

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

Abdulloh Munir, Andi Riansyah, Sam Farisa Chaerul Haviana

Teknik Informatika, Fakultas Teknologi Industri, Universitas Islam Sultan Agung Semarang

Email : abdullohmunir@std.unissula.ac.id, andi@unissula.ac.id, sam@unissula.ac.id

Correspondence Author: abdullohmunir@std.unissula.ac.id

Abstract

Kecepatan dalam mengakses di internet menjadikan jurnal ilmiah berkembang pesat dan banyak. Adanya permasalahan seperti menemukan jurnal sesuai dengan bidang ilmu. Dari permasalahan tersebut perlu metode pengklasifikasian pencarian jurnal sesuai isi topik yang dituju, salah satunya dengan metode data mining klasifikasi dengan *naïve bayes*. Data Mining adalah proses menemukan pola bertujuan melakukan klasifikasi, prediksi, dan klaster. *Naïve Bayes* merupakan metode atas dasar probabilitas statistika dengan menerapkan teorema bayes, memprediksi peluang hasil yang akan didapatkan dengan memperhitungkan berdasarkan data pengalaman dimasa sebelumnya. Tujuan penelitian ini yaitu menerapkan metode *naïve bayes* dalam pengklasifikasian jurnal ke dalam 5 bidang ilmu dapat merumuskan masalah berupa memudahkan pengelompokan jurnal ilmiah sesuai akan topik. Hasil akhir pada penelitian ini yaitu 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 44,2%, recall 24% dan presisi 20,2%.

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

1. PENDAHULUAN

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 tahap akhir yang penting dilakukan dalam penulisan karya ilmiah. Dengan adanya publikasi seorang penulis dalam mempresentasikan apa yang sudah 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. Pengklasifikasian merupakan suatu proses pengelompokan atau pemilahan meunut jenis kesamaan atau perbedaan. Salah satu metode yang dapat digunakan adalah menggunakan metode *Naïve Bayes*. Klasifikasi merupakan suatu proses pengelompokan atau mengumpulkan benda/entitas yang sama serta memisahkan benda/entitas yang tidak sama, yang bertujuan untuk mendefinisikan kelas dari sebuah objek yang belum diketahui kelasnya, pada klasifikasi terlebih dahulu akan dilakukan proses training dan testing, pada proses tersebut akan digunakan dataset yang telah diketahui kelas objeknya. [1] 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 analitis, ini digunakan sebagai alat penelitian mendukung beragam tugas ilmiah di berbagai domain pengetahuan serta himpunan data untuk studi intensif data skala besar. [2] 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 menghitung sekumpulan probabilitas atau kemungkinan dengan menjumlahkan pasangan nilai frekuensi dari kumpulan data yang diberikan. [3]

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 menentukan kategori bidang keilmuan apa yang dimiliki jurnal tersebut tanpa harus membaca secara keseluruhan, dimana jurnal itu bisa masuk dari satu, dua atau bahkan lebih bidang keilmuan hanya dilihat dari judulnya saja, sehingga perlu dilakukan

pengelompokan atau klasifikasi untuk menentukan kategori atau bidang keilmuan yang sesuai dengan jurnal tersebut, sehingga akan mempermudah dalam mencari ataupun memahami isi jurnal. 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 terkait dengan judul penelitian ini antara lain adalah sebagai berikut :

Dalam penelitian sebelumnya oleh Nurdin pada tahun 2021 tentang Klasifikasi Karya Ilmiah (Tugas Akhir) Mahasiswa Menggunakan Metode *Naive Bayes Classifier* (NBC). Pada penelitian tersebut menggunakan data dari rincian 170 artikel ilmiah yang terbagi menjadi 150 data latih dan 20 data uji. Hasil penelitian ini menunjukkan bahwa algoritma Naive Bayes Classifier merupakan algoritma yang sederhana dan mudah. Dapat digunakan dalam klasifikasi artikel ilmiah dengan akurasi rata-rata 86,68%. Waktu proses rata-rata untuk setiap tes adalah 5,7406 detik/tes. [4] Dalam penelitian sebelumnya oleh Devita dan kawan-kawan pada tahun 2018 tentang Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor Untuk Klasifikasi Artikel Bahasa Indonesia. Pada penelitian tersebut, metode Naive Bayesian dipilih karena dapat mencapai akurasi maksimal dengan data training yang lebih sedikit. Metode K-nearest neighbor dipilih karena kuat terhadap data yang berisik. Hasil perbandingan kinerja kedua metode menentukan metode mana yang lebih cocok untuk klasifikasi dokumen. Dan hasil yang diperoleh menunjukkan bahwa metode Naive Bayes memberikan nilai akurasi yang lebih tinggi dibandingkan dengan metode K-Nearest Neighbor yaitu hingga 70% untuk Naive Bayes dan 40% untuk K-Nearest Neighbor. [3] Dalam penelitian sebelumnya oleh Fitria dan Azis pada tahun 2018 tentang Analisis Kinerja Sistem Klasifikasi Skripsi menggunakan Metode Naive Bayes Classifier. Dalam penelitian tersebut, hasil pengujian data dengan menggunakan metode Naive Bayes berupa judul skripsi menghasilkan akurasi sebesar 50%, dan banyaknya data yang digunakan dapat mempengaruhi hasil data yang diuji. [5] Dalam penelitian sebelumnya oleh Erdwika Putri pada tahun 2018 tentang Klasifikasi Data Scimago Journal dan Country Rank menggunakan Algoritma C4.5. Pada penelitian klasifikasi jurnal internasional yang terindeks di Scimago Journal & Country Rank of Scopus tersebut, klasifikasi menggunakan pohon keputusan dengan algoritma C4.5, yang menggunakan metode ini karena hasil klasifikasi dengan pohon keputusan orang biasa dan pengguna dapat dengan mudah memahami algoritma C4.5, karena algoritma ini dapat menangani data numerik atau kategori. [6]

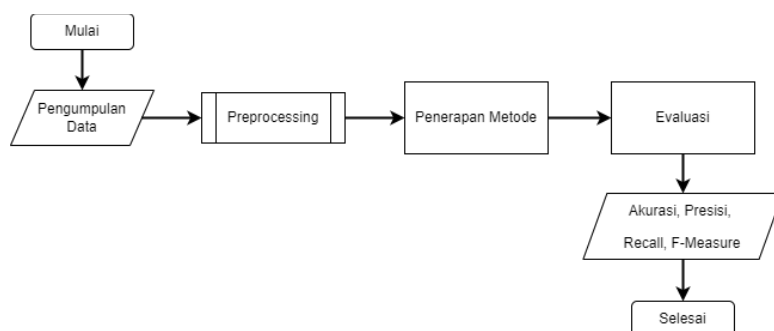
Dalam penelitian sebelumnya oleh Hardianti dan kawan-kawan tahun 2018 tentang Penerapan Metode Naive Bayes pada Klasifikasi Judul Jurnal. Pada penelitian tersebut digunakan metode Naive Bayesian untuk mengklasifikasikan judul jurnal berdasarkan disiplin ilmu dunia komputer yang termasuk dalam materi jurnal Seminar Nasional Ilmu Komputer (SNRIK) Tahun 2016 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 penelitian yang dilakukan, keakuratan data sebesar 50%, memberikan judul majalah kategori yang benar. [7] Dalam penelitian sebelumnya oleh Nuraeni dan kawan-kawan pada tahun 2021 dengan judul Perbandingan Algoritma Naive Bayes Classifier Dan Algoritma Decision Tree Untuk Analisa Sistem Klasifikasi Judul Skripsi. Pada penelitian tersebut menggunakan metode naive bayes dan decision tree dengan hasil perbandingan indeks akurasi perhitungan, naive bayes sebesar 55% pada metode 30 pohon keputusan, sehingga disimpulkan bahwa pohon keputusan memiliki persentase akurasi yang lebih tinggi. Penelitian ini dikategorikan menurut gejala umum. Dari hasil perbandingan tingkat akurasi, algoritma pohon keputusan memberikan tingkat akurasi yang lebih tinggi dibandingkan dengan algoritma naive bayes yaitu akurasi sebesar 90% berbanding 89,58. [8]

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 klasifikasi jurnal ilmiah bidang ilmu komputer dalam bahasa Inggris dilakukan menurut pembagian kategori penelitian berdasarkan Web of Science yaitu kecerdasan buatan, sibernetika, perangkat keras dan arsitektur, sistem informasi, aplikasi interdisipliner, rekayasa perangkat lunak, teori dan metode serta pencitraan dan fotografi. Dari masing-masing kategori diambil 50 jurnal, sehingga total 400 jurnal, sebanyak 360 jurnal menjadi data latih, dan 40 jurnal menjadi data uji. Proses klasifikasi menggunakan langkah-langkah text mining yang merupakan bagian dari data mining. Teknik klasifikasi menggunakan dua metode yaitu Naive Bayes Classifier dan Support Vector Machine. Proses penambangan teks menggunakan perangkat lunak rapidminer 8.0, yang ukuran kinerjanya adalah Recall, Accuracy, F-Measure dan Precision. Dibandingkan dengan metode Support Vector Machine, metode Naive Bayes Classifier memberikan nilai kinerja yang lebih baik dengan nilai kembalian sebesar 64,90%, akurasi sebesar 69,23%, F-Measure 66,99% dengan akurasi sebesar 64,42%. [9]

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 *Naive Bayes*, dimana dengan pengklasifikasian berupa menerapkan metode naive 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. 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 1. Tahapan proses penelitian

Tahapan dalam metode penelitian tugas akhir ini ditunjukkan pada gambar 1 mencakup beberapa aktivitas yaitu Pengumpulan Data, Pre-processing, Penerapan Metode, dan Evaluasi. Dengan penjelasan sebagai berikut :

2.1 Pengumpulan data

Pengumpulan data adalah tahap mengumpulkan data-data yang akan digunakan dalam proses klasifikasi *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 data yang akan digunakan untuk eksperimen pada penelitian ini sejumlah masing-masing 1000 judul publikasi yang mencakup 5 bidang ilmu.

2.2 Preprocessing

Tahap preprocessing dalam penelitian ini yaitu pengkodean data pada judul artikel di SINTA yang terindeks Web of Science. Berikut ini adalah tahapan-tahapan yang akan dilakukan pada tahapan data preprocessing:

a. Data Cleaning

Data Cleaning merupakan proses membersihkan data dengan menghapus karakter yang bukan huruf baca, symbol ataupun icon yang tidak diperlukan.[10]

b. Case folding

Case Folding merupakan proses perubahan semua huruf, mulai dari huruf kapital besar ke bentuk huruf kecil, meratakan seluruh isi teks menjadi lowercase.[11]

c. Tokenizing

Tokenizing merupakan proses memecah dokumen atau data menjadi satuan kata-per kata. tahap pemotongan kalimat menjadi lebih kecil atau perkata token.[12]

d. Stopword

Stopword merupakan tahap pemilihan dan penghapusan kata-kata yang muncul dengan rating persentase nilai tinggi di dalam dokumen akan tetapi tidak memuat nilai atau makna informasi yang tepat.

e. Stemming

Stemming merupakan tahap perubahan kata dari hasil proses filtering (penyaringan) ke dalam bentuk dasar awal, yaitu lewat mengeliminasi kata tambahan yang ada dalam kata tersebut.[13]

2.3 Penerapan *Naïve Bayes*

Penerapan metode pada penelitian ini menggunakan metode *naïve bayes*. Dalam metode naïve bayes hanya membutuhkan jumlah data pelatihan (training data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Dalam penerapan metode naïve bayes akan menggunakan perhitungan naïve bayes classifier.

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 :

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

C = Class

W_i = kata ke i

(w_i,c) = jumlah kata w_i dalam C

Count (C) = jumlah kata di class C

| V | = jumlah vocabulary

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 dan data testing terlihat pada tabel 1

Tabel 1 Rincian data testing dan data training

Data	Teks	Class
Training	D1 Keberagaman, 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 terlihat pada tabel 2

Tabel 2 Pemecahan data training dan data testing menjadi perkata(token)

No	Bidang Ilmu	Perkata (token)
D1	Art and Humanities	Keberagaman / kenyamanan / dan / kemarahan / siswa / Indonesia / terhadap / Tuhan / selama / pandemi / COVID-19
D2	Engineering and Technology	Contextual / Based / Elearning (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 per kata ke dalam 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) seperti terlihat pada tabel 3

Tabel 3 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 pada persamaan (2)

Persamaan (2) rumus Naive Bayes

$$P(W_i|C) = \frac{\text{count}(W_i, C) + 1}{\text{count}(C) + |V|} \quad (2)$$

Dari persamaan 2 tersebut, langkah selanjutnya yaitu perhitungan dengan data testing dengan acuan perkata(token) dari data training, dengan rincian hasil perhitungan terlihat pada tabel 4, 5, 6, 7 dan 8. Bidang ilmu 1 (Art and Humanities)

Tabel 4 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 5 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 6 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 7 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 8 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. Penerapan kode program metode naïve bayes dapat dilihat pada gambar 2 pada gambar tersebut pada gambar tersebut 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 2 Gambar penerapan metode naïve bayes

2.4 Evaluasi

Tahap ini mencangkup hasil eksperimen berdasarkan recall presisi dan akurasi. Tahapan ini digunakan untuk memperoleh prediksi menggunakan model algoritma yang sudah ditentukan. Tahapan ini mewakili langkah penting dalam proses membangun sebuah model. Tahapan ini mewakili langkah penting dalam proses membangun sebuah model. Untuk membuktikan kinerja algoritma yang digunakan menggunakan persamaan sebagai berikut:

- a. Akurasi di dalam pengklasifikasian yaitu berupa nilai persentase keakuratan dari data yang diuji.

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

- b. Precision merupakan penjabaran porsi dari nilai rediksi berupa positif dan positif benar dalam data mentah, tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh system.

$$Precision = \frac{TP}{TP+FP} \times 100 \% \quad (4)$$

- c. Recall merupakan pembagian pembahasan akan positif benar, diprediksikan dengan positif dengan tepat, tingkat keberhasilan system dalam menemukan kembali sebuah informasi.

$$Recall = \frac{TP}{TP+FN} \times 100 \% \quad (5)$$

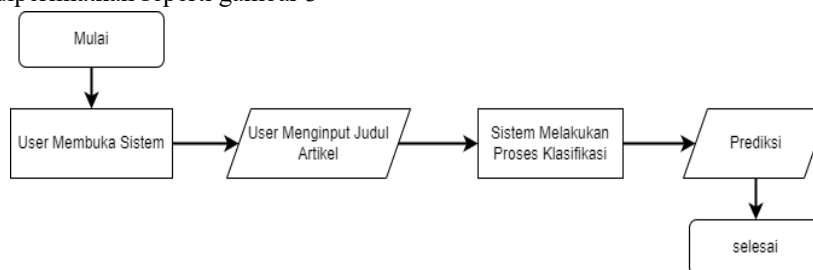
- d. F1 Score merupakan perbandingan rata-rata dari presisi dan recall.

$$F1 - score = 2x \frac{precision \times recall}{(Precision + Recall)} \quad (6)$$

Setelah hasil klasifikasi diperoleh, selanjutnya hasil dievaluasi menggunakan Cross Validation (confusion matrix) untuk melihat akurasi, presisi dan recall yang dihasilkan oleh model yang diusulkan.

2.5 Flowchart Sistem

Flowchart sistem dalam memprediksi judul publikasi dari lima bidang ilmu untuk menentukan akurasi persentase tertinggi yaitu dengan kesesuaian judul artikel dalam SINTA yang sudah terindeks oleh Web Of Science dalam 5 bidang ilmu tersebut diperlihatkan seperti gambar 3



Gambar 3 Flowchart sistem

Flowchart sistem dalam gambar 3 memprediksi judul publikasi dari lima bidang ilmu untuk menentukan akurasi persentase tertinggi yaitu dengan kesesuaian judul artikel dalam SINTA yang sudah terindeks oleh Web Of Science dalam 5 bidang ilmu tersebut yaitu :

1. User membuka sistem.
2. User Memasukkan Judul.

3. Setelah user memasukkan judul, tekan tombol prediksi untuk menyelesaikan Proses Klasifikasi, dimana sistem mengklasifikasikan judul dengan menggunakan metode Naïve Bayes
4. Menampilkan hasil prediksi judul publikasi dari lima bidang ilmu tersebut, maka alur proses sistem ini telah selesai.

3. HASIL DAN ANALISA

3.1 Implementasi *User Interface* (UI)

Terdapat halaman predict pada gambar 4, merupakan halaman yang 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.

a. Tampilan Halaman Prediksi

Halaman predict

Masukkan Judul

Naive Bayes Classification

Predict

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 Hasil Prediksi

Pada gambar 4 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.

3.2 Pengujian Sistem

Tahapan pengujian sistem dalam penelitian ini yaitu dengan memakai black box testing. Black box testing merupakan metode pengujian software dengan tujuan menguji 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 fitur yang terdapat pada sistem, hasil uji fungsional sistem ini terdapat dalam tabel 9

Tabel 9 Pengujian sistem

Skenario Pengujian	Kasus Pengujian	Hasil Pengujian	Kesimpulan
<i>Tampil dataset</i>	Halaman menampilkan <i>dataset</i>	Sesuai	Normal
<i>Input judul</i>	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 9 merupakan hasil 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.

3.3 Analisa Hasil Validasi Penerapan Algoritma *Naive Bayes*

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 10 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 10 Rincian data

Nama Data	Jumlah (%)	Jumlah angka
Data Training	90%	4500
Data Testing	10%	500
Total Data : 5000		

Dari tabel 10 menunjukkan rincian data sampel dan data testing yang terbagi ke dalam 5 bidang ilmu.

Tabel 11 Rincian data dari lima kategori bidang ilmu

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

Dari tabel 11, data tersebut nantinya akan menghasilkan akurasi berupa confusion matrix, dari confusion matrix ini akan digunakan untuk mencari nilai akurasi, presisi, recall dan F1 score. Hasil training yang telah di uji coba menghasilkan confusion matrix dengan nilai accuracy, precision, recall dan F1 score terdapat pada tabel 12 dan gambar 5.

Tabel 12 Hasil Accuracy, Recall, Precision, F1 score

Accuracy	Precision	Recall	F1 score
0.44	0.24	0.20	0.07

```
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 5 Output hasil klasifikasi 5000 data

Dalam gambar 5 menunjukkan hasil pengujian dengan sampel data 5000 judul artikel menghasilkan output beberapa nilai antara lain yaitu nilai akurasi sebanyak 0.442 dengan presisi, recall, F1 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 13 dan gambar 6 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 13 Hasil Confusion Matrix

	Positive	Negative
True	101	404
False	399	399

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

Gambar 6 Output hasil Confusion Matrix

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%).

3.4 Pembahasan 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 :

Adanya judul yang membahas tentang topik yang sama akan tetapi terbagi pada bidang ilmu yang berbeda. Seperti contoh pada tabel 14 Topik tentang “Business” yang seharusnya lebih mengacu ke arah bidang (Social Science and Management).

Tabel 14 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	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	Life Sciences & Medicine
4	- Automatic diagnosis of alcohol use disorder using EEG features	Social Sciences and Management

4 KESIMPULAN

Dari penelitian ini dapat diambil kesimpulan bahwa 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 pada “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.

ACKNOWLEDGEMENTS

Tugas akhir ini merupakan bagian dari penelitian dosen Teknik Informatika Universitas Islam Sultan Agung berjudul Pengembangan Sistem Klasifikasi Otomatis Publikasi Ilmiah pada database SINTA Science and Technology Index yang dibiayai dari Kementerian Pendidikan dan Kebudayaan, Pendidikan Tinggi, Riset dan Teknologi Republik Indonesia tahun 2022.

DAFTAR PUSTAKA

- [1] P. Bimo, N. Setio, D. Retno, S. Saputro, dan B. Winarno, “Klasifikasi dengan Pohon Keputusan Berbasis Algoritme,” vol. 3, hal. 64–71, 2020.
- [2] K. Li, J. Rollins, dan E. Yan, “Web of Science use in published research and review papers 1997–2017: a selective, dynamic, cross-domain, content-based analysis,” *Scientometrics*, vol. 115, no. 1, hal. 1–20, 2018, doi: 10.1007/s11192-017-2622-5.
- [3] R. N. Devita, H. W. Herwanto, dan A. P. Wibawa, “Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, hal. 427, 2018, doi: 10.25126/jtiik.201854773.
- [4] N. Nurdin, M. Suhendri, Y. Afrilia, dan R. Rizal, “Klasifikasi Karya Ilmiah (Tugas Akhir) Mahasiswa Menggunakan Metode Naive Bayes Classifier (NBC),” *Sistemasi*, vol. 10, no. 2, hal. 268, 2021, doi: 10.32520/stmsi.v10i2.1193.
- [5] A. Fitria dan H. Azis, “Analisis Kinerja Sistem Klasifikasi Skripsi menggunakan Metode Naïve Bayes Classifier,” *Pros. Semin. Nas. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 2, hal. 102–106, 2018.
- [6] D. I. D. Merangin *dkk.*, “Analisis struktur ovarium indikator terkait kesehatan pada orang tua di rumah dengan fokus pada persepsi kesehatan subjektif,” *J. Stud. Lingkung. Asyut*, vol. 2, no. 2, hal. 2016, 2018, [Daring]. Tersedia pada: <https://dspace.uin.ac.id/handle/123456789/7761>
- [7] A. T. Hardianti, A. R. Manga, dan H. Darwis, “Penerapan metode Naive Bayes pada klasifikasi judul jurnal,” *Pros. Semin. Nas. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 2, hal. 97, 2018, [Daring]. Tersedia pada: <http://e-journals.unmul.ac.id/index.php/SAKTI/article/view/1838/pdf>
- [8] R. Nuraeni, A. Sudiarjo, dan R. Rizal, “Perbandingan Algoritma Naïve Bayes Classifier dan Algoritma Decision Tree

- untuk Analisa Sistem Klasifikasi Judul Skripsi,” *Innov. Res. Informatics*, vol. 3, no. 1, hal. 26–31, 2021, doi: 10.37058/innovatics.v3i1.2976.
- [9] S. Widaningsih, U. Suryakencana, A. Suheri, dan U. Suryakencana, “KLASIFIKASI JURNAL ILMU KOMPUTER BERDASARKAN PEMBAGIAN WEB OF,” no. March, 2018.
- [10] H. Lubis, “Implementasi K-NN Dalam Analisa Sentimen Riba Pada Bunga Bank Berdasarkan Data Twitter,” vol. 4, no. April, hal. 369–376, 2020, doi: 10.30865/mib.v4i2.2051.
- [11] F. Rahutomo, I. Y. R. Pratiwi, dan D. M. Ramadhani, “Eksperimen Naïve Bayes Pada Deteksi Berita Hoax Berbahasa Indonesia,” *J. Penelit. Komun. Dan Opini Publik*, vol. 23, no. 1, 2019, doi: 10.33299/jpkop.23.1.1805.
- [12] A. D. Wibisono, S. Dadi Rizkiono, dan A. Wantoro, “FILTERING SPAM EMAIL MENGGUNAKAN METODE NAIVE BAYES,” *TELEFORTECH J. Telemat. Inf. Technol.*, vol. 1, no. 1, hal. 9–17, 2020, doi: 10.33365/tft.v1i1.685.
- [13] J. S. Dwi Normawati, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” vol. 5, no. September, hal. 697–711, 2021.