cutaneous amyloidosis: A clinical diagnosis, Life Sciences and Medicine.

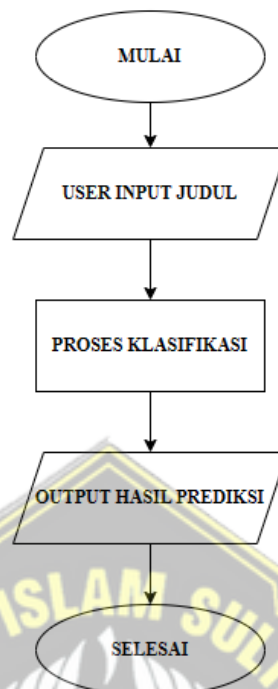
3.5.3 Analisis Alur Sistem

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 3.6



Gambar 3. 6 Alur perancangan sistem

Pada gambar 3.6 Diperlihatkan urutan dalam perancangan sistem dimana pada tahap pertama terdapat, pengumpulan data dari database SINTA, tahapan selanjutnya validasi data, kemudian jika data benar langkah selanjutnya yaitu perancangan alur sistem dan dilanjutkan ke dalam implementasi *coding* dan selanjutnya adalah penerapan metode. Setelah itu, pengujian sistem dan yang terakhir adalah pelaporan. Tahap selanjutnya yaitu alur dari sebuah berjalannya program atau alur sistem seperti dalam gambar 3.7



Gambar 3. 7 Flowchart Sistem

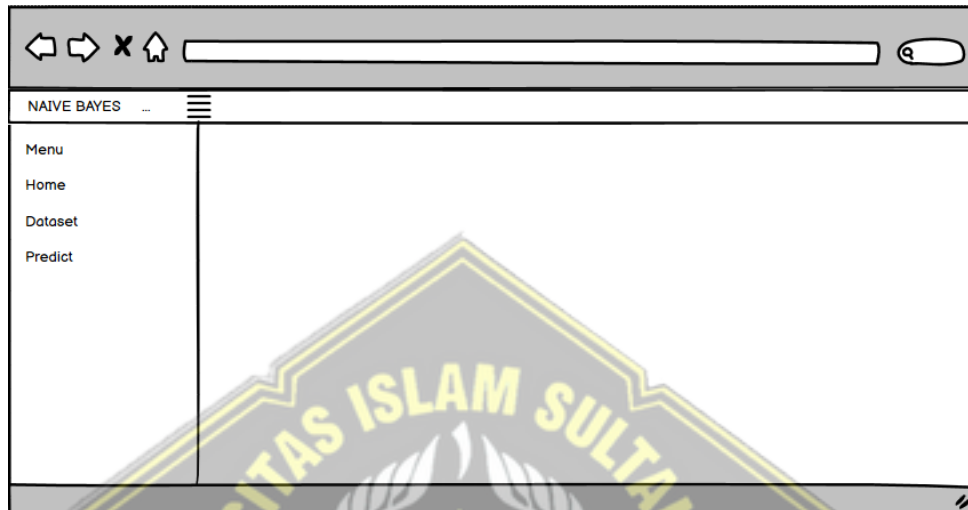
Pada gambar 3.7 Adalah flowchart dari sistem ini nantinya, alur dari penggunaan sistem ini nantinya akan diawali dengan user memulai menggunakan device yang akan mereka gunakan, dan dalam penelitian ini, penulis menggunakan laptop sebagai device utama, pada bagian awal dari sistem ini, user diharuskan membuka sistem ini, setelah itu user akan melihat beberapa menu seperti Home, halaman dataset, dan halaman predict. User kemudian memilih halaman predict. Dalam halam predict user diharuskan memasukkan judul yang akan diprediksi. Setelah user memasukkan judul, dengan menekan tombol prediksi selanjutnya adalah proses klasifikasi dimana sistem mengklasifikasi judul dengan metode Naïve Bayes. Sistem ini dapat melakukan prediksi judul publikasi dari lima bidang ilmu yang disebutkan, maka alur proses dari sistem ini telah selesai.

3.6 Perancangan Antarmuka

Pada bagian perancangan antarmuka adalah desain mockup yang akan dibuat dalam Sistem. Berikut adalah rancangan antarmuka dari Sistem klasifikasi bidang ilmu pada publikasi terindeks *Web of Science* menggunakan Metode Naïve Bayes.

1. Halaman utama

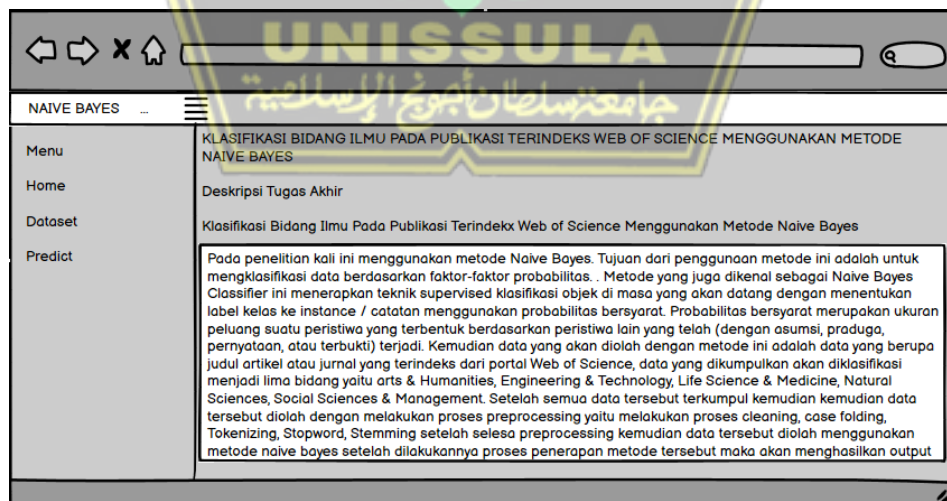
Halaman utama website penelitian ini bisa dilihat dalam gambar 3.8 Didalam halaman ada beberapa tampilan opsi pilihan dengan fungsi tersendiri, antara lain yaitu Home, Dataset, dan Predict.



Gambar 3. 8 Halaman utama

2. Halaman Home

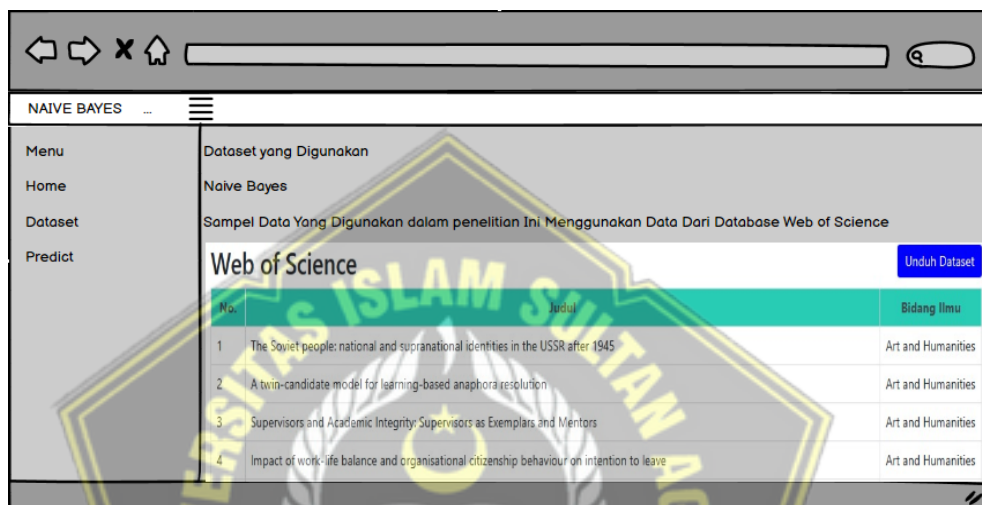
Halaman home merupakan halaman dimana terdapat penjelasan singkat deskripsi dari penelitian ini. Desain rancangan halaman home dapat dilihat pada gambar 3.9



Gambar 3. 9 Halaman utama

3. Halaman Dataset

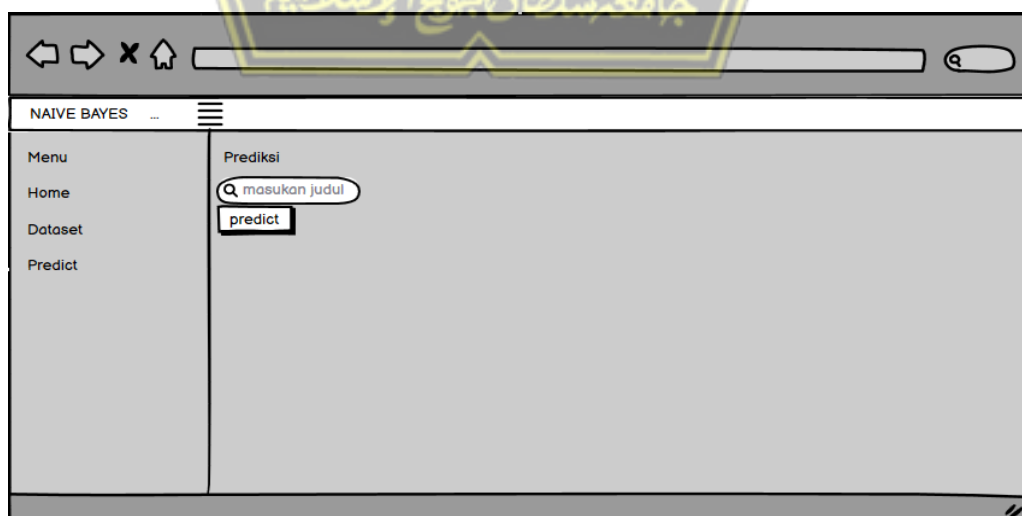
Halaman Dataset merupakan halaman dimana terdapat data berupa judul-judul yang telah terpublikasi indeks Web of Science pada SINTA sejumlah 5000 judul, dalam halaman Dataset juga terdapat fungsi Unduh Dataset dimana data dapat diakses dengan mengunduhnya. Desain rancangan halaman Dataset bisa dilihat dalam gambar 3.10



Gambar 3. 10 Halaman menu dataset

4. Halaman Predict

Halaman Predict memuat fungsi-fungsi dari website yang dibuat, yaitu fungsi untuk mencari hasil prediksi dari sebuah inputan judul dari dataset. Desain rancangan halaman predict terdapat dalam gambar 3.11



Gambar 3. 11 Halaman predict

BAB IV

HASIL DAN ANALISIS PENELITIAN

4.1 User Interface dan Penggunaan Sistem

Tampilan *user* atau bisa disebut dengan *user interface*, penting karena merupakan bagian yang sering dilihat dan diinteraksikan. berikut adalah tampilan dari user interface dan penggunaan sistem dari sistem berbasis web yang telah dibuat oleh peneliti:

A. Halaman Home

Menu pertama adalah Home. Pada halaman ini menampilkan gambar dan tulisan penjelasan dari deskripsi tugas akhir, *user* hanya dapat melihat bagaimana tahapan sistem yang dibuat peneliti beserta penjelasan deskripsi tugas akhir dari sistem tersebut.

Gambar 4.1 merupakan gambaran dari menu home dimana menampilkan berisi judul 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 science, social sciences & management*.

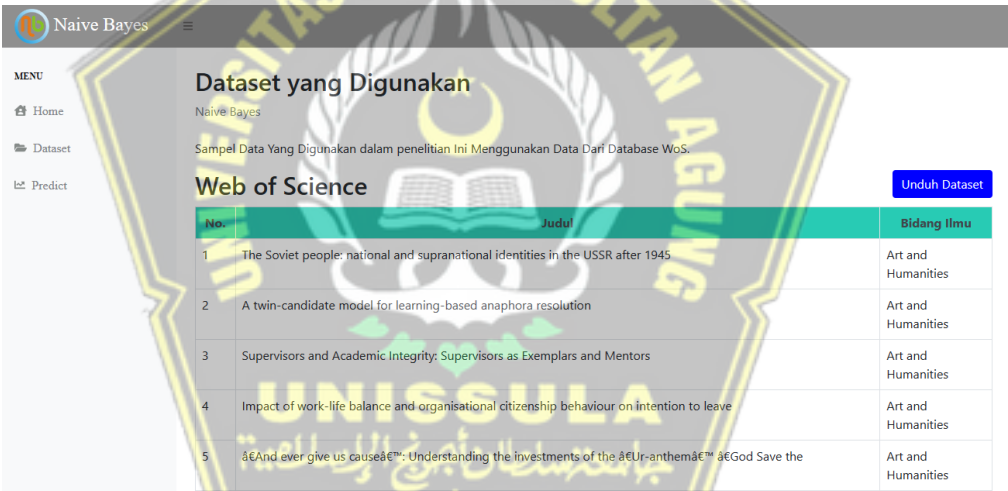


Gambar 4. 1 Halaman Home

B. Halaman Data Set

Menu kedua adalah Dataset. Pada halaman ini memiliki fitur unduh dan menampilkan tabel berisikan data *Web of Science* dimana tabel menjabarkan judul dari masing-masing 5 bidang ilmu. *User* dapat melihat hasil data pada tabel dan mengunduh dataset tersebut pada tombol yang telah disediakan. Hasil dari unduhan berupa format excel.

Gambar 4.2 merupakan tampilan dari dataset dimana pada halaman dataset tersebut menampilkan data Web of science yang telah diberi label sesuai dengan bidang ilmunya. Data tersebut berjumlah 5000 judul dengan masing-masing bidang ilmu terdapat 1000 judul artikel jurnal publikasi. pada halaman dataset juga terdapat menu unduh dataset yang berfungsi untuk mendownload dataset tersebut.



No.	Judul	Bidang Ilmu
1	The Soviet people: national and supranational identities in the USSR after 1945	Art and Humanities
2	A twin-candidate model for learning-based anaphora resolution	Art and Humanities
3	Supervisors and Academic Integrity: Supervisors as Exemplars and Mentors	Art and Humanities
4	Impact of work-life balance and organisational citizenship behaviour on intention to leave	Art and Humanities
5	“And ever give us cause”: Understanding the investments of the “Ur-anthem” “God Save the	Art and Humanities

Gambar 4. 2 Tampilan halaman dataset

Dalam gambar 4.2 juga terdapat menu download button, dimana tombol tersebut berfungsi untuk mendownload dataset yang terdapat pada halaman tersebut, dimana *user* cukup hanya menekan tombol tersebut, kemudian dataset secara otomatis akan terdownload ke dalam device user.



Unduh Dataset

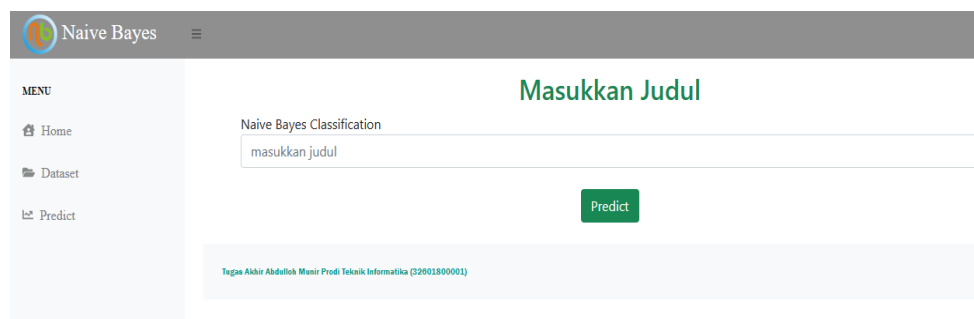
Gambar 4. 3 Tampilan fungsi unduh dataset

Dalam gambar 4.3 merupakan tampilan setelah *user* berhasil mendownload dataset tersebut, kemudian dengan otomatis data tersebut akan terunduh ke dalam folder download di perangkat pengguna *user* dengan format data excel CSV.

C. Halaman Predict

Menu terakhir adalah menu predict. Halaman ini memiliki fitur menampilkan hasil data predict yaitu dengan cara *user* memasukkan judul yang diinginkan kemudian klik tombol predict untuk melihat hasil dari data tersebut. Hasil data berupa judul, akurasi, dan bidang ilmu nya.

Pada gambar 4.4 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.



Gambar 4. 4 Tampilan halaman predict sebelum input judul

Halaman predict

Masukkan Judul

Naive Bayes Classification

The Soviet people: national and supranational identities in the USSR after 1945
 Art and Humanities = 0.99996400129596
 Life Sciences & Medicine = 8.9996760112349E-6
 Social Sciences and Management = 8.9996760112349E-6
 Natural Sciences = 8.9996760112349E-6
 Engineering & Technology = 8.9996760112349E-6

Gambar 4. 5 Hasil Tampilan output hasil predict

Pada gambar 4.5 merupakan tampilan output hasil prediksi dimana tampilan output tersebut berupa judul artikel beserta persentase akurasi di setiap masing-masing bidang ilmu, dengan menampilkan hasil akurasi judul artikel tersebut masuk kedalam bidang ilmu dari nilai yang paling tinggi dari setiap hasil.

4.2 Analisa dan Pengujian

Dalam tahapan Analisa Pengujian merupakan tahap dimana pengujian yang digunakan pada penelitian ini yaitu dengan memakai pengujian berupa black box testing, black box testing adalah metode pengujian sebuah software dengan tujuan menguji apakah sistem berjalan normal sesuai fungsi yang direncanakan, dalam black box testing memiliki beberapa 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 fitur yang terdapat pada sistem, hasil uji fungsional sistem ini terdapat dalam tabel 4.1

Tabel 4. 1 Pengujian sistem

Skenario Pengujian	Kasus Pengujian	Hasil Pengujian	Kesimpulan
Unduh Dataset	Mengunduh Sampel Data.	berhasil	sesuai
Memasukkan Judul	Memasukkan judul publikasi	berhasil	sesuai

Memprediksi Judul	Sistem memprediksi judul	berhasil	sesuai
Hasil Prediksi Judul	Output hasil prediksi judul	berhasil	sesuai

Dari hasil diatas dapat dilihat bahwa pengujian yang dilakukan sistem Klasifikasi bidang ilmu pada publikasi terindeks WoS menggunakan Metode Naïve Bayes, sudah sesuai dan dapat dijalankan setiap fungsinya. Tahapan testing berjalan dengan baik sesuai dengan fungsi utama.

4.3 Analisis Akurasi

Selanjutnya adalah Tahap Analisis Akurasi, tujuan dilakukannya tahap ini adalah untuk dapat mengetahui seberapa tingkat akurasi sistem yang telah dibuat ini, tahap uji akan menggunakan data testing, yang mana berikut adalah pembagian dari data testing dan data sampel. Tahap ini merupakan bagian dari tahapan untuk mengetahui sejauh mana tingkat akurasi sistem yang telah dibuat.

Pada tabel 4.2 merupakan rincian dari jumlah data sampel dan data testing yang digunakan, dengan data sampel sebanyak 90% = 4500 data dan data testing sebanyak 10% = 500 data.

Tabel 4. 2 Jumlah Data

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

Total Data : 5000

Dari tabel 4.3 menunjukkan rincian data sampel dan data testing yang terbagi ke dalam 5 bidang ilmu.

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 Sampel	900	900	900	900	900
Data Testing	100	100	100	100	100

Dari data tersebut nantinya akan menghasilkan akurasi berupa confusion matrix, dari confusion matrix ini akan digunakan untuk mencari nilai akurasi, presisi, recall dan F1 score. Dari hasil training yang telah di uji coba menghasilkan confusion matrix dengan nilai accuracy, precision, recall dan F1 score terdapat pada tabel 4.4

Tabel 4. 4 Hasil accuracy, precision, recall dan F1 score

Accuracy	Precision	Recall	F1 score
0.442	0.240	0.202	0.070

```
PS C:\xampp\htdocs\ta_munir> php training-single-csv.php
Training model... Model disimpan... OK {
  "breakdown": {
    "overall": {
      "accuracy": 0.4428925373134328,
      "balanced accuracy": 0.50125,
      "f1 score": 0.07073835931172416,
      "precision": 0.24008016032064128,
      "recall": 0.202,
```

Gambar 4. 6 Output hasil klasifikasi 5000 data

Dalam gambar 4.6 menunjukkan hasil pengujian dengan sampel data 5000 judul artikel menghasilkan output beberapa nilai antara lain yaitu nilai akurasi sebanyak 0.442 dengan presisi, recall, F 1 masing-masing mendapat nilai 0.240, 0.202, dan 0.070, dengan demikian menunjukkan bahwa metode Naive Bayes masih kurang maksimal untuk melakukan klasifikasi judul artikel jurnal publikasi garuda dimana metode tersebut hanya menghasilkan akurasi 0.4428 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 kemiripan judul antara bidang ilmu satu dengan bidang ilmu lainnya.

Dalam tabel 4.5 dan gambar 4.7 adalah hasil nilai-nilai dari confusion matrix dengan rincian nilai TP (true positive) = 101, TN (true negative) = 404, FP (false positive) = 399 dan FN (false negative) = 399.

Tabel 4. 5 Hasil confusion matrix

	Positive	Negative
True	101	404
False	399	399

```
"true positives": 101,
"true negatives": 404,
"false positives": 399,
"false negatives": 399,
```

Gambar 4. 7 Output hasil (TP, TN, FP, FN)

Berdasarkan pengujian-pengujian yang telah dilakukan dalam penelitian ini, kesimpulan hasil dari uji perhitungan menjelaskan bahwa, dalam pengujian menggunakan jumlah data total sebanyak 5000 data, dengan rincian pembagian data training sebanyak 90% = 4500 dan data testing sebanyak 10% = 500. Menghasilkan akurasi total yaitu 0.4428 (44%).

4.4 Pembahasan Hal Yang Mempengaruhi Hasil Akurasi

Dari hasil pengujian yang telah dilakukan, ternyata menghasilkan nilai akurasi yang cukup rendah, yaitu sebesar 0.4428 (44%). Maka dari itu dalam pembahasan hal yang mempengaruhi hasil akurasi, menjelaskan faktor-faktor apa saja yang mungkin mempengaruhi rendahnya hasil nilai akurasi pada pengujian yang telah dilakukan, Adapun faktor-faktor tersebut yang mungkin mempengaruhi hasil nilai akurasi adalah sebagai berikut :

- I. Adanya judul yang membahas tentang topik yang sama akan tetapi terbagi pada bidang ilmu yang berbeda.
 - a) Topik tentang “Covid-19” yang seharusnya lebih mengacu ke arah bidang (Art & Humanities dan Life Science & Medicine)

Tabel 4. 6 Judul dengan topik “Covid-19”

No	Judul	Bidang Ilmu
1	Personal Impact of Media Publicity During Covid-19 Pandemic and Awareness Among Generation Z in Jakarta Indonesia Genre-based visualization through an online teaching platform: A strategy to engage with academic texts during the covid-19 outbreak Words and images of Covid-19 prevention (A case study of tourism new normal protocols signs) Positive discourse analysis of the indonesian government spokespersons discursive strategies during the covid-19 pandemic	Art and Humanities
2	Information Dissemination of COVID-19 by Ministry of Health in Indonesia Setting Acceptance Criteria for a National Flocked Swab for Biological Specimens during the COVID-19 Pandemic The Benefits of e-Commerce before and during the Covid-19 Pandemic for Small Enterprises in Indonesia An integrated approach to sustainable development National Resilience and COVID-19 responses	Engineering and Technology

3	<ul style="list-style-type: none"> - Factors affecting public non-compliance with large-scale social restrictions to control COVID-19 transmission in Greater Jakarta Indonesia - Correlation between academic self-efficacy and burnout originating from distance learning among nursing students in Indonesia during the COVID-19 pandemic - Analysis of risk of exposure to COVID-19 in fishermen in Kenjeran - Anxiety and suicidal thoughts during the COVID-19 pandemic: Cross-country comparative study among Indonesian Taiwanese and Thai university students 	Life Sciences & Medicine
4	<ul style="list-style-type: none"> - Community behavior for mathematical model of coronavirus disease 2019 (COVID-19) - UBNNet: Deep learning-based approach for automatic X-ray image detection of pneumonia and COVID-19 patients 	Natural Sciences
5	<ul style="list-style-type: none"> - Transformation of the Education Sector during the COVID-19 Pandemic in Indonesia - Impact of covid-19 news on performance of indonesia stock market - Democracy and human rights during the COVID-19: the case of Indonesia 	Social Sciences and Management

- b) Topik tentang “Diagnosis” yang seharusnya lebih mengacu ke arah bidang (Life Science & Medicine)

Tabel 4. 7 Judul dengan topik “Diagnosis”

No	Judul	Bidang Ilmu
1	<ul style="list-style-type: none"> - The role of corporate reputation andamp; distinctive organization capability in developing business model innovation: Case study of Indonesian ICT firms in facing industry resolution 4.0 - Impact of long term planning on revenue on business - Management accounting information system in gas station business - The effectiveness of business partnership to the marketing of home industry products - Relevance of employability skills of informatics education students to the business and industrial sector 	Art and Humanities
2	<ul style="list-style-type: none"> - Recovery of business intelligence systems: Towards guaranteed continuity of patient centric healthcare 	Engineering and Technology
3	<ul style="list-style-type: none"> - Silk weaving business sustainability as a cultural heritage of Indonesia: A case study in Wajo Regency South Sulawesi 	Natural Sciences
4	<ul style="list-style-type: none"> - Relationship of perception and awareness towards utilization of government business support services - Business strategy of success factors of enterprise resource planning system (ERP) and the impact towards the quality of accounting information - Role of social media marketing in the successful implementation of business management - Business and data analytics: New innovations for the management of e-commerce 	Social Sciences and Management

- c) Topik tentang “Business” yang seharusnya lebih mengacu ke arah bidang (Social Science and Management)

Tabel 4. 8 Judul dengan topik "Business"

No	Judul	Bidang Ilmu
1	Incremental contribution of reported previous head injury to the prediction of diagnosis and cogniti	Art and Humanities
2	Automated diagnosis of Coronary Artery Disease affected patients using LDA PCA ICA and Discrete Wa High Density RF-DC Plasma Nitriding under Optimized Conditions by Plasma-Diagnosis	Engineering and Technology
3	Laboratory evaluation of a simple and rapid latex agglutination assay for the serodiagnosis of typhoid fever Evaluation for the Clinical Diagnosis of Pythium insidiosum Using a Single-Tube Nested PCR The importance of accurate diagnosis of dengue fever Primary localized cutaneous amyloidosis: A clinical diagnosis	Life Sciences & Medicine
4	Automatic diagnosis of alcohol use disorder using EEG features	Social Sciences and Management

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari hasil penelitian ini, ditarik kesimpulan bahwasanya dengan mengklasifikasikan bidang ilmu publikasi terindeks Web of Science menggunakan metode Naïve bayes menghasilkan output yaitu kesesuaian judul dari artikel dalam SINTA yang sudah terindeks Web of Science dalam 5 bidang ilmu tersebut. Hasil dari pengujian sistem berupa nilai akurasi, recall, presisi. Di dapatkan nilai akurasi 44,2%, recall 24% dan presisi 20,2%.

Namun hasil tersebut masih tergolong rendah jika digunakan pada sistem SINTA, dikarenakan dataset yang didapatkan kurang baik. Dari percobaan pengujian data yang telah dilakukan, hasil nilai yang didapatkan tidak sesuai dengan yang diharapkan, akan tetapi dari segi sistem, sistem dapat bekerja dengan baik sebagaimana fungsinya mampu mengklasifikasi dan sistem dapat bekerja dengan baik pengoperasiannya mampu memberikan kejelasan.

5.2 Saran

Dari uraian yang telah dijelaskan dan hasil kesimpulan, saran dalam penelitian ini antara lain sebagai berikut :

1. Untuk penelitian selanjutnya, perlu adanya tambahan metode algoritma lain, seperti metode K-nearest neighbors, sebagai acuan untuk membandingkan hasil nilai akurasi yang tertinggi.
2. Pemilihan judul yg lebih merujuk, judul sesuai dengan bidang ilmu yang dikategorikan.
3. Untuk sistem tepat pada desain lebih dirancang sedemikian menarik mungkin.
4. Pengembangan pada sistem bisa ditambahkan dengan berbagai fitur.
5. Sistem hanya dalam bentuk website, untuk kedepanya bisa diupgrade ke bentuk versi mobile.

DAFTAR PUSTAKA

- Bimo, P., Setio, N., Retno, D., Saputro, S., & Winarno, B. (2020). *Klasifikasi dengan Pohon Keputusan Berbasis Algoritma*. 3, 64–71.
- 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>
- Dwi Normawati, J. S. (2021). *Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter*. 5(September), 697–711.
- Fadhilaturrahmi, F., Erlinawati, E., & Ananda, R. (2020). Workshop Sinta 2 dan Google Scholar di Universitas Pahlawan Tuanku Tambusai. *Jurnal Abdidas*, 1(4), 203–209. <https://doi.org/10.31004/abdidas.v1i4.45>
- Fadila, R. R., Aprison, W., Musril, H. A., Putiah, K., Birugo, A., & Baleh, T. (2019). *Perancangan Perizinan Santri Menggunakan Bahasa Pemrograman PHP / MySQL Di SMP Nurul Ikhlas*. 11(2), 84–95.
- Fitria, A., & Azis, H. (2018). Analisis Kinerja Sistem Klasifikasi Skripsi menggunakan Metode Naïve Bayes Classifier. *Prosiding Seminar Nasional Ilmu Komputer dan Teknologi Informasi*, 3(2), 102–106.
- Hardianti, A. T., Manga, A. R., & Darwis, H. (2018). Penerapan metode Naive Bayes pada klasifikasi judul jurnal. *Prosiding Seminar Nasional Ilmu Komputer dan Teknologi Informasi*, 3(2), 97. <http://e-journals.unmul.ac.id/index.php/SAKTI/article/view/1838/pdf>
- Li, K., Rollins, J., & Yan, E. (2018). Web of Science use in published research and review papers 1997–2017: a selective, dynamic, cross-domain, content-based analysis. *Scientometrics*, 115(1), 1–20. <https://doi.org/10.1007/s11192-017-2622-5>
- Lubis, H. (2020). *Implementasi K-NN Dalam Analisa Sentimen Riba Pada Bunga Bank Berdasarkan Data Twitter*. 4(April), 369–376. <https://doi.org/10.30865/mib.v4i2.2051>

- Ma'rifah, H., Wibawa, A. P., & Akbar, M. I. (2020). Klasifikasi Artikel Ilmiah Dengan Berbagai Skenario Preprocessing. *Sains, Aplikasi, Komputasi dan Teknologi Informasi*, 2(2), 70. <https://doi.org/10.30872/jsakti.v2i2.2681>
- Merangin, D. I. D., Pattiselanno, F., Mentansan, G., Nijman, V., Nekaris, K. A. I., Pratiwi, A. I. N., Studi, P., Nutrisi, I., Makanan, D. A. N., Peternakan, F., Penulisan, P., Ilmiah, K., Berbagai, P., Cahaya, I., Lapangan, D. I., Eropa, A., Geometry, R., Analysis, G., Nasution, R. D., ... Bismark, M. (2018). Analisis struktur ovarium indikator terkait kesehatan pada orang tua di rumah dengan fokus pada persepsi kesehatan subjektif. *Jurnal Studi Lingkungan Asyut*, 2(2), 2016. <https://dspace.uui.ac.id/handle/123456789/7761>
- Muhabatin, H., Prabowo, C., Ali, I., Lukman Rohmat, C., Rizki Amalia, D., sitasi, C., & Rizki, D. (2021). Classification of Hoax News Using Naïve Bayes Algorithm Based on PSO. *Informatics for Educators and Professionals*, 5(2), 156–165.
- Najjichah, H., Syukur, A., & Subagyo, H. (2019). Pengaruh Text Preprocessing Dan Kombinasinya Pada Peringkat Dokumen Otomatis Teks Berbahasa Indonesia. *Jurnal Teknologi Informasi*, XV(1), 1–11.
- Nuraeni, R., Sudiarjo, A., & Rizal, R. (2021). Perbandingan Algoritma Naïve Bayes Classifier dan Algoritma Decision Tree untuk Analisa Sistem Klasifikasi Judul Skripsi. *Innovation in Research of Informatics (INNOVATICS)*, 3(1), 26–31. <https://doi.org/10.37058/innovatics.v3i1.2976>
- 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>
- Oktanisa, I., & Supianto, A. A. (2018). Perbandingan Teknik Klasifikasi Dalam Data Mining Untuk Bank a Comparison of Classification Techniques in Data Mining for. *Teknologi Informasi dan Ilmu Komputer*, 5(5), 567–576. <https://doi.org/10.25126/jtiik20185958>
- Prasetya, F. (2022). Analisis Data Mining Klasifikasi Berita Hoax COVID 19 Menggunakan Algoritma Naive Bayes. 4(September), 132–139. <https://doi.org/10.30865/json.v4i1.4852>

- Rahutomo, F., Pratiwi, I. Y. R., & Ramadhani, D. M. (2019). Eksperimen Naïve Bayes Pada Deteksi Berita Hoax Berbahasa Indonesia. *Jurnal Penelitian Komunikasi Dan Opini Publik*, 23(1). <https://doi.org/10.33299/jpkop.23.1.1805>
- Sabrani, A., Putu, I. G., Wedashwara, W., & Bimantoro, F. (2020). *METODE MULTINOMIAL NAÏVE BAYES UNTUK KLASIFIKASI ARTIKEL ONLINE TENTANG GEMPA DI INDONESIA (Multinomial Naïve Bayes Method for Classification of Online Article About Earthquake in Indonesia)*. 2(1), 89–100.
- Sholih, M., Muzakir, M., Al, I., Al, G., Teknik, M., Universitas, I., Yogyakarta, A., Utara, R. R., & Catur, C. (2021). *TEXT MINING UNTUK MENGLASIFIKASI JUDUL BERITA ONLINE STUDI KASUS RADAR BANJARMASIN*. 08(2), 199–208.
- Studi, P., & Kimia, T. (2018). *PELATIHAN PUBLIKASI KARYA ILMIAH BAGI GURU SMK AL-INABAHAH KECAMATAN*. 2(November).
- Studi, P., Teknik, S., Informasi, F. T., Hasyim, U., Studi, P., Teknik, S., Informasi, F. T., Hasyim, U., Studi, P., Manajemen, D., Informasi, F. T., & Hasyim, U. (2021). (*STUDI KASUS: FAKULTAS TEKNOLOGI INFORMASI*) *Feri Irfanto Aries Dwi Indriyanti Dharma Bagus Pratama Putra Abstrak*. 1–9.
- Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Dataset Penyakit Jantung. *Jurnal Media Informatika Budidarma*, 4(2), 437. <https://doi.org/10.30865/mib.v4i2.2080>
- Wibisono, A., Rizkiono, S. D., & Wantoro, A. (2020). *FILTERING SPAM EMAIL MENGGUNAKAN METODE NAIVE BAYES*. 1(1), 9–17.
- Wibowo, A. P., Saifudin, A., & Darmawan, A. S. (2022). *Naïve Bayes , Neural Network dan K-Nearest Neighbor untuk Klasifikasi Topik Tugas Akhir*. 661–667.
- Widaningsih, S., Suryakencana, U., Suheri, A., & Suryakencana, U. (2018). *KLASIFIKASI JURNAL ILMU KOMPUTER BERDASARKAN PEMBAGIAN WEB OF. March*.

Yaman, A., Yoganingrum, A., & Riyanto, S. (2019). *TINJAUAN PUSTAKA SISTEMATIS PADA BASIS DATA PUSTAKA DIGITAL: TREN RISET , METODOLOGI , DAN*. 9008(21), 1–20.

