

**Klasifikasi Kepakaran *Reviewer* Menggunakan Algoritma
K-Nearest Neighbor (KNN)**

LAPORAN TUGAS AKHIR

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



DISUSUN OLEH :

INTAN PUTRI NUR FAJARINI

NIM 32601800013

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNOLOGI INDUSTRI

UNIVERSITAS ISLAM SULTAN AGUNG

SEMARANG

DESEMBER 2022

FINAL PROJECT

***Classification of Reviewer Expertise Using Algorithms K-Nearest
Neighbor (KNN)***

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



Arranged By :

INTAN PUTRI NUR FAJARINI

NIM 32601800013

MAJORING OF INFORMATICS ENGINEERING

INDUSTRIAL TECHNOLOGY FACULTY

SULTAN AGUNG ISLAMIC UNIVERSITY

SEMARANG

DECEMBER 2022

LEMBAR PENGESAHAN PEMBIMBING

Laporan Tugas Akhir dengan judul “Klasifikasi Kepakaran *Reviewer* Menggunakan Algoritma *K-Nearest Neighbor* (KNN)” ini disusun oleh :

Nama : Intan Putri Nur Fajarini

NIM : 32601800013

Program Studi : Teknik Informatika

Telah disahkan oleh dosen pembimbing pada :

Hari : Kamis

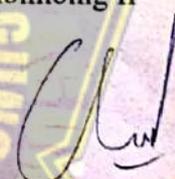
Tanggal : 29 Desember 2022

Mengesahkan,

Pembimbing I

Pembimbing II


Imam Much Ibnu Subroto, ST., M.Sc., Ph.D


Andi Riansyah, ST, M.Kom

NIDN. 0613037301

NIDN. 0609108802

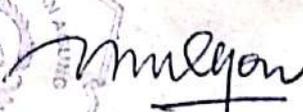
Mengetahui,

Ketua Program Studi Teknik Informatika

Fakultas Teknologi Industri

Universitas Islam Sultan Agung




Ir. Sri Mulyono, M.Eng

NIDN. 062606661

LEMBAR PENGESAHAN PENGUJI

Laporan tugas akhir dengan judul "**Klasifikasi Kepakaran *Reviewer* Menggunakan Algoritma *K-Nearest Neighbor (KNN)***" ini telah dipertahankan di depan dosen penguji Tugas Akhir pada :

Hari : Kamis.....

Tanggal : 29 Desember 2022.....

TIM PENGUJI

Penguji I

Badie'ah, ST, M.Kom

NIDN.0619018701

Penguji II

Ghufron, ST, M.Kom

NIDN.0602079005



SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Intan Putri Nur Fajarini

NIM : 32601800013

Judul Tugas Akhir : Klasifikasi Kepakaran *Reviewer* Menggunakan Algoritma
K-Nearest Neighbor (KNN)

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, 09 Januari 2023

Yang Menyatakan,



Intan Putri Nur Fajarini

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Intan Putri Nur Fajarini

NIM : 32601800013

Program Studi : Teknik Informatika

Fakultas : Teknologi industri

Alamat Asal : LK. Kauman RT 02 RW 07 Kec. Wirosari Kab. Grobogan

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul : Klasifikasi Kepakaran *Reviewer* Menggunakan Algoritma *K-Nearest Neighbor* (KNN).

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

Semarang, 09 Januari 2023

Yang menyatakan,


Intan Putri Nur Fajarini

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 Kepakaran *Reviewer* Menggunakan Algoritma *K-Nearest Neighbor* (KNN)” 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 Much Ibnu Subroto, ST., M.Sc., Ph.D yang telah meluangkan waktu dan memberi ilmu.
4. Dosen pembimbing II penulis Andi Riansyah, 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 sebutkan 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, 09 Januari 2023



Intan Putri Nur Fajarini

DAFTAR ISI

LAPORAN TUGAS AKHIR.....	i
<i>FINAL PROJECT</i>.....	ii
LEMBAR PENGESAHAN PEMBIMBING	iii
LEMBAR PENGESAHAN PENGUJI.....	iv
SURAT PERNYATAAN KEASLIAN TUGAS AKHIR.....	v
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xiii
ABSTRAK	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	3
1.3 Pembatasan Masalah	3
1.4 Tujuan	3
1.5 Manfaat	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI.....	6
2.1 Tinjauan Pustaka	6
2.2 Dasar Teori.....	11
2.2.1 Data <i>mining</i>	11
2.2.2 <i>Preprocessing</i>	11

2.2.3	Klasifikasi	12
2.2.4	<i>K-Nearest Neighbor (KNN)</i>	12
2.2.5	Evaluasi	14
2.2.6	Rubix ML.....	15
BAB III METODE PENELITIAN		16
3. 1	Metode Penelitian.....	16
a.	Dataset yang digunakan	17
b.	Studi kasus	18
c.	Koleksi data.....	18
d.	<i>Preprocessing</i>	19
e.	Penerapan metode	21
f.	Evaluasi.....	21
3. 2	Metode Perancangan Sistem	22
3.2.1	Analisis Kebutuhan	22
a.	Menampilkan data nama-nama <i>reviewer</i>	22
b.	Menampilkan data judul artikel <i>reviewer</i>	22
c.	Melakukan klasifikasi judul	22
d.	Menampilkan hasil akurasi	22
3.2.2	Alur Penggunaan Sistem	23
3. 3	Desain <i>Database</i>	24
3. 4	Perancangan <i>User Interface</i>	28
a.	Menu <i>Dashboard</i>	28
b.	Menu <i>Home</i>	28
c.	Menu Dataset	29
d.	Menu <i>Predict</i>	30

e.	Halaman <i>reviewer</i> yang menampilkan daftar judul artikel	30
f.	Halaman <i>reviewer</i> yang menampilkan label kepakaran.....	31
BAB IV HASIL DAN ANALISIS PENELITIAN		32
4.1	Analisis Penerapan Metode Penelitian.....	32
4.1.1	Dataset.....	32
4.1.2	<i>Preprocessing</i>	33
4.1.3	Penerapan metode KNN.....	34
4.2	Hasil Implementasi <i>User Interface</i>	35
4.3	Analisis Hasil Evaluasi	40
4.4	Perbandingan Performa Data <i>Training</i> dengan Data <i>Testing</i>	46
4.5	Hasil Pengujian Sistem	49
4.6	Pembahasan Faktor Hasil Akurasi	50
BAB V KESIMPULAN DAN SARAN		51
5.1	Kesimpulan	51
5.2	Saran.....	51
LAMPIRAN.....		55

DAFTAR TABEL

Tabel 2. 1 Model <i>Confusion matrix</i>	14
Tabel 3. 1 Contoh data	18
Tabel 3. 2 Proses Data <i>cleaning</i>	19
Tabel 3. 3 Proses <i>Case Folding</i>	19
Tabel 3. 4 Proses <i>Tokenizing</i>	19
Tabel 3. 5 Proses <i>Stopword</i>	20
Tabel 3. 6 Proses <i>Stemming</i>	21
Tabel 3. 7 Desain tabel nama	25
Tabel 3. 8 Desain tabel <i>article</i>	25
Tabel 3. 9 Desain tabel <i>article_author</i>	26
Tabel 3. 10 Desain tabel scopus.....	27
Tabel 4. 1 Tabel skenario data uji (Hendriyanto & Sari, 2022).....	40
Tabel 4. 2 Tabel <i>confusion matrix</i> K = 3 menggunakan split data 70% : 30%....	41
Tabel 4. 3 Tabel <i>confusion matrix</i> K = 5 menggunakan split data 70% : 30%....	41
Tabel 4. 4 Tabel <i>confusion matrix</i> K = 9 menggunakan split data 70% : 30%....	41
Tabel 4. 5 Tabel <i>confusion matrix</i> K = 20 menggunakan split data 70% : 30%..	41
Tabel 4. 6 Tabel <i>confusion matrix</i> K = 25 menggunakan split data 70% : 30%..	41
Tabel 4. 7 Tabel <i>confusion matrix</i> K = 30 menggunakan split data 70% : 30%..	42
Tabel 4. 8 Tabel <i>confusion matrix</i> K = 35 menggunakan split data 70% : 30%..	42
Tabel 4. 9 Tabel <i>confusion matrix</i> K = 3 menggunakan split data 80% : 20%....	42
Tabel 4. 10 Tabel <i>confusion matrix</i> K = 5 menggunakan split data 80% : 20%..	42
Tabel 4. 11 Tabel <i>confusion matrix</i> K = 9 menggunakan split data 80% : 20%..	42
Tabel 4. 12 Tabel <i>confusion matrix</i> K = 20 menggunakan split data 80% : 20%	43
Tabel 4. 13 Tabel <i>confusion matrix</i> K = 25 menggunakan split data 80% : 20%	43
Tabel 4. 14 Tabel <i>confusion matrix</i> K = 30 menggunakan split data 80% : 20%	43
Tabel 4. 15 Tabel <i>confusion matrix</i> K = 35 menggunakan split data 80% : 20%	43
Tabel 4. 16 Tabel <i>confusion matrix</i> K = 3 menggunakan split data 90% : 10%..	43
Tabel 4. 17 Tabel <i>confusion matrix</i> K = 5 menggunakan split data 90% : 10%..	44
Tabel 4. 18 Tabel <i>confusion matrix</i> K = 9 menggunakan split data 90% : 10%..	44

Tabel 4. 19 Tabel <i>confusion matrix</i> K = 20 menggunakan split data 90% : 10%	44
Tabel 4. 20 Tabel <i>confusion matrix</i> K = 25 menggunakan split data 90% : 10%	44
Tabel 4. 21 Tabel <i>confusion matrix</i> K = 30 menggunakan split data 90% : 10%	44
Tabel 4. 22 Tabel <i>confusion matrix</i> K = 35 menggunakan split data 90% : 10%	45
Tabel 4. 23 Tabel <i>Accuracy, Precision, Recall, F-Measure</i> menggunakan split data 70% : 30%	45
Tabel 4. 24 Tabel <i>Accuracy, Precision, Recall, F-Measure</i> menggunakan split data 80% : 20%	45
Tabel 4. 25 Tabel <i>Accuracy, Precision, Recall, F-Measure</i> menggunakan split data 90% : 10%	46
Tabel 4. 26 Perbandingan performa <i>training</i> dan <i>testing</i> dengan split data 70% : 30%	47
Tabel 4. 27 Perbandingan performa <i>training</i> dan <i>testing</i> dengan split data 80% : 20%	47
Tabel 4. 28 Perbandingan performa <i>training</i> dan <i>testing</i> dengan split data 90% : 10%	48
Tabel 4. 29 Tabel hasil pengujian sistem	49



DAFTAR GAMBAR

Gambar 3. 1 Tahap-tahap Penelitian	16
Gambar 3. 2 <i>Flowchart</i> tahap <i>Preprocessing</i>	17
Gambar 3. 3 <i>Flowchart</i> Sistem	23
Gambar 3. 4 <i>Database relation design</i>	24
Gambar 3. 5 Rancangan menu <i>Dashboard</i>	28
Gambar 3. 6 Rancangan menu <i>home</i>	28
Gambar 3. 7 Rancangan menu dataset	29
Gambar 3. 8 Rancangan menu <i>predict</i>	30
Gambar 3. 9 Rancangan halaman <i>reviewer</i> dan judul artikel	30
Gambar 3. 10 Rancangan halaman <i>reviewer</i> dan kepakaran	31
Gambar 4. 1 Code program untuk membaca dataset	33
Gambar 4. 2 Code program untuk <i>preprocessing</i> data	33
Gambar 4. 3 Code program untuk <i>preprocessing</i> data fungsi <i>stemmer</i>	34
Gambar 4. 4 Code program untuk membaca daftar <i>stopword</i>	34
Gambar 4. 5 Code program untuk metode KNN	34
Gambar 4. 6 Tampilan menu pada sistem klasifikasi	35
Gambar 4. 7 Tampilan halaman <i>Dashboard</i>	35
Gambar 4. 8 Tampilan halaman <i>home</i>	36
Gambar 4. 9 Tampilan halaman dataset.....	37
Gambar 4. 10 Tampilan halaman <i>predict</i>	37
Gambar 4. 11 Tampilan halaman <i>reviewer</i> dan judul artikel (1)	38
Gambar 4. 12 Tampilan halaman <i>reviewer</i> dan judul artikel (2)	38
Gambar 4. 13 Tampilan halaman <i>reviewer</i> dan label kepakaran (1)	39
Gambar 4. 14 Tampilan halaman <i>reviewer</i> dan label kepakaran (2)	40
Gambar 4. 15 Contoh data dengan kategori yang belum sesuai	50

ABSTRAK

Ada banyak hal yang bisa dilakukan dalam meningkatkan kualitas pendidikan di negara Indonesia salah satunya yaitu Kemdikbud membuat sebuah program untuk meningkatkan jumlah publikasi jurnal di Indonesia yang bereputasi internasional. Jika dilakukan upaya peningkatan publikasi jurnal maka diperlukan *reviewer* jurnal, yang mana bidang keilmuan *reviewer* harus sesuai dengan bidang ilmu dari jurnal yang akan direview, data *reviewer* ini diperoleh dari database SINTA. Bidang ilmu dari jurnal tersebut dikelompokkan ke dalam 5 bidang yaitu *Arts & Humanities, Engineering & Technology, Life Sciences & Medicine, Natural Science, dan Social Science & Management*. Saat ini belum ada penelitian yang menerapkan suatu metode untuk mengklasifikasikan *reviewer* dengan 5 bidang ilmu kepakaran, untuk memudahkan hal tersebut bisa memanfaatkan teknik yang ada saat ini yaitu menggunakan metode data *mining* untuk mengolah data tersebut, dalam penerapan data *mining* untuk klasifikasi akan digunakan algoritma KNN. Dalam penelitian ini tahap yang dilakukan adalah pengumpulan data, kemudian data tersebut akan diolah menggunakan metode *preprocessing* agar diperoleh data yang bersih dan memiliki kualitas yang baik, setelah itu akan diproses lagi dengan menerapkan metode klasifikasi menggunakan algoritma KNN. Setelah melakukan beberapa pengujian jumlah K dan menggunakan 3 skenario pengujian, pada penelitian ini menghasilkan akurasi 0,751 atau 75 %, *Precision* 0.628, *recall* 0.54, dan *f-measure* 0.528 dengan K = 30 pada pengujian skenario uji 80% : 20%.

Kata kunci : *reviewer*, data *mining*, klasifikasi, KNN

There are many things that can be done to improve the quality of education in Indonesia, one of which is that the Ministry of Education and Culture has created a program to increase the number of Indonesian journal publications that have an international reputation. If an effort is made to increase journal publications, a journal reviewer is needed, whose scientific field must be in accordance with the field of study of the journal to be reviewed. The reviewer's data is obtained from the SINTA database. The scientific fields of the journal are grouped into 5 fields: pers & humanities, engineering & technology, life sciences & medicine, natural science, and social science & management. Currently, there is no research that applies a method to classify reviewers with 5 areas of expertise. To make it easier for this to take advantage of existing techniques, namely using the data mining method to process the data, the KNN algorithm will be used. In this research, the data collection stage is carried out, then the data will be processed using the preprocessing method in order to obtain clean and good-quality data, after which it will be processed again by applying the classification method using the KNN algorithm. After testing the number K and using 3 test scenarios, this study produced an accuracy of 0.751 or 75%, a precision of 0.628, a recall of 0.54, and an f-measure of 0.528 with K = 30 in a test scenario of 80%: 20%.

Keywords: reviewer, data mining, classification, KNN.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kementerian Pendidikan dan Kebudayaan Republik Indonesia (Kemdikbud) merupakan kementerian pemerintah negara Indonesia yang mengurus segala urusan yang ada di bidang pendidikan, meliputi pendidikan anak usia dini, pendidikan dasar, pendidikan menengah, pembinaan masyarakat, dan budaya. Dalam hal ini Kemdikbud memiliki dan mengelola banyak data terkait pendidikan dan disiplin ilmu. Data-data tersebut dapat digunakan untuk meningkatkan kualitas pendidikan di Indonesia, jika diolah dengan baik dan benar maka akan menghasilkan sebuah informasi yang dapat berguna bagi kepentingan pendidikan.

Saat ini, Kemdikbud membuat sebuah program untuk meningkatkan jumlah publikasi jurnal di Indonesia yang bereputasi internasional. Untuk itu, diperlukan tenaga *reviewer* jurnal sesuai dengan bidang ilmu para peneliti dan dosen yang akan dipublikasi. Yang mana jumlah *reviewer* jurnal di Indonesia sangat banyak yaitu lebih dari 100 ribu dengan bidang ilmu yang berbeda-beda dari sejumlah 14.086 jurnal yang terindeks di Garuda Kemdikbud, dalam pembagiannya bidang ilmu yang berbeda-beda tersebut dikelompokkan menjadi 5 bidang keilmuan yaitu *Arts & Humanities, Engineering & Technology, Life Sciences & Medicine, Natural Sciences, dan Social Sciences & Management*. Oleh karena itu, diperlukan suatu metode untuk mengklasifikasikan secara otomatis sehingga bidang ilmu kepakaran dengan *reviewer* dapat sesuai. Selain itu belum adanya sebuah kebijakan untuk mengatur metode apa yang dapat digunakan untuk melakukan klasifikasi terhadap data *reviewer* tersebut, sehingga dari data-data yang ada masih sulit untuk dikelompokkan ke dalam 5 bidang ilmu.

Seiring berjalannya waktu, laju perkembangan teknologi semakin pesat, berbagai data dengan jumlah yang banyak dapat diolah menggunakan proses yang dikenal sebagai data *mining*. Data *mining* merupakan proses menambang informasi penting yang diperoleh dari sekumpulan data. Informasi yang penting tersebut diperoleh melalui proses yang sangat kompleks seperti penggunaan kecerdasan buatan, teknik statistik, dan lain-lain. Teknik yang kompleks ini mengidentifikasi

dan mengekstraksi informasi yang berguna dari *database* besar (Sudarsono et al., 2021). Ada beberapa metode di dalam data mining salah satunya adalah klasifikasi. Klasifikasi adalah proses mengklasifikasikan fitur atau kata yang mirip ke dalam sebuah kelas menggunakan data latih (data *training*) sebagai acuan pengelompokan dan data uji (data *testing*) sebagai pengujian. Hal tersebut bertujuan agar objek kelas yang belum memiliki label dan terklasifikasi dapat diprediksi (Fitrianti et al., 2019). Ada beberapa algoritma yang dapat digunakan dalam metode klasifikasi salah satunya adalah KNN, dimana objek metode diklasifikasikan berdasarkan data pembelajaran dengan cara menghitung kedekatan jarak satu data terhadap data yang lainnya, yang mana KNN ini mempunyai beberapa keunggulan diantaranya memiliki kemampuan untuk melakukan pelatihan dengan sangat cepat, termasuk algoritma yang sederhana dan mudah untuk dipelajari, tahan pada data yang memiliki derau, dan efektif pada data pelatihan yang besar. Namun algoritma KNN ini juga mempunyai kelemahan diantaranya jumlah nilai k dapat menjadi bias, komputasinya kompleks, memori terbatas, serta mudah dikelabui dengan fitur yang tidak relevan (Supriatman, 2021).

Adapun penelitian sebelumnya yang membahas mengenai klasifikasi laporan masyarakat menggunakan algoritma KNN, dalam penelitian tersebut menyebutkan bahwa algoritma KNN menghasilkan akurasi yang baik dalam mengklasifikasikan data laporan masyarakat ke dalam tiga kategori yaitu pengaduan, permintaan informasi, dan aspirasi. Data yang digunakan adalah 2235 data yang telah dikumpulkan oleh pemerintah kota Semarang selama tahun 2017. Setelah dilakukan pengolahan data diperoleh 930 data latih, masing-masing 310 data latih untuk setiap kelas laporan masyarakat dan 100 data uji, dengan uji evaluasi dan validasi menggunakan *confusion matrix* diketahui bahwa pada nilai $k = 11$ memperoleh akurasi tertinggi yaitu 82% dalam penentuan kategori kelas laporan masyarakat (Hadi & Sukamto, 2020).

Dari permasalahan tersebut penelitian ini mengusulkan metode KNN untuk klasifikasi kepakaran *reviewer*. Pada penelitian ini menggunakan data judul publikasi jurnal terindeks scopus sebagai data *training* dan data judul publikasi jurnal *reviewer* sebagai data *testing*. Sehingga bidang ilmu kepakaran dengan

reviewer dapat sesuai. Dalam proses klasifikasi data *reviewer* ini akan menggunakan algoritma KNN dengan metode perhitungan jarak *euclidean distance*.

1.2 Perumusan Masalah

Adapun rumusan masalah dalam penelitian ini diantaranya yaitu :

- a. Belum adanya penelitian untuk mengklasifikasikan bidang ilmu kepakaran dengan *reviewer* terhadap data *reviewer*.
- b. Bagaimana mengklasifikasikan bidang ilmu kepakaran dengan data *reviewer* menggunakan algoritma KNN.

1.3 Pembatasan Masalah

Adapun pembatasan masalah dalam penelitian ini diantaranya yaitu:

- a. Melakukan klasifikasi bidang ilmu kepakaran dengan *reviewer* terhadap data *reviewer* yang diambil dari *database* SINTA.
- b. Klasifikasi kepakaran pada penelitian ini hanya dibatasi pada 5 bidang kepakaran yaitu *Arts & Humanities, Engineering & Technology, Life Sciences & Medicine, Natural Sciences, dan Social Sciences & Management*.
- c. Data kepakaran *reviewer* yang diolah dalam penelitian ini diambil berdasarkan bidang kepakaran atau bidang ilmu yang ditulis sendiri oleh para *reviewer* pada jurnal yang terkait.

1.4 Tujuan

Adapun tujuan dari dilakukannya penelitian ini adalah untuk mengklasifikasikan bidang ilmu kepakaran dengan *reviewer* terhadap data *reviewer* dari *database* SINTA ke dalam 5 bidang ilmu menggunakan metode klasifikasi data *mining* dengan algoritma KNN.

1.5 Manfaat

Adapun manfaat dari dilakukannya penelitian ini adalah :

- a. Mengklasifikasikan kepakaran *reviewer* ke dalam 5 bidang ilmu kepakaran secara otomatis.
- b. Memudahkan pencarian pakar pada bidang tertentu untuk tujuan penelitian baik dilevel perorangan, perguruan tinggi maupun secara nasional.

1.6 Sistematika Penulisan

Sistematika penulisan laporan yang digunakan dalam menyusun laporan tugas akhir ini adalah sebagai berikut :

BAB I : PENDAHULUAN

Pada bab ini berisi tentang latar belakang penelitian dalam pembuatan tugas akhir, perumusan masalah, pembatasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan untuk laporan tugas akhir.

BAB II : TINJAUAN PUSTAKA DAN DASAR TEORI

Pada bab ini berisi tentang tinjauan pustaka yang memuat penelitian-penelitian yang pernah dilakukan sebelumnya terkait judul tugas akhir yang diambil, yang juga memuat penjelasan tentang konsep dan prinsip dasar yang diperlukan untuk memecahkan masalah yang menjadi topik tugas akhir berbagai referensi yang dijadikan landasan pada kegiatan penelitian yang sedang dilakukan. Memuat dasar teori yang berisi penjelasan tentang teori-teori yang digunakan dalam penelitian.

BAB III : METODE PENELITIAN

Pada bab ini menguraikan secara rinci metode penelitian yang digunakan yaitu penggunaan metode KNN untuk mengklasifikasikan kepakaran *reviewer*, yang mana metode penelitian tersebut dipakai sebagai pendekatan untuk mendapatkan solusi.

BAB IV : HASIL DAN ANALISIS PENELITIAN

Pada bab ini memuat hasil penelitian yang telah dilakukan yaitu tingkat akurasi, presisi, dan *recall* terhadap kesesuaian 5 bidang ilmu dengan data *reviewer*.

BAB V : KESIMPULAN DAN SARAN

Pada bab ini memuat kesimpulan dari seluruh uraian penelitian yang dilakukan dari awal hingga selesai, serta berisi saran dari hasil penelitian dan diharapkan dapat bermanfaat dalam pengembangan selanjutnya.



BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Berikut ini adalah beberapa penelitian sebelumnya yang dijadikan sebagai sumber referensi dan berkaitan dengan judul penelitian tugas akhir ini :

Pada penelitian sebelumnya membahas tentang implementasi dari algoritma KNN dalam menganalisis sentimen *review* sebuah restoran menggunakan teks berbahasa Indonesia. Penelitian tersebut melakukan perhitungan akurasi menggunakan *confusion matrix*, visualisasi *wordcloud*, histogram serta tampilan dengan package shiny R Studio. Data yang digunakan sebanyak 205 data *review* pelanggan restoran dan setelah dilakukan *preprocessing* berkurang menjadi 199 data disebabkan data mengalami reduksi atau pengurangan karena memiliki noise atau beberapa kata dalam bahasa Indonesia yang tidak tepat. Menggunakan 70% data latih yang terdiri dari 140 data dan data uji 30% yaitu 59 data dengan membagi data latih dan data uji secara acak, dari pengujian dengan *confusion matrix* diperoleh hasil akurasi 96,61%, dan jumlah nilai $k = 1$ serta *error rate* 3,39% (Fitrianti et al., 2019).

Penelitian berikutnya mengenai penerapan algoritma KNN untuk mengklasifikasikan dokumen tugas akhir teknik informatika berdasarkan data abstrak penelitian tugas akhir. Dalam penelitian ini menggunakan 504 data judul, 454 data abstrak berbahasa Inggris, dan 50 data abstrak berbahasa Indonesia. Dalam penelitian ini pengklasifikasian data intisari berbahasa Indonesia memperoleh hasil *accuracy* lebih besar tanpa melalui tahap *stemming*, dengan perbandingan data *training* dan data *testing* 9 : 1 menggunakan pembagian data *K-fold cross validation* 80,0% sehingga memperoleh nilai akurasi 100,0% (Akromunnisa & Hidayat, 2019).

Penelitian berikutnya mengenai penggunaan algoritma KNN dalam mengklasifikasikan calon mahasiswa bidikmisi. Dalam penelitian menggunakan data sebanyak 358 data, untuk menentukan atribut yang akan digunakan pada penelitian ini dengan cara melakukan wawancara terhadap pihak terkait yang bertugas menjadi tim penerima mahasiswa bidikmisi, dari wawancara tersebut

diperoleh atribut diantaranya nilai ujian, pendidikan orang tua, profesi orang tua, penghasilan orang tua, dan jumlah saudara. Dalam penelitian ini melakukan beberapa pengujian nilai $k = 1$ sampai dengan $k = 10$, dan diperoleh hasil tertinggi dengan $k = 1$ berupa akurasi sebesar 97,2067%, presisi 0,972, dan *recall* 0,972 (Haerul Jaman, Jajam; Fahlevi, 2019).

Penelitian berikutnya mengenai penerapan algoritma KNN dalam mengklasifikasikan penentuan pengajuan kartu kredit, penelitian ini bertujuan untuk mempermudah pihak bank dalam penentuan kategori kartu kredit yang sesuai bagi nasabah, yang mana terdapat 4 kategori kartu kredit yaitu tidak diterima, *classic*, *gold*, dan platinum. Data yang digunakan adalah data nasabah Bank BNI Syariah Surabaya sebanyak 250 data, dengan rincian 250 data latih dan 100 data uji dengan mengambil sampel dari data latih yang diperoleh sebelumnya. Dalam penelitian ini menggunakan beberapa atribut untuk faktor pendukung diantaranya jenis kelamin, status rumah, status, jumlah tanggungan berupa anak, pekerjaan, serta penghasilan tiap tahun. Dalam penelitian ini diperoleh hasil nilai akurasi dengan rata-rata 93%, presisi 92%, dan *recall* 83% (Kurniawan & Barokah, 2020).

Penerapan berikutnya mengenai sebuah analisa penerapan algoritma KNN dalam menentukan nilai akurasi terhadap kepuasan pelanggan yang mana merupakan studi kasus pada PT. Trigatra Komunikatama. Dalam penelitian ini menggunakan data kuisisioner yang diberikan kepada *customer* PT. Trigatra Komunikatama, sehingga diperoleh data dari 1000 responden dan dilakukan reduksi dengan meminimalkan adanya duplikasi data sehingga menjadi 150 *record* sebagai data *training*. Penelitian ini memperoleh hasil akurasi cukup baik yaitu 83,33% (Fansyuri, 2020).

Penelitian berikutnya mengenai analisis judul skripsi untuk menentukan peminatan mahasiswa jurusan teknik informatika UNJANI dengan membandingkan penggunaan *Vector Space Model* dan metode KNN. Pada jurusan teknik informatika UNJANI memiliki 2 peminatan yaitu Sistem Cerdas Data Mining (SCDM) dan Sistem Informasi *Enterprise* (SIE), sesuai dengan kebijakan UNJANI judul skripsi yang dikerjakan oleh setiap mahasiswa harus berdasarkan dengan topik serta peminatan yang diambil oleh mahasiswa, hal ini membuat para

pengambil kebijakan kesulitan sehingga judul, topik, dan peminatan yang dipilih mahasiswa sering kali tidak sesuai. Penelitian ini bertujuan untuk mempermudah pihak jurusan dengan membuat sebuah pengambil keputusan terhadap judul skripsi yang diajukan oleh setiap mahasiswa agar dapat sesuai dengan peminatan. Pada penelitian ini menerapkan metode *5-fold cross validation*, dengan menggunakan 400 data judul skripsi sebagai data *training* dan data *testing*. Sehingga pada penelitian ini penggunaan algoritma KNN memiliki akurasi lebih baik dan cukup efektif dalam melakukan prediksi dan klasifikasi data judul skripsi dengan atribut topik yaitu menghasilkan akurasi 96,85% (Atmaja & Mandala, 2020).

Penelitian berikutnya mengenai klasifikasi data produk baru pada toko mebel Aneka Jaya menggunakan metode KNN untuk mengklasifikasikan data produk baru dengan kriteria yang digunakan yaitu nilai tinggi, nilai panjang, nilai lebar, dan nilai harga. Subjek dalam penelitian ini menggunakan 188 data yang merupakan data *training* dan data *testing* pada tahun 2019-2021. berdasarkan pengujian metode KNN dengan *confusion matrix* memperoleh akurasi 90% dengan perbandingan data prediksi yang digunakan adalah data tahun 2019 dengan data *real* yang digunakan adalah data tahun 2020. Kemudian memperoleh hasil akurasi sebesar 80% dengan perbandingan data prediksi yang digunakan adalah data tahun pada 2020 dengan data *real* yang digunakan adalah data pada tahun 2021 (Maulana et al., 2021).

Penelitian berikutnya yaitu mengenai klasifikasi untuk pemberian sanksi pajak menggunakan metode KNN di KPP Pratama Cirebon Dua yang mana dapat digunakan untuk memudahkan dalam mengetahui perbedaan denda wajib pajak, penelitian tersebut menggunakan 229 data. Dari hasil pengujian akhir penerapan algoritma KNN dengan data *training* sebesar 80% yaitu 183 data serta jumlah data *testing* sebesar 20% yaitu 46 data menghasilkan nilai akurasi 97,78%, nilai presisi 97,78%, dan *recall* 100,00% (Saadah et al., 2021).

Penelitian berikutnya yaitu penggunaan algoritma KNN setelah dilakukan pembobotan TF-IDF terhadap judul penelitian yang dilakukan oleh dosen yang digunakan untuk dasar dari klasifikasi (studi kasus Universitas Siliwangi), pada penelitian tersebut dilakukan pengelompokkan mengenai judul penelitian yang

telah dilakukan untuk mencapai tujuan yang diharapkan. Dengan menggunakan pembobotan TF-IDF untuk *text mining* terhadap kumpulan data judul penelitian dosen di Universitas Siliwangi maka dapat disimpulkan nilai K optimal dalam penelitian tersebut yaitu $k = 4$ menghasilkan akurasi sebesar 90,7% dan menghasilkan 115 judul *science*, 142 judul *social*, dan 98 judul *education* dari jumlah keseluruhan judul penelitian dosen yaitu 355 (Supriatman, 2021).

Penelitian berikutnya yaitu penggunaan algoritma KNN untuk klasifikasi opini pada berita vaksinasi di media sosial *twitter*, penelitian tersebut melakukan split dataset untuk memperoleh data latih dan data uji yang mana menerapkan tiga skenario uji yaitu 80% : 20%, 70% : 30%, dan 60% : 40%. Dalam penelitian ini menggunakan beberapa nilai k yaitu $k = 3, 5, 7, 9$ dan setiap pengujian akan menggunakan 3 skenario *split* data, sehingga didapatkan akurasi terbaik dari hasil pengujian serta analisa yang dilakukan yaitu pada split data 80% untuk data latih dan 20% data uji dengan nilai $k = 3$ dan $k = 5$ yang mana keduanya menghasilkan akurasi sebesar 90.00% (Sandi et al., 2022).

Penelitian berikutnya mengenai analisis dari klasifikasi terhadap sentimen ulasan shopee menggunakan *word cloud* dengan menerapkan metode *naïve bayes* dan KNN, pada penelitian tersebut menjelaskan bahwa sentimen ulasan positif ataupun negatif memberikan cerminan pendapat pengguna mengenai berbagai produk dan layanan yang tersedia, pembeli yang membaca ulasan tersebut secara signifikan bisa terpengaruhi dan dapat memilih produk maupun layanan yang diinginkan. Penelitian tersebut menggunakan sebanyak 500 data yang diklasifikasikan dengan salah satu *tool* yaitu *orange* dengan metode *naïve bayes* dan KNN, yang mana digunakan juga metode *word cloud* untuk memperoleh informasi mengenai topik yang paling banyak diulas oleh pengguna. Dengan penerapan metode *naïve bayes* diperoleh hasil akurasi 0,914, *Precision* 0,915, *recall* 0,914, dan *FI score* 0,916. Kemudian dengan penerapan KNN diperoleh hasil akurasi 0,928, *Precision* 0,929, *recall* 0,928, dan *FI score* 0,926. Hal tersebut menunjukkan bahwa dalam penelitian ini penerapan metode KNN lebih baik. Dari hasil metode *word cloud* diperoleh informasi kata yang paling sering diulas dengan kategori sentimen positif yaitu mengenai kata gratis, bagus, suka, murah, mudah,

dan cepat. Sedangkan informasi kata yang paling sering diulas dengan kategori sentimen negatif yaitu mengenai kata kecewa, jelek, mahal, bohong, ribet, dan perbaiki (Limbong et al., 2022).

Penelitian berikutnya mengenai penggunaan algoritma Naïve Bayes dan KNN dalam melakukan klasifikasi penjurusan siswa kelas XI. Dalam penelitian ini menggunakan data nilai siswa pada semester 2 sebanyak 277 data, serta menentukan 4 atribut nilai mata pelajaran diantaranya PAI, prakarya, sejarah, dan PPKN. Penelitian ini menggunakan *tool* rapidminer untuk melakukan pengolahan data, dan dilakukan *split* data dengan rincian data latih 80% dan data uji 20%. Hasil akurasi yang diperoleh dalam penelitian ini adalah 81,82% dengan algoritma Naïve Bayes dan 92,73% dengan algoritma KNN, keduanya menggunakan sampel 55 data dari sejumlah 277 data (Putra & Putri, 2022).

Penelitian berikutnya mengenai penerapan algoritma KNN untuk mengklasifikasikan judul berita hoax di internet berdasarkan kategori berita hoax atau tidak. Dalam penelitian ini menggunakan data yang telah dikumpulkan secara manual oleh peneliti yaitu sebanyak 75 judul berita fakta yang diperoleh dari situs detikcom dan cnnindonesia, kemudian 75 judul berita hoax yang diperoleh dari situs turnbackhoax, pengumpulan data tersebut dilakukan pada tanggal 24 November 2021 sampai 22 Januari 2022. Pada penelitian ini menggunakan 3 jenis skenario pengujian yaitu 90% : 10%, 80% : 20%, dan 70% : 30%. Sehingga diperoleh akurasi terbaik pada skenario 90 : 10 dengan akurasi 93,33%, *precision* 100%, *recall* 80%, dan *f1-score* 88,89% (Hendriyanto & Sari, 2022).

Penelitian berikutnya mengenai perbandingan penggunaan algoritma Naïve Bayes dan KNN dalam mengklasifikasikan status gizi balita. Status gizi balita bisa diketahui dengan menggunakan sebuah metode yaitu antroperti, namun menghitung dengan cara manual membuat proses membutuhkan waktu lebih lama dan kemungkinan lebih besar terjadi kesalahan pada saat memasukkan data, penelitian ini bertujuan untuk membuat sebuah *website* keputusan status gizi balita menjadi lebih cepat. Data yang digunakan sebanyak 412 data yang diperoleh dari beberapa posyandu di Jakarta Utara pada jangka waktu 2017 sampai 2020. Sehingga diperoleh hasil *accuracy* sebesar 91,79%, *recall* 91,79%, dan *precision*

91,17% menggunakan algoritma KNN dan diperoleh hasil *accuracy* sebesar 80,60%, *recall* 80,60%, *precision* 76,66% menggunakan algoritma Naïve Bayes (Setiawan & Triayudi, 2022).

2.2 Dasar Teori

2.2.1 Data mining

Data *mining* atau penambangan data yaitu suatu langkah dalam *Knowledge Discovery in Database s* (KDD). *Knowledge discovery* sebagai sebuah proses yang terdiri dari pembersihan data, integrasi data, pemilihan data, transformasi data, data *mining*, evaluasi pola (*pattern evaluation*) dan *knowledge presentation*. Data *mining* adalah proses penggalian informasi dari kumpulan data yang sangat besar (Firdaus, 2017). Data *mining* telah berkembang menjadi sebuah disiplin ilmu yang dibangun diatas bidang kecerdasan buatan (AI) dan rekayasa pengetahuan (KE). Penambangan data berakar pada *machine learning* dan statistika, akan tetapi data *mining* telah berkembang dan memasuki bidang yang lainnya dalam ilmu komputer dan disiplin ilmu yang lainnya seperti biologi, lingkungan, keuangan, jaringan, dan lain-lain (Sudarsono et al., 2021).

2.2.2 Preprocessing

Dalam penerapan data mining terdapat sebuah proses pengolahan data yang disebut preprocessing. *Preprocessing* yaitu tahapan-tahapan untuk memproses data mentah yang nantinya akan dimasukkan ke dalam sebuah metode klasifikasi data *mining*, pengolahan data ini berfungsi untuk mengoptimalkan kualitas suatu data, dan membuat proses penambangan data menjadi lebih efisien. (Irfa et al., 2018). Pada *preprocessing* akan dilakukan beberapa tahapan untuk mengolah data yang masih mentah diantaranya yaitu data *cleaning*, *case folding*, *tokenizing*, *stopword*, dan *stemming*.

- *Cleaning* merupakan tahapan untuk menghapus karakter yang tidak berpengaruh terhadap pengolahan data seperti tanda baca, angka, link, dan lain-lain.
- *Case folding* merupakan tahapan untuk mengubah semua huruf menjadi huruf besar atau kecil semua.

- *Tokenizing* merupakan tahapan untuk memisahkan teks yang berupa kalimat, paragraf maupun dokumen menjadi potongan kata yang disebut token, tujuannya adalah untuk memperoleh potongan kata yang dijadikan entitas dan mempunyai nilai dalam matriks dokumen teks yang akan dianalisis.
- *Stopword* merupakan tahapan untuk menghilangkan kosa kata atau simbol yang tidak memiliki informasi yang berguna seperti kata penghubung dan kata keterangan.
- *Stemming* merupakan tahapan untuk mendapatkan kata dasar dari sebuah token yaitu dengan menghapus kata awalan, akhiran, sisipan, dan *confixes* (kombinasi kata awalan dan akhiran) (Hadi & Sukamto, 2020)(Irfa et al., 2018)(Limbong et al., 2022).

2.2.3 Klasifikasi

Pada data mining terdapat beberapa metode salah satunya yaitu klasifikasi, setelah data diolah menggunakan tahapan-tahapan preprocessing selanjutnya akan dilakukan tahap klasifikasi. Klasifikasi yaitu sebuah bentuk analisis data yang mengekstraksi pola-pola yang mendeskripsikan kelas pada data, algoritma klasifikasi memprediksi label kelas kategorikal terhadap suatu data untuk mengklasifikasikannya ke dalam salah satu kelas yang telah teridentifikasi sebelumnya (Nugroho & Religia, 2021). Klasifikasi adalah proses untuk menemukan pola yang menggambarkan dan membedakan antara kelas yang satu dengan yang lain dan berfungsi untuk menunjukkan suatu objek termasuk dalam label tertentu yang telah teridentifikasi sebelumnya. Klasifikasi terbagi menjadi 2 proses, pertama yaitu proses pembelajaran (*training*) yang mana sebuah algoritma klasifikasi dibangun untuk menganalisis data latih kemudian diungkapkan kedalam bentuk aturan klasifikasi, kedua yaitu proses klasifikasi yang mana data uji akan digunakan untuk mengevaluasi tingkat akurasi dari aturan klasifikasi tersebut (Maulana et al., 2021).

2.2.4 *K-Nearest Neighbor* (KNN)

Dalam proses klasifikasi data terdapat beberapa algoritma yang bisa digunakan salah satunya yaitu KNN. KNN yaitu suatu metode yang dimanfaatkan untuk mengklasifikasikan suatu objek berlandaskan data pembelajaran yang

memiliki jarak terdekat dengan objek tersebut, tujuannya adalah untuk melakukan klasifikasi sebuah objek baru yang sesuai dengan atribut dan pola pelatihan (Limbong et al., 2022). KNN merupakan salah satu metode yang digunakan untuk mengklasifikasikan data, cara kerja dari algoritma KNN ini adalah dengan mengklasifikasikan data berdasarkan jarak terdekat satu data dengan data yang lain, dekat atau jauhnya jarak tersebut bisa dihitung menggunakan besaran jarak (Irfa et al., 2018).

Algoritma KNN merupakan algoritma *supervised learning*, yang mana *supervised learning* adalah sebuah pembelajaran yang terawasi apabila *output* yang diinginkan telah teridentifikasi sebelumnya, pada umumnya pembelajaran tersebut digunakan pada data yang telah mempunyai label. Algoritma KNN melakukan proses klasifikasi objek dengan mengacu pada atribut dan sampel dari data *training* (data latih), jika algoritma ini diberikan titik *query* maka algoritma ini akan menemukan sekumpulan k objek atau titik latih yang memiliki jarak paling dekat dengan titik *query*. Klasifikasi menggunakan jumlah suara terbanyak dari k klasifikasi objek, dan algoritma KNN menggunakan klasifikasi tetangga untuk nilai prediksi dari *query instance* baru. Terdapat metode perhitungan jarak pada algoritma KNN salah satunya yaitu *Euclidean Distance* yang merupakan perhitungan jarak paling dekat, perhitungan jarak paling dekat ini digunakan untuk memperoleh jumlah kesamaan yang dihitung berdasarkan kesamaan kata yang terdapat dalam suatu paragraf (Hardiyanto & Rahutomo, 2016). Adapun tahapan implementasi dari algoritma KNN yaitu yang pertama dengan mendefinisikan jumlah nilai k sebagai ketetanggaan yang merupakan jarak terdekat dari data *training* ke data *testing*, tahap yang kedua yaitu menghitung jarak antara nilai vektor data *testing* dengan semua nilai vektor data *training* menggunakan perhitungan jarak *Euclidean Distance* yang ditunjukkan pada persamaan (1) dan (2).

$$d(x, y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (1)$$

Atau

$$d_{(i,j)} = \sqrt{(|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{ip} - x_{jp}|^2)} \quad (2)$$

Keterangan :

$d_{(i,j)}$ = Jarak dokumen i ke j.

$x_{i(n)}$ = Kata ke-n pada dokumen i.

$x_{j(n)}$ = Kata ke-n pada dokumen j.

Kemudian pada tahap ketiga yaitu mengambil sejumlah nilai parameter k dari data latih yang terdekat, dan tahap terakhir yaitu menentukan kategori yang dominan sebagai kategori untuk data uji (Hadi & Sukanto, 2020).

2.2.5 Evaluasi

Setelah penerapan algoritma KNN untuk klasifikasi data maka akan dilakukan proses evaluasi hasil akhir. Evaluasi algoritma KNN menggunakan *confusion matrix* terhadap akurasi, presisi, *recall*, dan *F-Measure* yang mana *confusion matrix* merupakan sebuah bentuk perhitungan suatu perbandingan yang mendeskripsikan nilai positif dan nilai negatif dari hasil pengklasifikasian data. Hasil dari perhitungan evaluasi dan validasi akan digunakan untuk menghitung nilai dari *Precision*, *recall*, *f-measure* dan *accuracy* sebagai mana yang ditunjukkan pada tabel 1.

Tabel 2. 1 Model *Confusion matrix*

		Nilai Prediksi	
		<i>Positive</i>	<i>Negative</i>
Nilai Aktual	<i>Positive</i>	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
	<i>Negative</i>	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

- *Precision* adalah perbandingan dari hasil *True Positive* (TP) dengan seluruh dokumen yang diprediksi sebagai *positive* oleh model klasifikasi, ditunjukkan oleh persamaan (3).

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

- *Recall* adalah perbandingan hasil *True Positive* (TP) dengan seluruh dokumen yang memiliki nilai *positive*, ditunjukkan oleh persamaan (4).

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

- *F-Measure* adalah gabungan dari hasil perhitungan *precision* dan *recall* yang digunakan untuk melihat estimasi kinerja klasifikasi, ditunjukkan oleh persamaan (5).

$$f - measure = \frac{2TP}{2TP + FP + FN} \quad (5)$$

- *Accuracy* adalah hasil dari persentase keseluruhan dokumen yang diprediksi benar, ditunjukkan oleh persamaan (6) (Hadi & Sukamto, 2020).

$$Accuracy = \frac{\sum TP + \sum TN}{\sum Data Uji} \quad (6)$$

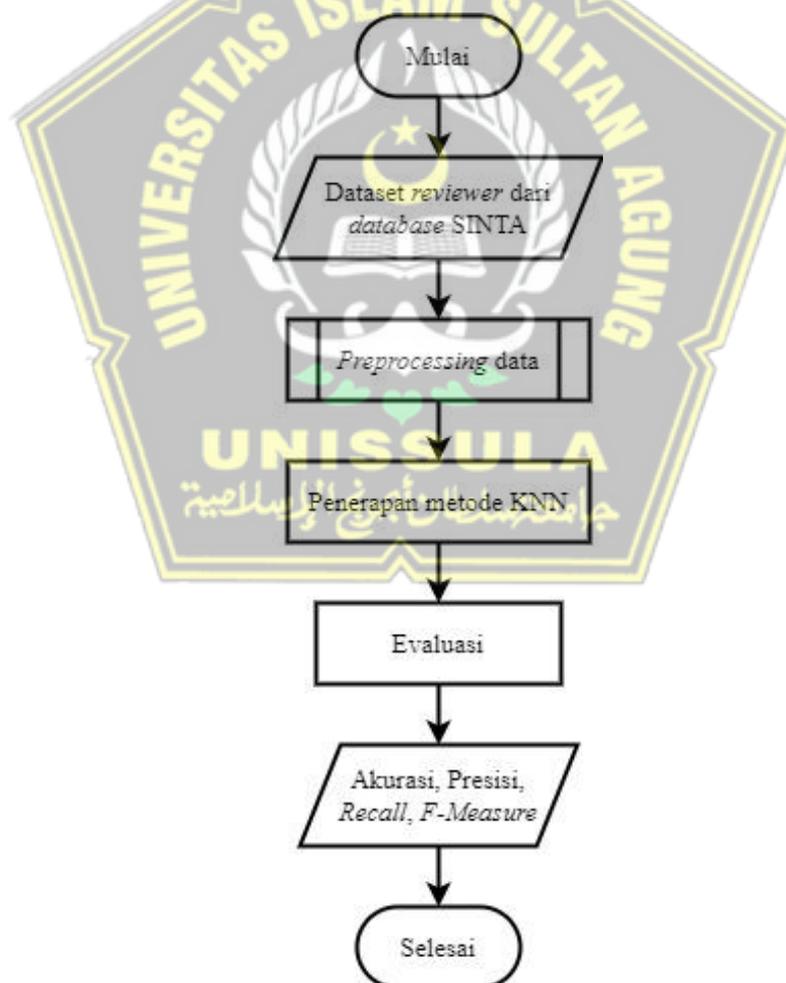
2.2.6 Rubix ML

Rubix ML merupakan sebuah pustaka pembelajaran mesin (*Machine Learning*) yang bersifat gratis dan memungkinkan untuk membuat program melalui pembelajaran data menggunakan bahasa pemrograman PHP. *Library* rubix ML menyediakan berbagai alat untuk pembelajaran mesin mulai dari ELT sampai pelatihan, *cross* validasi, dan tersedia lebih dari 40 algoritma yang dapat digunakan baik pembelajaran yang terawasi dan pembelajaran yang tidak diawasi. Pada *library* ini juga menyediakan sebuah tutorial dan konten belajar yang membantu untuk menyelesaikan sebuah *project* menggunakan *machine learning* (DalPino, 2022).

BAB III METODE PENELITIAN

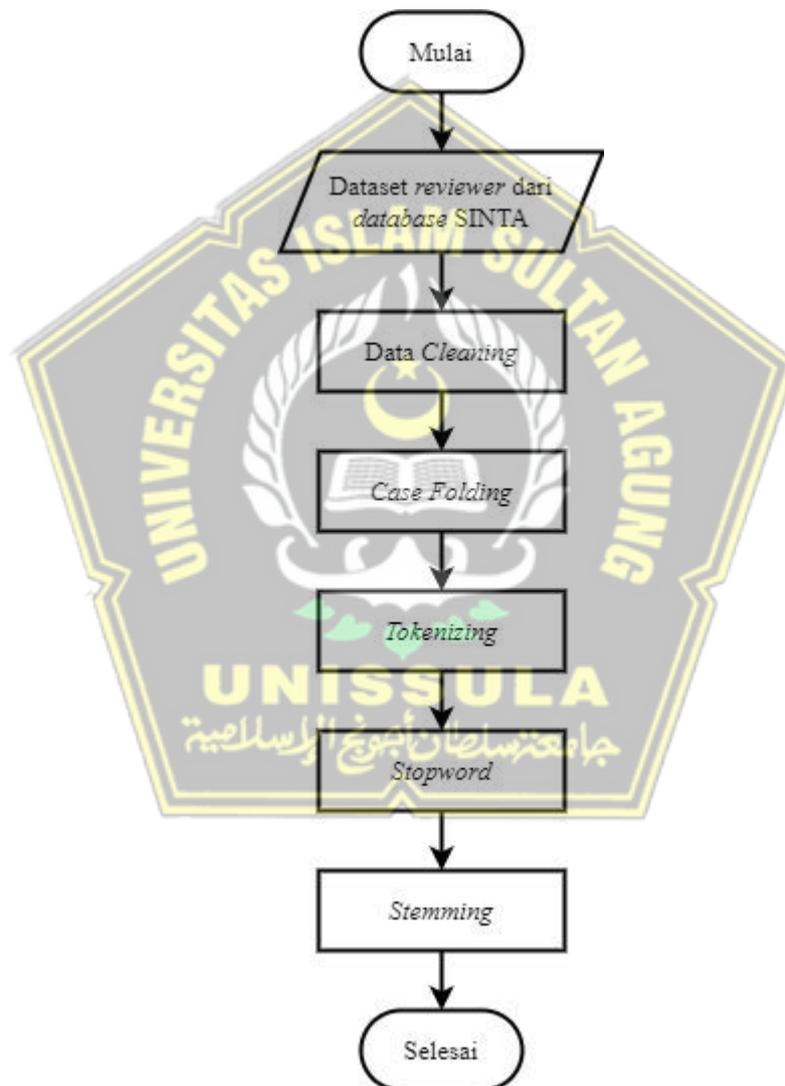
3.1 Metode Penelitian

Pada penelitian tugas akhir ini meliputi beberapa tahapan yang dilakukan diantaranya yaitu pengumpulan data, kemudian dilakukan pengolahan data yang disebut dengan *preprocessing*, data yang sudah melalui *preprocessing* akan diolah menggunakan metode klasifikasi data *mining* dengan menerapkan algoritma KNN, kemudian dari penerapan metode tersebut akan dihasilkan sebuah evaluasi akhir yaitu tingkat akurasi, presisi, *recall*, dan *F-Measure*. Adapun tahapan penelitian dalam tugas akhir ini ditunjukkan pada gambar 3.1.



Gambar 3. 1 Tahap-tahap Penelitian

Pada *preprocessing* data terdapat beberapa tahapan untuk mengolah data diantaranya yaitu data *cleaning* untuk menghilangkan karakter-karakter yang tidak memiliki pengaruh, *case folding* untuk mengubah karakter huruf menjadi besar atau kecil semua, *tokenizing* untuk memisahkan teks menjadi potongan kata yang disebut token, *stopword* untuk menghilangkan kosa kata atau simbol, dan terakhir yaitu *stemming* untuk memperoleh kata dasar dari suatu token. Adapun tahap-tahap *preprocessing* tersebut ditunjukkan pada gambar 3.2.



Gambar 3.2 Flowchart tahap *Preprocessing*

a. Dataset yang digunakan

Pada penelitian ini menggunakan data judul publikasi jurnal terindeks scopus sebagai data *training* dan data judul publikasi jurnal *reviewer* sebagai data

testing yang diperoleh dari *database* SINTA. Yang mana data tersebut telah divalidasi terlebih dahulu oleh pihak yang memiliki izin untuk mengakses data tersebut.

b. Studi kasus

Studi kasus dalam penelitian tugas akhir ini adalah klasifikasi bidang ilmu dengan *reviewer* jurnal yang dikelompokkan ke dalam 5 bidang ilmu yaitu *Arts & Humanities, Engineering & Technology, Life Sciences & Medicine, Natural Sciences, dan Social Sciences & Management* menggunakan metode klasifikasi data *mining* dengan menerapkan algoritma KNN.

c. Koleksi data

Pada proses koleksi data akan dilakukan labelisasi yaitu klasifikasi kepakaran *reviewer* ke dalam 5 bidang ilmu utama sebagaimana telah ditetapkan sebelumnya yang akan dijadikan sebagai label. Dalam penelitian ini menggunakan data sebanyak 500 dataset dengan rincian 100 data untuk setiap pengelompokan bidang ilmu. Adapun contoh datanya ditunjukkan pada tabel 3.1.

Tabel 3. 1 Contoh data

No.	Area	Bidang Ilmu Kepakaran
1.	<i>Political Mediation in Ireland: Campaigning between Traditional and Tabloid Markets.</i>	<i>Social Science and Management</i>
2.	<i>Electrical Engineering Education in Indonesia.</i>	<i>Engineering and Technology</i>
3.	<i>Interpreting the atmospheric neutrino anomaly.</i>	<i>Natural Science</i>
4.	<i>Small doses of subcutaneous heparin preventing deep-vein thrombosis in major surgery.</i>	<i>Life Science and Medicine</i>
5.	<i>Audiovisual patterns in Wayang Kulit.</i>	<i>Arts & Humanity</i>

d. *Preprocessing*

Pada tahap *preprocessing* akan dilakukan pengolahan data *reviewer* untuk memperoleh kualitas data yang baik dan bersih menggunakan beberapa tahapan diantaranya yaitu data *cleaning*, *case folding*, *tokenizing*, *stopword*, dan *stemming* agar diperoleh data yang bersih dan siap untuk dilakukan proses klasifikasi dengan menerapkan algoritma KNN. Adapun contoh input data beserta *output*-nya dalam tahapan *preprocessing* data, ditunjukkan pada tabel 3.2 sampai dengan tabel 3.6.

Tabel 3. 2 Proses Data *cleaning*

No.	Input	Output
1.	<i>Political Mediation in Ireland: Campaigning between Traditional and Tabloid Markets.</i>	<i>Political Mediation in Ireland Campaigning between Traditional and Tabloid Markets</i>
2.	<i>Electrical Engineering Education in Indonesia.</i>	<i>Electrical Engineering Education in Indonesia</i>

Tabel 3. 3 Proses *Case Folding*

No.	Input	Output
1.	<i>Political Mediation in Ireland Campaigning between Traditional and Tabloid Markets</i>	<i>political mediation in ireland campaigning between traditional and tabloid markets</i>
2.	<i>Electrical Engineering Education in Indonesia</i>	<i>electrical engineering education in indonesia</i>

Tabel 3. 4 Proses *Tokenizing*

No.	Input	Output
1.	<i>political mediation in ireland campaigning between traditional and tabloid markets</i>	<i>political mediation in ireland</i>

		<i>campaigning between traditional and tabloid markets</i>
2.	<i>electrical engineering education in indonesia</i>	<i>electrical engineering education in indonesia</i>

Tabel 3. 5 Proses Stopword

No.	Input	Output
1.	<i>political mediation in ireland campaigning between traditional and tabloid markets</i>	<i>political mediation ireland campaigning traditional tabloid markets</i>
2.	<i>electrical engineering education in indonesia</i>	<i>electrical engineering education indonesia</i>

Tabel 3. 6 Proses *Stemming*

No.	<i>Input</i>	<i>Output</i>
1.	<i>political</i> <i>mediation</i> <i>ireland</i> <i>campaigning</i> <i>traditional</i> <i>tabloid</i> <i>markets</i>	<i>political</i> <i>mediation</i> <i>ireland</i> <i>campaign</i> <i>traditional</i> <i>tabloid</i> <i>market</i>
2.	<i>electrical</i> <i>engineering</i> <i>education</i> <i>indonesia</i>	<i>electrical</i> <i>engineering</i> <i>education</i> <i>indonesia</i>

e. Penerapan metode

Dalam penelitian ini menerapkan metode data *mining* yaitu klasifikasi menggunakan algoritma KNN, dengan algoritma KNN ini setiap data uji akan dibandingkan jaraknya berdasarkan dengan data pembelajaran atau data latih untuk menentukan jarak terdekatnya sehingga dapat diketahui labelnya berdasarkan 5 bidang ilmu utama.

f. Evaluasi

Tahap akhir yaitu evaluasi yang mana pada tahap ini akan menghasilkan seberapa besar akurasi penggunaan algoritma KNN dalam mengklasifikasikan data judul artikel *reviewer* ke dalam 5 bidang ilmu yang telah ditentukan dengan melihat hasil akurasi, presisi, *recall*, dan *F-Measure*, yang mana penelitian ini akan menunjukkan apakah metode KNN tersebut layak digunakan dalam proses klasifikasi data *reviewer* terhadap 5 bidang ilmu kepakaran.

3.2 Metode Perancangan Sistem

3.2.1 Analisis Kebutuhan

Analisis kebutuhan adalah tahapan untuk menganalisis fitur-fitur yang diperlukan dan dibutuhkan dalam sistem yang akan dibuat pada penelitian ini, agar sistem dapat menampilkan *output* sesuai dengan apa yang diharapkan. Adapun beberapa fitur yang harus dibuat adalah sebagai berikut :

a. Menampilkan data nama-nama *reviewer*

Fungsi pertama yang harus bisa dilakukan oleh sistem klasifikasi yaitu menampilkan nama-nama *reviewer* yang judul artikel nya akan diklasifikasi menggunakan metode KNN. Yang mana data nama *reviewer* tersebut disimpan pada *database* PhpMyAdmin.

b. Menampilkan data judul artikel *reviewer*

Fungsi selanjutnya yang harus bisa dilakukan oleh sistem yaitu menampilkan judul artikel yang ditulis oleh *reviewer*, data judul juga disimpan pada *database* PhpMyAdmin. Sistem harus bisa menampilkan semua judul artikel yang ditulis oleh salah satu *reviewer*, yang mana seorang *reviewer* bisa memiliki lebih dari satu judul artikel.

c. Melakukan klasifikasi judul

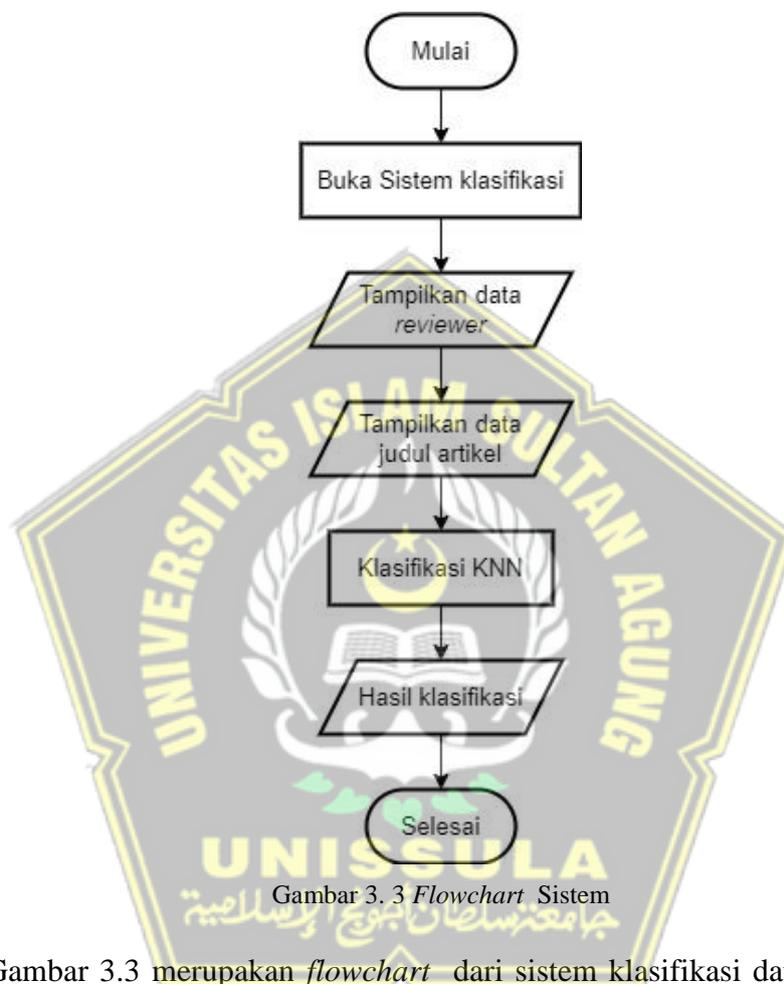
Fungsi selanjutnya yaitu sistem melakukan klasifikasi terhadap judul-judul artikel yang telah berhasil ditampilkan. Pada sistem ini setiap judul akan dicari tetangga dengan jarak terdekat nya berdasarkan data *training* nya menggunakan metode KNN dan perhitungan jarak *eucliden distance*.

d. Menampilkan hasil akurasi

Fungsi selanjutnya yaitu menampilkan hasil akurasi, setelah judul-judul artikel diklasifikasi menggunakan metode KNN maka akan mengasilkan akurasi berupa pengelompokannya ke dalam salah satu dari 5 bidang ilmu yang telah ditentukan.

3.2.2 Alur Penggunaan Sistem

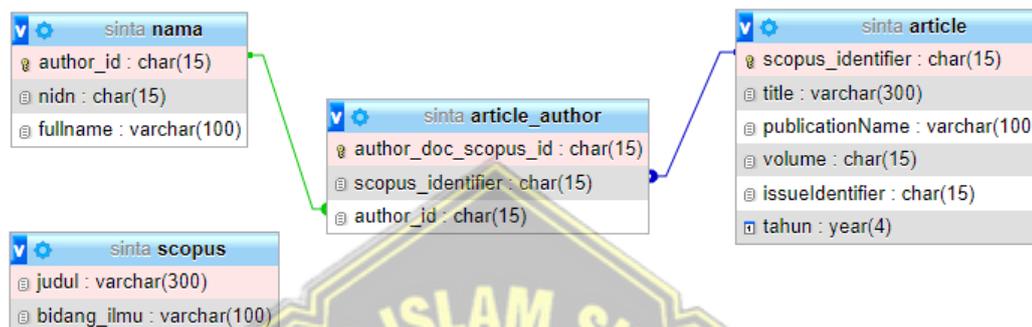
Analisis sistem merupakan tahapan untuk menganalisis cara kerja dari sistem yang dibuat pada penelitian ini, adapun *flowchart* sistem dari penelitian ditunjukkan pada gambar 3.3.



Gambar 3.3 merupakan *flowchart* dari sistem klasifikasi data *reviewer* ke dalam 5 bidang ilmu yang telah ditentukan. Langkah pertama yaitu masuk pada *website* sistem klasifikasi, kemudian sistem akan menampilkan data-data *reviewer* dan data judul-judul arikel yang ditulis oleh *reviewer*. Yang mana seorang *reviewer* memiliki atau dapat menulis lebih dari satu judul artikel. Kemudian judul-judul artikel tersebut akan dilakukan klasifikasi menggunakan metode KNN. Setelah itu sistem akan menampilkan hasil dari klasifikasi berupa pengelompokannya ke dalam salah satu dari 5 bidang ilmu yang telah ditentukan

3.3 Desain Database

Dalam merancang *database* diperlukan sebuah diagram yang berfungsi untuk menunjukkan hubungan atau relasi antara entitas dengan entitas yang lain, dan setiap entitas memiliki atribut. *Database relation design* pada penelitian ini ditunjukkan pada gambar 3.4.



Gambar 3.4 Database relation design

Gambar 3.4 merupakan tampilan *database relation design* pada penelitian. Terdapat entitas berupa nama, *article*, *article_author*, dan *scopus*, kemudian setiap entitas memiliki atributnya masing-masing seperti yang ada pada gambar. Yang mana pada gambar *database relation design* tersebut menjelaskan sebuah relasi yaitu *one to many* dimana seorang *reviewer* atau penulis dapat menulis *article* lebih dari satu. Dan terdapat satu tabel yang tidak memiliki relasi atau berdiri sendiri yaitu tabel *scopus*, karena tabel *scopus* berisi data judul jurnal terindeks *scopus* dan label bidang ilmu yang digunakan sebagai data *training*.

Pada penelitian ini menggunakan data dengan format *file* CSV dan sql yang kemudian data-data tersebut diimport ke dalam *database* PhpMyAdmin, sehingga dibuatlah 1 *database* dengan nama *sinta* dengan 4 tabel yaitu tabel nama, tabel *article*, tabel *article_author*, dan tabel *scopus*. adapun desain tabel pada *database* pada penelitian ini ditunjukkan pada tabel 3.7 sampai dengan 3.10.

1. Tabel nama

Tabel nama menampung id *author* atau *reviewer*, NIDN dan nama lengkap dari *reviewer*, yang mana data pada tabel ini digunakan untuk menampilkan nama *reviewer* pada halaman prediksi.

Tabel 3. 7 Desain tabel nama

Atribut	Tipe Data	Keterangan
<i>author_id</i>	CHAR	Berfungsi untuk menampung karakter dan menyimpan <i>id author</i> .
<i>nidn</i>	CHAR	Berfungsi untuk menampung karakter dan menyimpan NIDN <i>reviewer</i> .
<i>fullname</i>	VARCHAR	Berfungsi untuk menampung karakter dan menyimpan nama lengkap <i>reviewer</i> .

2. Tabel *article*

Tabel *article* menampung *scopus identifier*, judul jurnal yang ditulis *reviewer*, nama jurnal publikasi, volume jurnal, *issue identifier*, dan tahun diterbitkannya jurnal. Yang mana data pada tabel ini digunakan sebagai data *testing*.

Tabel 3. 8 Desain tabel *article*

Atribut	Tipe Data	Keterangan
<i>scopus_identifier</i>	CHAR	Berfungsi untuk menampung karakter dan menyimpan <i>scopus identifier</i> .
<i>title</i>	VARCHAR	Berfungsi untuk menampung karakter dan menyimpan judul jurnal yang ditulis <i>reviewer</i> .
<i>publicationName</i>	VARCHAR	Berfungsi untuk menampung karakter dan

		menyimpan nama publikasi.
volume	CHAR	Berfungsi untuk menampung karakter dan menyimpan volume jurnal.
<i>issueIdentifier</i>	CHAR	Berfungsi untuk menampung karakter dan menyimpan <i>issue identifier</i> .
tahun	YEAR	Berfungsi untuk menampung data berupa tahun.

3. Tabel *article_author*

Tabel *article_author* mempunyai 3 atribut berupa *author_doc_scopus_id*, *scopus_identifier*, dan *author_id*.

Tabel 3. 9 Desain tabel *article_author*

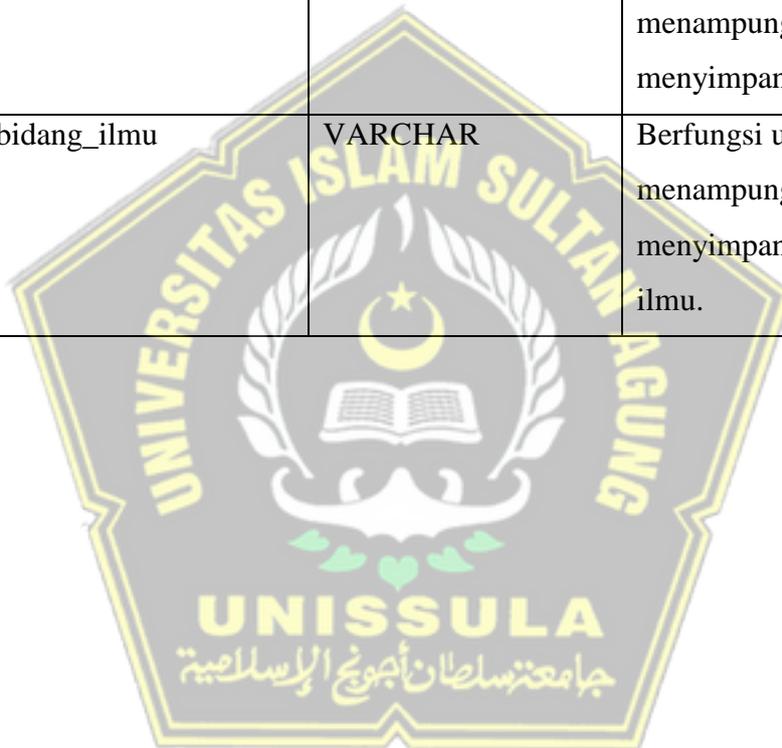
Atribut	Tipe Data	Keterangan
<i>author_doc_scopus_id</i>	CHAR	Berfungsi untuk menampung karakter dan menyimpan id dari tabel <i>article_author</i> .
<i>scopus_identifier</i>	CHAR	Berfungsi untuk menampung karakter dan menyimpan <i>scopus identifier</i> .
<i>author_id</i>	CHAR	Berfungsi untuk menampung karakter dan menyimpan <i>id author</i> .

4. Tabel scopus

Tabel scopus memiliki 2 atribut yaitu judul dan bidang_ilmu. Tabel scopus ini menampung data-data judul jurnal yang terindeks pada scopus beserta label pengelompokan bidang ilmunya, yang mana data scopus ini digunakan sebagai data *training* dalam penelitian.

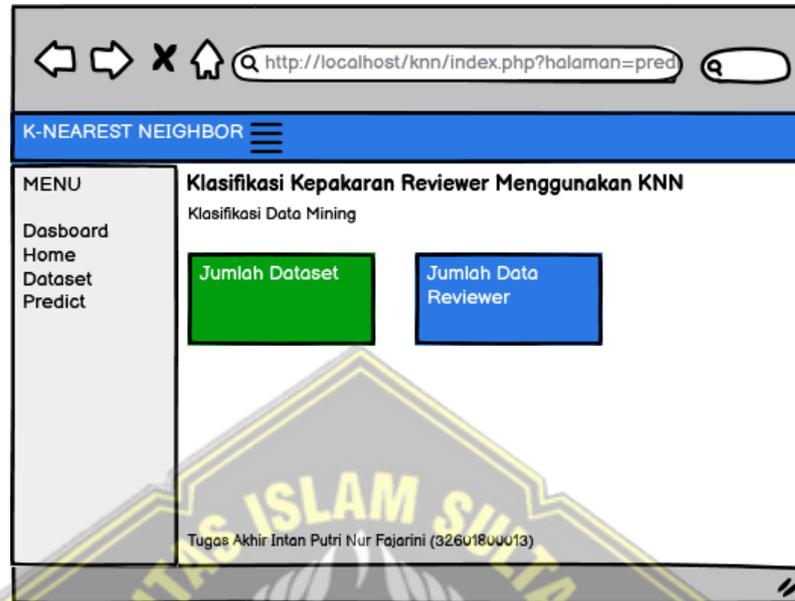
Tabel 3. 10 Desain tabel scopus

Atribut	Tipe Data	Keterangan
judul	VARCHAR	Berfungsi untuk menampung karakter dan menyimpan judul jurnal.
bidang_ilmu	VARCHAR	Berfungsi untuk menampung karakter dan menyimpan label bidang ilmu.



3.4 Perancangan *User Interface*

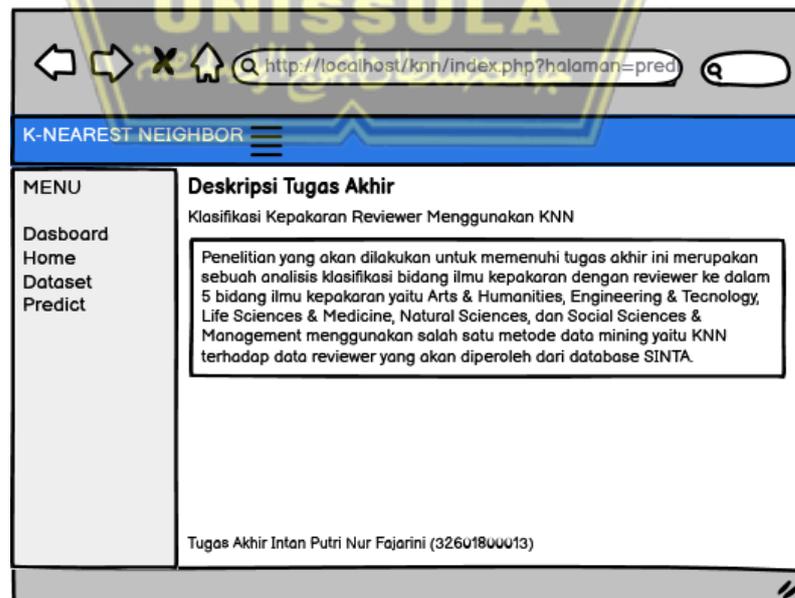
a. Menu *Dashboard*



Gambar 3. 5 Rancangan menu *Dashboard*

Gambar 3.5 merupakan rancangan *user interface* untuk menu *Dasboard* pada sistem klasifikasi, pada menu ini akan menampilkan jumlah dataset yang digunakan dalam penelitian tugas akhir, dan jumlah data *reviewer* yang judul artikelnya akan diklasifikasi secara otomatis menggunakan KNN.

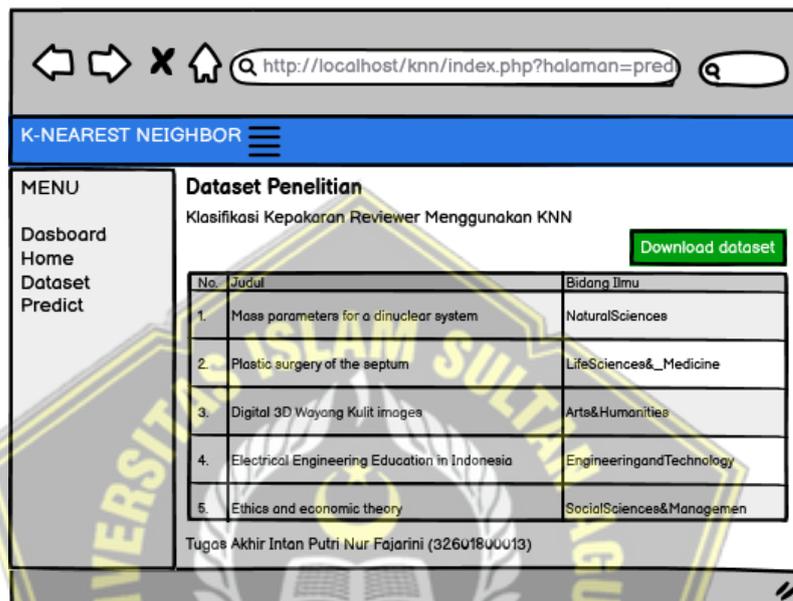
b. Menu *Home*



Gambar 3. 6 Rancangan menu *home*

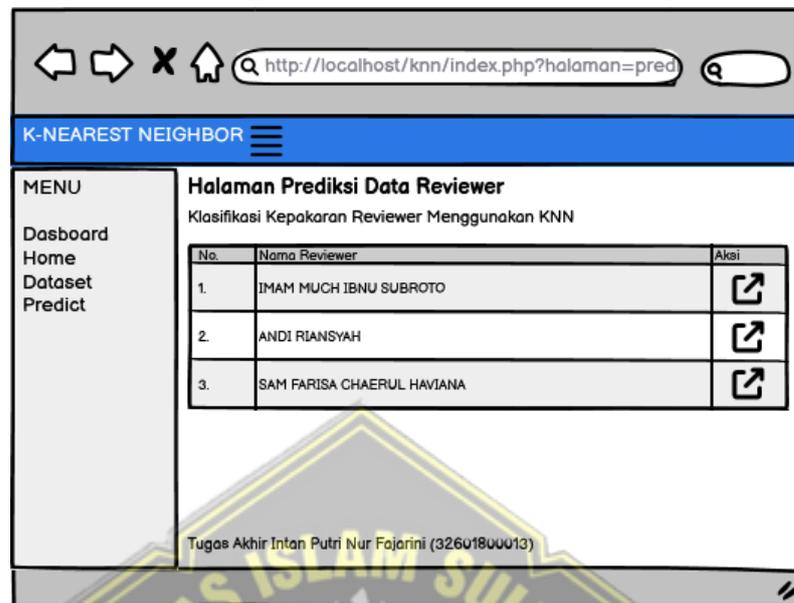
Gambar 3.6 merupakan rancangan *user interface* untuk menu *home* pada sistem klasifikasi, pada menu ini akan menampilkan deskripsi mengenai tugas akhir. Adanya menu *home* ini agar saat *user* masuk pada sistem dapat mengetahui secara singkat fungsi dibuatnya sistem klasifikasi *reviewe*.

c. Menu Dataset

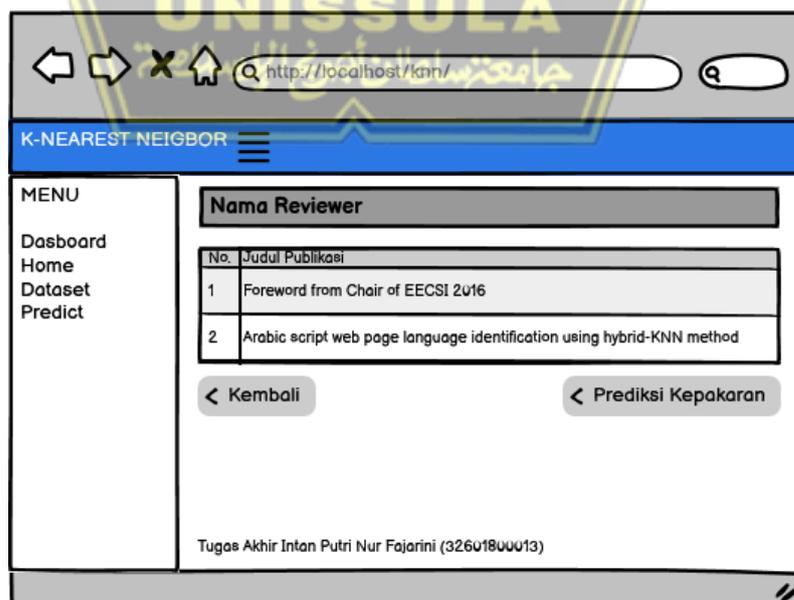


Gambar 3. 7 Rancangan menu dataset

Gambar 3.7 merupakan rancangan *user interface* untuk menu dataset pada sistem klasifikasi, pada menu ini akan menampilkan contoh atau sampel dataset scopus yang digunakan untuk data *training*, terdapat tabel dengan kolom judul dan kolom bidang ilmu. Pada menu ini juga terdapat tombol *download* dataset scopus untuk melihat keseluruhan datanya.

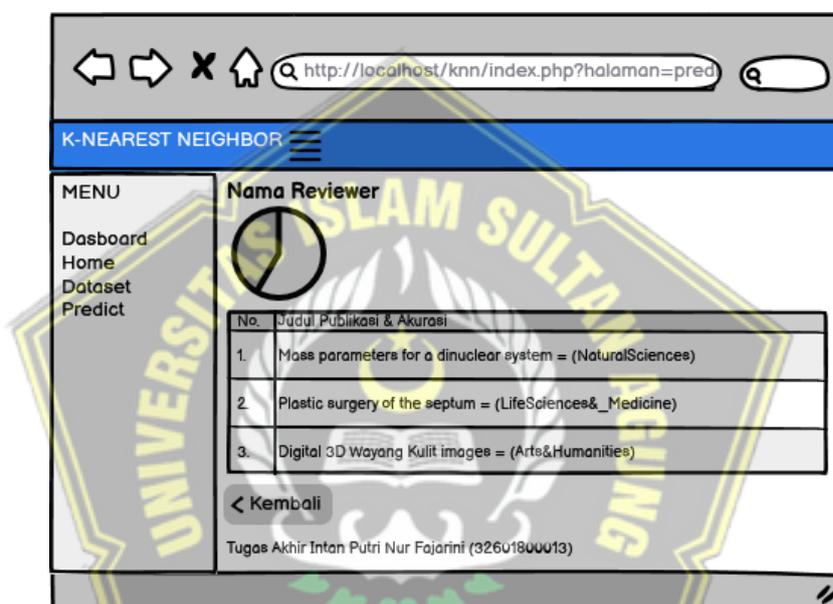
d. Menu *Predict*Gambar 3. 8 Rancangan menu *predict*

Gambar 3.8 merupakan rancangan *user interface* untuk menu prediksi pada sistem klasifikasi, pada menu ini akan menampilkan nama-nama *reviewer* yang judul artikelnnya akan diklasifikasi untuk mengetahui bidang kepakaran *reviewer*, terdapat tabel dengan kolom nama untuk nama *reviewer* dan kolom aksi untuk menampilkan judul-judul artikel pada halaman berbeda.

e. Halaman *reviewer* yang menampilkan daftar judul artikelGambar 3. 9 Rancangan halaman *reviewer* dan judul artikel

Gambar 3.9 merupakan rancangan *user interface* untuk halaman *reviewer* dan judul-judul artikel yang ditulis oleh *reviewer*. halaman ini akan menampilkan nama *reviewer* dan sebuah tabel yang berisi daftar judul artikel. Kemudian pada bagian bawah halaman terdapat tombol kembali dan tombol untuk memprediksi keparakan dari judul-judul artikel tersebut yang akan ditampilkan pada halaman yang berbeda lagi.

f. Halaman *reviewer* yang menampilkan label keparakan



Gambar 3. 10 Rancangan halaman *reviewer* dan keparakan

Gambar 3.10 merupakan rancangan *user interface* untuk halaman *reviewer* dan hasil keparannya pada sistem klasifikasi. Halaman ini merupakan halaman lanjutan pada menu *predict*, saat tombol pada kolom aksi diklik akan muncul halaman *reviewer*. Pada halaman *reviewer* akan menampilkan sebuah diagram lingkaran untuk menunjukkan persentase paling banyak dari hasil klasifikasi judul-jurnal artikel yang ditulis oleh *reviewer*, diagram ini bertujuan untuk mempermudah *user* dalam mengambil kesimpulan. Kemudian dibawah diagram tetap akan ditampilkan tabel dengan kolom judul beserta hasil akurasi terbaiknya.

BAB IV

HASIL DAN ANALISIS PENELITIAN

4.1 Analisis Penerapan Metode Penelitian

4.1.1 Dataset

Dalam penelitian ini menggunakan data *training* dan data *testing* untuk pengujian, dataset yang digunakan sebagai data *training* adalah data judul jurnal terindeks di scopus yang berjumlah 500 data dengan rinciang 100 data *Arts & Humanities*, 100 data *Engineering & Technology*, 100 data *Life Sciences & Medicine*, 100 data *Natural Sciences*, dan 100 data *Social Sciences & Management*. Kemudian untuk data *testing* kepakaran menggunakan data *reviewer* yang berjumlah 171 data, dan data judul artikel dengan jumlah keseluruhan data 716 data.

Adapun proses *input* data dilakukan secara manual, yaitu dengan melakukan *import* data-data yang dimiliki dan diperlukan untuk penelitian berupa data scopus untuk proses data *training*, data nama *reviewer*, data judul *article*, dan data *article_author* ke dalam *database* PhpMyAdmin. Sebelum data diimport ke *database*, harus membuat sebuah *database* baru pada PhpMyAdmin dengan nama “sinta”, setelah itu membuat tabel-tabel yang diperlukan sesuai dengan *file* data yang akan diimport beserta kolom-kolomnya dengan memberikan tipe data yang sesuai agar saat data diimport dapat dibaca atau dikenali oleh *database*. Kemudian untuk menampilkan data pada sistem klasifikasi kepakaran yaitu dengan cara memanggil dan membuat koneksi ke *database* menggunakan bahasa pemrograman PHP serta *library* Rubixml. Adapun baris program untuk memanggil dan membuat koneksi data ke *database* PhpMyAdmin ditunjukkan pada Gambar 4.1.

```

training.php > ...
33 // Load dataset dari database sql
34 $servername = "localhost";
35 $username = "root";
36 $password = "";
37
38 try {
39     $conn = new PDO("mysql:host=$servername;dbname=sinta", $username, $password);
40     // set the PDO error mode to exception
41     $conn->setAttribute(PDO::ATTR_ERRMODE, PDO::ERRMODE_EXCEPTION);
42     echo "Connected successfully";
43 } catch (PDOException $e) {
44     echo "Connection failed: " . $e->getMessage();
45 }
46
47 $query = $conn->prepare('SELECT judul, bidang_ilmu FROM scopus');
48
49 $query->execute();
50
51 $data = $query->fetchAll(PDO::FETCH_NUM);
52 $samples = [];
53 $labels = [];
54 foreach ($data as $value) {
55     $samples[] = [$value[0]];
56     $labels[] = $value[1];
57 }
58 $dataset = new Labeled($samples, $labels);

```

Gambar 4. 1 Code program untuk membaca dataset

4.1. 2 Preprocessing

Pada tahap *preprocessing* data menggunakan salah satu bahasa pemrograman yaitu PHP dengan *library* Rubixml untuk melakukan pengolahan data. Yang mana terdapat beberapa tahapan *preprocessing* yaitu data *cleaning*, *case folding*, *tokenizing*, *stopword*, dan *stemming*. Gambar 4.2 sampai dengan 4.4 adalah baris code program yang digunakan.

```

training.php > ...
111 // pipeline untuk preprocessing dan klasifikasi
112 $estimator = new PersistentModel(
113     new Pipeline([
114         //PREPROCESSING
115         new RegexFilter([
116             RegexFilter::EXTRA_WHITESPACE,
117             RegexFilter::EXTRA_WORDS, RegexFilter::EXTRA_CHARACTERS
118         ]), //text cleaning
119         new MultibyteTextNormalizer(), //case folding - lower case
120         new StopWordFilter($stopwords_en), //stopwords
121         new WordCountVectorizer(), //tokenizer
122     ])

```

Gambar 4. 2 Code program untuk *preprocessing* data

```

64 // fungsi untuk stemmer bahasa inggris
65 $stemmeringgris = function (&$sample, $offset, $context) {
66     $stemmer_en = new WordStemmer('english');
67     $words = new Word();
68     $arrwords = [];
69     foreach ($sample as $s) {
70         if ($s !== null) {
71             $array_words = $words->tokenize($s);
72
73             foreach ($array_words as $aw) {
74                 if ($aw !== null) {
75                     $stemmres = $stemmer_en->tokenize($aw);
76                     $arrwords[] = $stemmres[0];
77                 }
78             }
79             $arrimploded = implode(" ", $arrwords);
80         }
81         $sample[0] = $arrimploded;
82     }
83 };
84
85 // stemming datasets (pilih sesuai bahasa)
86 $dataset->apply(new LambdaFunction($stemmeringgris, 'stemmer'));

```

Gambar 4. 3 Code program untuk *preprocessing* data fungsi *stemmer*

```

training.php > ...
95 //daftar stopwords sesuai bahasa
96 $stopwords_en = file("../stopwords-en.txt", FILE_IGNORE_NEW_LINES);

```

Gambar 4. 4 Code program untuk membaca daftar *stopword*

4.1. 3 Penerapan metode KNN

Pada tahap penerapan metode KNN menggunakan bahasa pemrograman PHP dan *library* Rubixml untuk melakukan klasifikasi data. Gambar 4.5 adalah baris kode program yang digunakan.

```

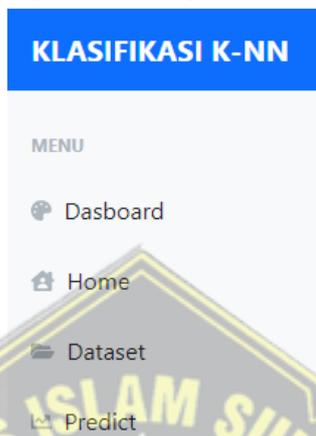
training.php > ...
111 // ALGORITMA KLASIFIKASI
112 ], new KNearestNeighbors(25)),

```

Gambar 4. 5 Code program untuk metode KNN

4.2 Hasil Implementasi *User Interface*

Tahap ini merupakan hasil implementasi dari rancangan *user interface* yang telah dirancang sebelumnya, yaitu berupa tampilan *website*. Berikut ini adalah beberapa gambar hasil dari implementasinya.



Gambar 4. 6 Tampilan menu pada sistem klasifikasi

Gambar 4.6 menampilkan menu apa saja yang terdapat pada sistem klasifikasi yang dibuat pada penelitian tugas akhir, terdapat 4 menu yaitu menu *Dashboard*, menu *home*, menu *dataset*, dan menu *predict* yang akan menampilkan data *reviewer* dan judul artikel *reviewer* beserta klasifikasinya berdasarkan dari hasil penerapan klasifikasi menggunakan metode KNN.



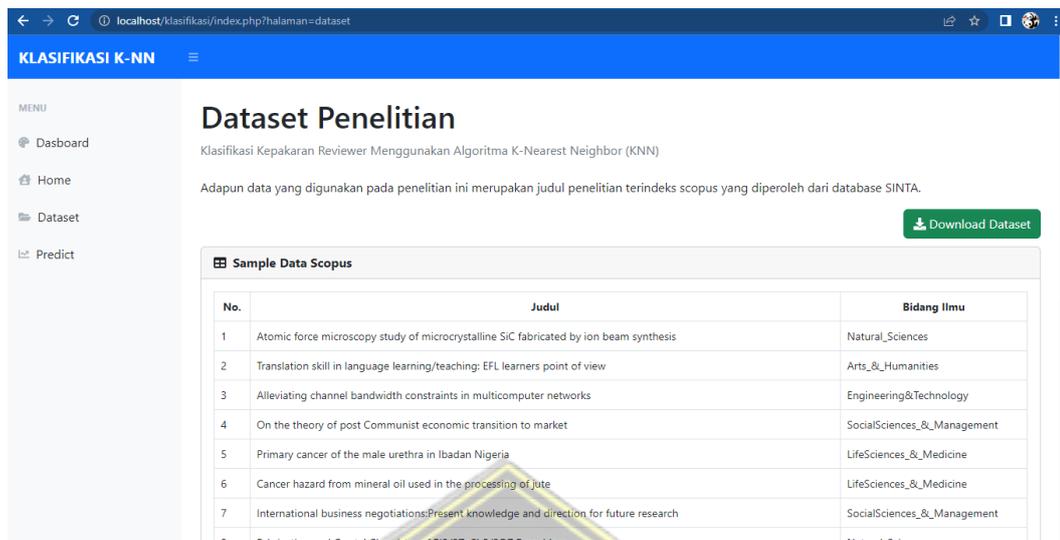
Gambar 4. 7 Tampilan halaman *Dashboard*

Gambar 4.7 merupakan hasil implementasi *user interface* dari menu *Dashboard*, yang mana menu *Dashboard* ini merupakan halaman yang pertama kali akan muncul pada saat *user* membuka sistem klasifikasi yang berupa *website*. Pada halaman ini akan memberikan informasi kepada *user* mengenai jumlah keseluruhan dataset yang dipakai sebagai data *training* dan jumlah keseluruhan data *reviewer* yang terdata pada sistem ini.



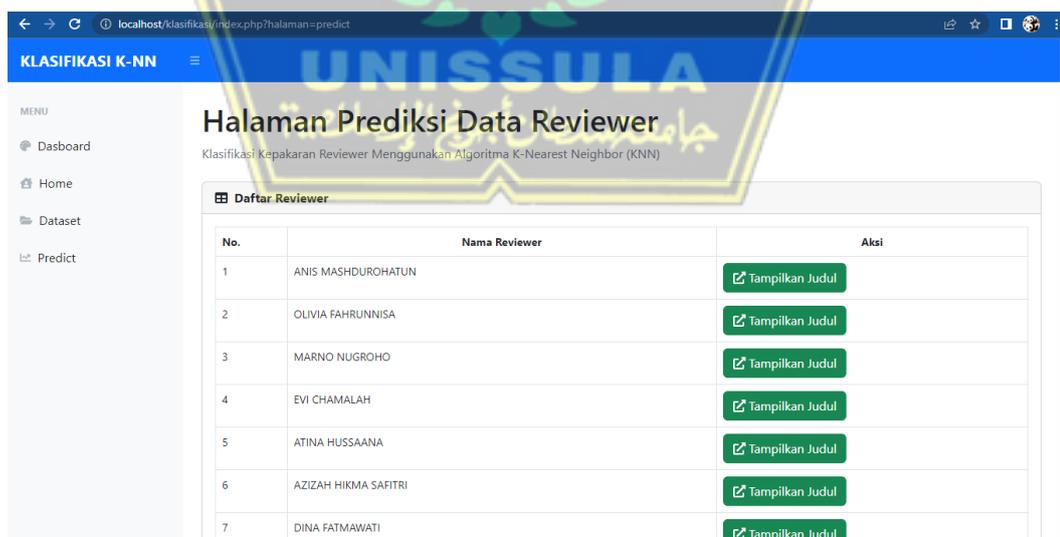
Gambar 4. 8 Tampilan halaman *home*

Gambar 4.8 merupakan hasil implementasi *user interface* dari menu *home*, halaman ini akan menampilkan deskripsi singkat mengenai penelitian tugas akhir serta memberikan informasi pada *user*. Dibuatnya menu *home* ini agar saat *user* masuk pada sistem dapat mengetahui secara singkat fungsi dibuatnya sistem klasifikasi *reviewer*.



Gambar 4. 9 Tampilan halaman dataset

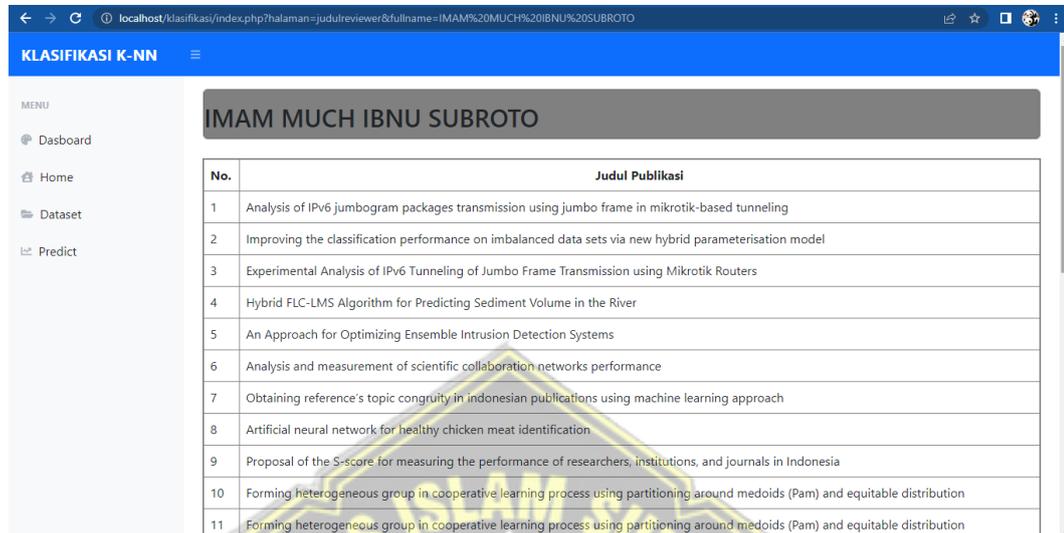
Gambar 4.9 merupakan hasil implementasi *user interface* dari menu dataset, halaman ini menampilkan keterangan mengenai asal perolehan data yang digunakan dalam penelitian. Data yang ditampilkan adalah data *training* yang merupakan judul jurnal terindeks scopus dari database SINTA. Pada halaman ini terdapat tabel dengan kolom judul yang berisi judul-judul jurnal serta kolom bidang ilmu yang berisi label bidang ilmu. Terdapat tombol *download* agar *user* dapat melihat keseluruhan dataset.



Gambar 4. 10 Tampilan halaman *predict*

Gambar 4.10 merupakan hasil implementasi *user interface* dari menu *predict* pada sistem klasifikasi. Halaman ini menampilkan daftar nama-nama *reviewer*,

terdapat 2 kolom untuk daftar nama dan untuk tombol aksi. Pada saat tombol “Tampilkan Judul” diklik akan muncul halaman *reviewer*.



No.	Judul Publikasi
1	Analysis of IPv6 jumbogram packages transmission using jumbo frame in mikrotik-based tunneling
2	Improving the classification performance on imbalanced data sets via new hybrid parameterisation model
3	Experimental Analysis of IPv6 Tunneling of Jumbo Frame Transmission using Mikrotik Routers
4	Hybrid FLC-LMS Algorithm for Predicting Sediment Volume in the River
5	An Approach for Optimizing Ensemble Intrusion Detection Systems
6	Analysis and measurement of scientific collaboration networks performance
7	Obtaining reference's topic congruity in Indonesian publications using machine learning approach
8	Artificial neural network for healthy chicken meat identification
9	Proposal of the S-score for measuring the performance of researchers, institutions, and journals in Indonesia
10	Forming heterogeneous group in cooperative learning process using partitioning around medoids (Pam) and equitable distribution
11	Forming heterogeneous group in cooperative learning process using partitioning around medoids (Pam) and equitable distribution

Gambar 4. 11 Tampilan halaman *reviewer* dan judul artikel (1)

Gambar 4.11 merupakan hasil implementasi *user interface* dari menu *predict* pada sistem klasifikasi, halaman ini merupakan halaman lanjutan menu *predict*. Pada tampilan halaman sebelumnya terdapat tombol di kolom aksi, jika di klik maka halaman nama *reviewer* dan judul artikel ini lah yang akan muncul. Halaman ini menampilkan nama *reviewer* beserta daftar judul artikel milik *reviewer*.

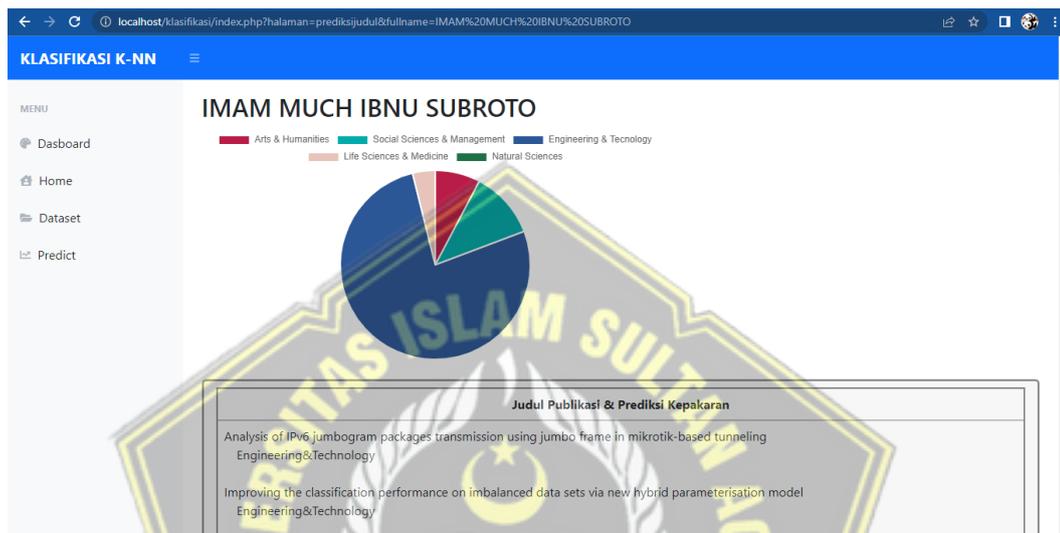


15	Designing Business System Model using System Modeling Approach to the Small and Medium Enterprises (SME) of Furniture in Indonesia
16	WhatsApp, viber and telegram: Which is the best for instant messaging?
17	A position controller model on color-based object tracking using fuzzy logic
18	Conveyor performance based on motor DC 12 volt Eg-530ad-2f using K-means clustering
19	Designing business system model using system modeling approach to the small and medium enterprises (SME) of furniture in Indonesia
20	Foreword from chair of EECSI 2016
21	Comparing performance of data mining algorithms in prediction heart diseases
22	Fortifying big data infrastructures to face security and privacy issues
23	Plagiarism detection through internet using hybrid artificial neural network and support vectors machine
24	The architecture of Indonesian publication index: A major Indonesian academic database
25	Arabic script web page language identification using hybrid-KNN method
26	Decision Support System using Weighting Similarity Model for Constructing Ground-Truth Dataset

Gambar 4. 12 Tampilan halaman *reviewer* dan judul artikel (2)

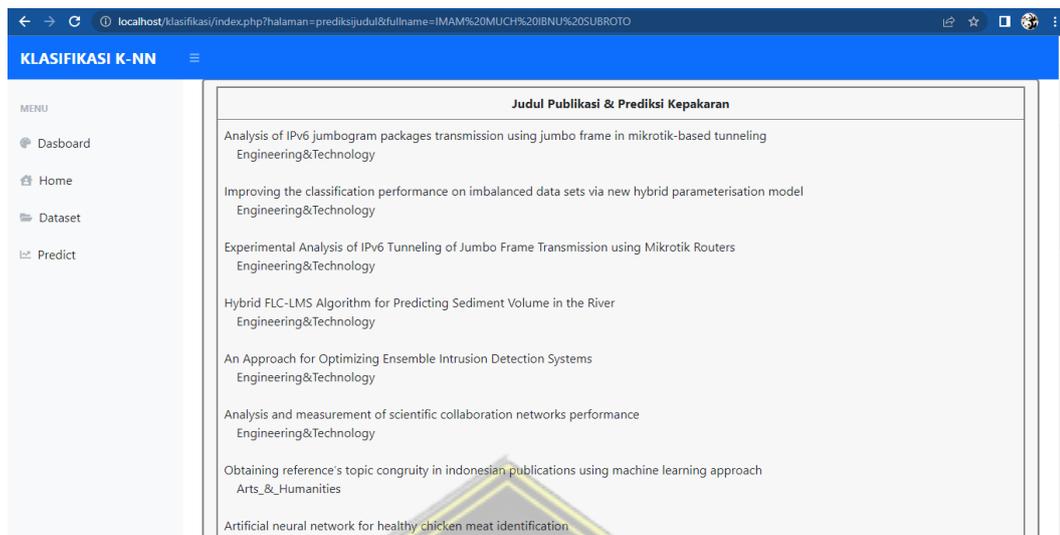
Gambar 4.12 masih merupakan tampilan dari halaman nama *reviewer* dan judul-judul artikel yang ditulis oleh *reviewer*, yang mana pada bagian bawah daftar

judul artikel terdapat 2 tombol yaitu tombol kembali yang berada disebelah kiri bawah dan tombol untuk prediksi kepakaran yang berada disebelah kanan bawah. Jika tombol prediksi kepakaran diklik akan menampilkan klasifikasi judul-judul arikel tersebut ke dalam 5 bidang ilmu yang telah ditentukan pada halaman yang berbeda lagi.



Gambar 4. 13 Tampilan halaman *reviewer* dan label kepakaran (1)

Gambar 4.13 merupakan hasil implementasi *user interface* dari menu *predict* pada sistem klasifikasi, halaman ini merupakan halaman lanjutan pada menu *predict*. Halaman prediksi merupakan halaman yang paling penting yaitu untuk menampilkan hasil prediksi. Pada halaman ini akan menampilkan sebuah diagram lingkaran (*pie chart*) untuk menunjukkan presentasi paling banyak dari hasil klasifikasi judul-jurnal artikel yang ditulis oleh *reviewer*, diagram ini bertujuan untuk mempermudah *user* dalam mengambil kesimpulan. Dengan begitu dapat diketahui seorang *reviewer* memiliki keahlian dalam salah satu bidang diantara 5 bidang ilmu yang telah ditentukan sebelumnya. Kemudian dibawah diagram tetap akan ditampilkan daftar judul-judul artikel seperti pada halaman sebelumnya, namun di halaman ini judul-judul artikel tersebut sudah diklasifikasin oleh sistem kedalam salah satu dari 5 bidang ilmu yang telah ditentukan.



Gambar 4. 14 Tampilan halaman *reviewer* dan label kepakaran (2)

Gambar 4.14 masih merupakan halaman yang menampilkan judul-judul artikel beserta klasifikasi bidang ilmunya ke dalam salah satu 5 bidang ilmu yang telah ditentukan.

4.3 Analisis Hasil Evaluasi

Pada tahap analisis hasil evaluasi bertujuan untuk mengetahui tingkat akurasi dari penerapan metode klasifikasi menggunakan algoritma KNN terhadap data *testing*. Adapun tabel rincian dari pembagian data *testing* dan data *training* menggunakan dataset scopus dengan beberapa skenario uji, ditunjukkan pada tabel 4.1.

Tabel 4. 1 Tabel skenario data uji (Hendriyanto & Sari, 2022)

Nama data	Split data	Jumlah data
Data <i>training</i> : data <i>testing</i>	70% : 30%	350 : 150
Data <i>training</i> : data <i>testing</i>	80% : 20%	400 : 100
Data <i>training</i> : data <i>testing</i>	90% : 10%	450 : 50

Evaluasi hasil akhir pada penelitian ini menggunakan *confusion matrix* terhadap hasil akurasi, presisi, *recall*, dan *f-measure* dengan melakukan berbagai percobaan penentuan jumlah tetangga terdekat (*k*) yaitu $K = 3, 5, 9, 20, 25, 30, 35$.

Tabel 4.2 sampai dengan 4.22 adalah tabel untuk *confusion matrix* dari setiap jumlah k dengan menggunakan 3 skenario split data.

a. Hasil perhitungan *confusion matrix* menggunakan split data 70% : 30%.

$K = 3$

Tabel 4. 2 Tabel *confusion matrix* $K = 3$ menggunakan split data 70% : 30%

	<i>Predicted Positive</i>	<i>Predicted Negatif</i>
<i>Actually Positive</i>	46 (TP)	104 (FN)
<i>Actually Negatif</i>	104 (FP)	184 (TN)

$K = 5$

Tabel 4. 3 Tabel *confusion matrix* $K = 5$ menggunakan split data 70% : 30%

	<i>Predicted Positive</i>	<i>Predicted Negatif</i>
<i>Actually Positive</i>	45 (TP)	105 (FN)
<i>Actually Negatif</i>	105 (FP)	180 (TN)

$K = 9$

Tabel 4. 4 Tabel *confusion matrix* $K = 9$ menggunakan split data 70% : 30%

	<i>Predicted Positive</i>	<i>Predicted Negatif</i>
<i>Actually Positive</i>	52 (TP)	98 (FN)
<i>Actually Negatif</i>	98 (FP)	208 (TN)

$K = 20$

Tabel 4. 5 Tabel *confusion matrix* $K = 20$ menggunakan split data 70% : 30%

	<i>Predicted Positive</i>	<i>Predicted Negatif</i>
<i>Actually Positive</i>	56 (TP)	94 (FN)
<i>Actually Negatif</i>	94 (FP)	224 (TN)

$K = 25$

Tabel 4. 6 Tabel *confusion matrix* $K = 25$ menggunakan split data 70% : 30%

	<i>Predicted Positive</i>	<i>Predicted Negatif</i>
<i>Actually Positive</i>	69 (TP)	81 (FN)
<i>Actually Negatif</i>	81 (FP)	276 (TN)

$K = 30$

Tabel 4. 7 Tabel *confusion matrix* $K = 30$ menggunakan split data 70% : 30%

	<i>Predicted Positive</i>	<i>Predicted Negatif</i>
<i>Actually Positive</i>	65 (TP)	85 (FN)
<i>Actually Negatif</i>	85 (FP)	260 (TN)

$K = 35$

Tabel 4. 8 Tabel *confusion matrix* $K = 35$ menggunakan split data 70% : 30%

	<i>Predicted Positive</i>	<i>Predicted Negatif</i>
<i>Actually Positive</i>	66 (TP)	84 (FN)
<i>Actually Negatif</i>	84 (FP)	264 (TN)

b. Hasil perhitungan *confusion matrix* menggunakan split data 80% : 20%.

$K = 3$

Tabel 4. 9 Tabel *confusion matrix* $K = 3$ menggunakan split data 80% : 20%

	<i>Predicted Positive</i>	<i>Predicted Negatif</i>
<i>Actually Positive</i>	34 (TP)	66 (FN)
<i>Actually Negatif</i>	66 (FP)	136 (TN)

$K = 5$

Tabel 4. 10 Tabel *confusion matrix* $K = 5$ menggunakan split data 80% : 20%

	<i>Predicted Positive</i>	<i>Predicted Negatif</i>
<i>Actually Positive</i>	29 (TP)	71 (FN)
<i>Actually Negatif</i>	71 (FP)	116 (TN)

$K = 9$

Tabel 4. 11 Tabel *confusion matrix* $K = 9$ menggunakan split data 80% : 20%

	<i>Predicted Positive</i>	<i>Predicted Negatif</i>
<i>Actually Positive</i>	36 (TP)	64 (FN)
<i>Actually Negatif</i>	64 (FP)	144 (TN)

K = 20

Tabel 4. 12 Tabel *confusion matrix* K = 20 menggunakan split data 80% : 20%

	<i>Predicted Positive</i>	<i>Predicted Negatif</i>
<i>Actually Positive</i>	48 (TP)	52 (FN)
<i>Actually Negatif</i>	52 (FP)	192 (TN)

K = 25

Tabel 4. 13 Tabel *confusion matrix* K = 25 menggunakan split data 80% : 20%

	<i>Predicted Positive</i>	<i>Predicted Negatif</i>
<i>Actually Positive</i>	49 (TP)	51 (FN)
<i>Actually Negatif</i>	51 (FP)	196 (TN)

K = 30

Tabel 4. 14 Tabel *confusion matrix* K = 30 menggunakan split data 80% : 20%

	<i>Predicted Positive</i>	<i>Predicted Negatif</i>
<i>Actually Positive</i>	54 (TP)	46 (FN)
<i>Actually Negatif</i>	46 (FP)	216 (TN)

K = 35

Tabel 4. 15 Tabel *confusion matrix* K = 35 menggunakan split data 80% : 20%

	<i>Predicted Positive</i>	<i>Predicted Negatif</i>
<i>Actually Positive</i>	37 (TP)	63 (FN)
<i>Actually Negatif</i>	63 (FP)	148 (TN)

c. Hasil perhitungan *confusion matrix* menggunakan split data 90% : 10%.

K = 3

Tabel 4. 16 Tabel *confusion matrix* K = 3 menggunakan split data 90% : 10%

	<i>Predicted Positive</i>	<i>Predicted Negatif</i>
<i>Actually Positive</i>	17 (TP)	33 (FN)
<i>Actually Negatif</i>	33 (FP)	68 (TN)

$K = 5$

Tabel 4. 17 Tabel *confusion matrix* $K = 5$ menggunakan split data 90% : 10%

	<i>Predicted Positive</i>	<i>Predicted Negatif</i>
<i>Actually Positive</i>	13 (TP)	37 (FN)
<i>Actually Negatif</i>	37 (FP)	52 (TN)

$K = 9$

Tabel 4. 18 Tabel *confusion matrix* $K = 9$ menggunakan split data 90% : 10%

	<i>Predicted Positive</i>	<i>Predicted Negatif</i>
<i>Actually Positive</i>	14 (TP)	36 (FN)
<i>Actually Negatif</i>	36 (FP)	56 (TN)

$K = 20$

Tabel 4. 19 Tabel *confusion matrix* $K = 20$ menggunakan split data 90% : 10%

	<i>Predicted Positive</i>	<i>Predicted Negatif</i>
<i>Actually Positive</i>	18 (TP)	32 (FN)
<i>Actually Negatif</i>	32 (FP)	72 (TN)

$K = 25$

Tabel 4. 20 Tabel *confusion matrix* $K = 25$ menggunakan split data 90% : 10%

	<i>Predicted Positive</i>	<i>Predicted Negatif</i>
<i>Actually Positive</i>	19 (TP)	31 (FN)
<i>Actually Negatif</i>	31 (FP)	76 (TN)

$K = 30$

Tabel 4. 21 Tabel *confusion matrix* $K = 30$ menggunakan split data 90% : 10%

	<i>Predicted Positive</i>	<i>Predicted Negatif</i>
<i>Actually Positive</i>	21 (TP)	29 (FN)
<i>Actually Negatif</i>	29 (FP)	84 (TN)

$K = 35$

Tabel 4. 22 Tabel *confusion matrix* $K = 35$ menggunakan split data 90% : 10%

	<i>Predicted Positive</i>	<i>Predicted Negatif</i>
<i>Actually Positive</i>	25 (TP)	25 (FN)
<i>Actually Negatif</i>	25 (FP)	100 (TN)

Setelah diperoleh data *confusion matrix* nya, maka akan diketahui hasil dari nilai *accuracy*, *Precision*, *recall*, dan *f-measure* dari beberapa percobaan jumlah k sebagai tetangga terdekatnya. Tabel untuk nilai *accuracy*, *Precision*, *recall*, dan *f-measure* dari setiap jumlah k dan menggunakan 3 skenario split data ditunjukkan pada tabel 4.23 sampai dengan 4.25.

a. Hasil perhitungan *Accuracy*, *Precision*, *Recall*, *F-Measure* menggunakan split data 70% : 30%.

Tabel 4. 23 Tabel *Accuracy*, *Precision*, *Recall*, *F-Measure* menggunakan split data 70% : 30%

Jumlah K	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
3	0.539	0.452	0.306	0.265
5	0.545	0.590	0.3	0.265
9	0.602	0.637	0.346	0.304
20	0.606	0.484	0.373	0.346
25	0.682	0.490	0.459	0.448
30	0.658	0.432	0.433	0.426
35	0.667	0.447	0.440	0.413

b. Hasil perhitungan *Accuracy*, *Precision*, *Recall*, *F-Measure* menggunakan split data 80% : 20%

Tabel 4. 24 Tabel *Accuracy*, *Precision*, *Recall*, *F-Measure* menggunakan split data 80% : 20%

Jumlah K	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
3	0.580	0.356	0.34	0.314
5	0.535	0.386	0.29	0.270
9	0.612	0.668	0.36	0.312
20	0.698	0.487	0.48	0.473

25	0.711	0.588	0.490	0.481
30	0.751	0.628	0.54	0.528
35	0.619	0.610	0.37	0.359

c. Hasil perhitungan *Accuracy*, *Precision*, *Recall*, *F-Measure* menggunakan split data 90% : 10%.

Tabel 4. 25 Tabel *Accuracy*, *Precision*, *Recall*, *F-Measure* menggunakan split data 90% : 10%

Jumlah K	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
3	0.584	0.484	0.339	0.368
5	0.496	0.382	0.260	0.230
9	0.544	0.531	0.28	0.267
20	0.597	0.517	0.36	0.322
25	0.612	0.384	0.38	0.354
30	0.652	0.464	0.420	0.410
35	0.716	0.514	0.5	0.483

Dari perhitungan *accuracy*, *Precision*, *recall*, dan *f-measure* menerapkan 3 skenario split data dengan percobaan beberapa nilai parameter K yang telah dicatat pada tabel-tabel diatas, maka dapat diketahui bahwa hasil akurasi tertinggi diperoleh pada split data 80% : 20% yaitu 0,751, *Precision* 0.628, *recall* 0.54, dan *f-measure* 0.528 pada percobaan menggunakan nilai K = 30.

4. 4 Perbandingan Performa Data *Training* dengan Data *Testing*

Pada tahap ini akan dilakukan pengamatan terhadap perbandingan performa antara data *training* dengan data *testing*, pengamatan ini dilakukan untuk mengetahui apakah pembelajaran data pada mesin *learning* mengalami *overfitting* atau tidak. Perbandingan performa data *training* dan data *testing* ditunjukkan pada tabel 4.26 sampai dengan 4.28.

a. Perbandingan performa dengan split data 70% : 30%

Tabel 4. 26 Perbandingan performa *training* dan *testing* dengan split data 70% : 30%

Data <i>Training</i>	Jumlah K	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
	3	0.539	0.452	0.306	0.265
	5	0.545	0.590	0.3	0.265
	9	0.602	0.637	0.346	0.304
	20	0.606	0.484	0.373	0.346
	25	0.682	0.490	0.459	0.448
	30	0.658	0.432	0.433	0.426
	35	0.667	0.447	0.440	0.413
Data <i>testing</i>	Jumlah K	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
	3	0.633	0.512	0.386	0.367
	5	0.630	0.55	0.373	0.343
	9	0.634	0.593	0.373	0.331
	20	0.686	0.649	0.453	0.459
	25	0.699	0.494	0.48	0.478
	30	0.735	0.571	0.52	0.521
	35	0.683	0.700	0.439	0.428

b. Perbandingan performa dengan split data 80% : 20%

Tabel 4. 27 Perbandingan performa *training* dan *testing* dengan split data 80% : 20%

Data <i>Training</i>	Jumlah K	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
	3	0.580	0.356	0.34	0.314
	5	0.535	0.386	0.29	0.270
	9	0.612	0.668	0.36	0.312
	20	0.698	0.487	0.48	0.473
	25	0.711	0.588	0.490	0.481
	30	0.751	0.628	0.54	0.528
	35	0.619	0.610	0.37	0.359
	Jumlah K	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>

Data testing	3	0.581	0.483	0.333	0.293
	5	0.541	0.496	0.293	0.299
	9	0.612	0.459	0.36	0.315
	20	0.743	0.562	0.533	0.515
	25	0.839	0.780	0.666	0.675
	30	0.763	0.773	0.546	0.551
	35	0.697	0.763	0.453	0.443

c. Perbandingan performa dengan split data 90% : 10%.

Tabel 4. 28 Perbandingan performa *training* dan *testing* dengan split data 90% : 10%

Data Training	Jumlah K	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
	3	0.584	0.484	0.339	0.368
	5	0.496	0.382	0.260	0.230
	9	0.544	0.531	0.28	0.267
	20	0.597	0.517	0.36	0.322
	25	0.612	0.384	0.38	0.354
	30	0.652	0.464	0.420	0.410
	35	0.716	0.514	0.5	0.483
Data testing	Jumlah K	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
	3	0.706	0.424	0.48	0.433
	5	0.666	0.456	0.426	0.405
	9	0.626	0.499	0.373	0.340
	20	0.673	0.579	0.439	0.449
	25	0.690	0.512	0.466	0.453
	30	0.868	0.775	0.72	0.730
	35	0.753	0.645	0.546	0.520

Dari perbandingan performa yang telah dilakukan terhadap data *training* dan data *testing* dapat disimpulkan bahwa data *training* mampu membentuk model *machine learning* dengan cukup baik, karena saat digunakan untuk melakukan

testing terhadap suatu data dapat menghasilkan akurasi yang tidak jauh berbeda yaitu sama atau lebih besar dari hasil performa yang dihasilkan oleh data *training*.

4.5 Hasil Pengujian Sistem

Pada tahap pengujian sistem klasifikasi yang dibuat dalam penelitian tugas akhir ini menggunakan metode *Black Box Testing*. Tujuan dari pengujian ini adalah untuk mengamati hasil dari *input* dan *output* dari sistem yang telah dibuat tanpa mengetahui tentang detail dari struktur kode pemrogramannya. Dan pengujian ini untuk meninjau dan mengamati apakah di dalam sistem terdapat kesalahan atau *bug* serta meninjau apakah sistem sudah berjalan sesuai dengan yang diharapkan. Gambar 4.26 adalah tabel dari hasil pengujian *black box testing* pada sistem klasifikasi kepakaran *reviewer*.

Tabel 4. 29 Tabel hasil pengujian sistem

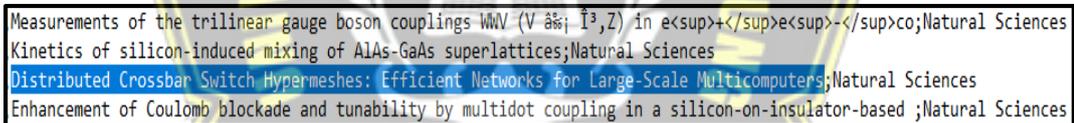
Skenario Pengujian	Kasus Pengujian	Hasil Pengujian	Keterangan
Menampilkan deskripsi tugas akhir	Menampilkan deskripsi tugas akhir pada sistem	Sesuai	Sukses
Menampilkan dataset	Menampilkan sampel dataset penelitian pada halaman dataset	Sesuai	Sukses
<i>Download</i> dataset	Melakukan <i>download</i> dataset pada halaman dataset	Sesuai	Sukses
Menampilkan daftar <i>reviewer</i>	Menampilkan daftar nama-nama <i>reviewer</i> pada sistem	Sesuai	Sukses
Menampilkan judul-judul artikel dari tiap <i>reviewer</i>	Menampilkan judul-judul artikel dari tiap <i>reviewer</i> pada saat tombol “Tampilkan Judul” diklik	Sesuai	Sukses

Prediksi kepakaran	Melakukan prediksi kepakaran <i>reviewer</i>	Sesuai	Sukses
Hasil prediksi	Menampilkan hasil prediksi kepakaran <i>reviewer</i> .	Sesuai	Sukses

4.6 Pembahasan Faktor Hasil Akurasi

Dalam penelitian ini telah dapat memberikan hasil akurasi sebesar 0,751 pada percobaan jumlah $K = 30$ menggunakan skenario uji 80% : 20%. Hasil akurasi tersebut cukup baik namun bisa ditingkatkan lagi untuk dapat melakukan klasifikasi data dengan tepat dan sesuai. Adapun beberapa alasan yang bisa menyebabkan hasil akurasi dalam penelitian ini, diantaranya yaitu :

1. Dataset yang digunakan kurang bersih, dikarenakan dalam 1 data terdapat judul yang sama namun memiliki kategori bidang ilmu yang berbeda.
2. Terdapat beberapa judul yang kategorinya belum sesuai.



Measurements of the trilinear gauge boson couplings WWV ($V = \gamma, Z$) in $e^+e^- \rightarrow e^+e^- \gamma \gamma$;Natural Sciences
 Kinetics of silicon-induced mixing of AlAs-GaAs superlattices;Natural Sciences
 Distributed Crossbar Switch Hypermeshes: Efficient Networks for Large-Scale Multicomputers;Natural Sciences
 Enhancement of Coulomb blockade and tunability by multidot coupling in a silicon-on-insulator-based ;Natural Sciences

Gambar 4. 15 Contoh data dengan kategori yang belum sesuai

Gambar 4.15 adalah salah satu contoh pengkategorian judul yang belum sesuai, karena *Computer Networks* termasuk dari bidang *Engineering & Technology*.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Adapun kesimpulan yang dapat diambil dari penelitian ini adalah sistem klasifikasi yang dibangun dengan menerapkan metode KNN memperoleh hasil akurasi paling tinggi setelah dilakukan beberapa percobaan jumlah tetangga terdekat $K = 30$ menggunakan skenario pengujian 80% : 20% yaitu sebesar 0,751 atau 75 %, serta diperoleh juga hasil *Precision* sebesar 0.628, *recall* sebesar 0.54, dan *f-measure* sebesar 0.528.

5.2 Saran

Adapun saran yang diberikan setelah melakukan penelitian dan pembuatan klasifikasi kepakaran *reviewer* ini adalah sebagai berikut :

1. Melakukan pembersihan dan pengolahan data dengan lebih teliti dan lebih baik lagi agar data benar-benar bersih dan siap untuk dilanjutkan ke tahap penerapan metode.
2. Menggunakan parameter nilai K yang lebih bervariasi lagi.
3. Penerapan metode klasifikasi yang benar agar memperoleh hasil akurasi sesuai dengan yang diharapkan, sehingga sistem dapat digunakan sebagaimana mestinya.
4. Pengembangan sistem dengan menambah fitur lainnya agar sistem menjadi lebih kompleks.

DAFTAR PUSTAKA

- Akromunnisa, K., & Hidayat, R. (2019). Klasifikasi Dokumen Tugas Akhir (Skripsi) Menggunakan K-Nearest Neighbor. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 4(1), 69. <https://doi.org/10.14421/jiska.2019.41-07>
- Atmaja, D. M. U., & Mandala, R. (2020). Analisa Judul Skripsi untuk Menentukan Peminatan Mahasiswa Menggunakan Vector Space Model dan Metode K-Nearest Neighbor. *IT for Society*, 4(2), 1–6. <https://doi.org/10.33021/itfs.v4i2.1182>
- DalPino, A. (2022). *Rubix ML Documentation*. Rubixml.Com. <https://rubixml.com/>
- Fansyuri, M. (2020). Analisa algoritma klasifikasi k-nearest neighbor dalam menentukan nilai akurasi terhadap kepuasan pelanggan (study kasus pt. Trigatra komunikatama). *Humanika: Jurnal Ilmu Sosial, Pendidikan, Dan Humaniora*, 3(1), 29–33.
- Firdaus, D. (2017). Penggunaan Data Mining dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer. *Jurnal Format*, 6(2), 91–97.
- Fitrianti, R. P., Kurniawati, A., & Agustien, D. (2019). Implementasi Algoritma K-Nearest Neighbor Terhadap Analisis Sentimen Review Restoran Dengan Teks Bahasa Indonesia. *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*.
- Hadi, H. P., & Sukanto, T. S. (2020). Klasifikasi Jenis Laporan Masyarakat Dengan K-Nearest Neighbor Algorithm. *JOINS (Journal of Information System)*, 5(1), 77–85. <https://doi.org/10.33633/joins.v5i1.3355>
- Haerul Jaman, Jajam; Fahlevi, S. A. (2019). Klasifikasi Calon Mahasiswa Bidikmisi dengan Algoritma K-Nearest Neighbor. *Proshiding Annual Research Seminar 2019*, 1–5.
- Hardiyanto, E., & Rahutomo, F. (2016). STUDI AWAL KLASIFIKASI ARTIKEL WIKIPEDIA BAHASA INDONESIA DENGAN MENGGUNAKAN METODA K NEAREST NEIGHBOR. *Seminar Nasional Terapan Riset Inovatif*, 01, 15–16.
- Hendriyanto, M. D., & Sari, B. N. (2022). Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Judul Berita Hoax. *Jurnal Ilmiah Informatika*, 10(02), 80–84. <https://ejournal.upbatam.ac.id/index.php/jif/article/view/5477>

- Irfa, A. A., Adiwijaya, A., & Mubarok, M. S. (2018). Klasifikasi Topik Berita Berbahasa Indonesia Menggunakan k-Nearest Neighbor. *EProceedings of Engineering*, 5(2).
- Kurniawan, Y. I., & Barokah, T. I. (2020). Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan K-Nearest Neighbor. *Jurnal Ilmiah Matrik*, 22(1), 73–82. <https://doi.org/10.33557/jurnalmatrik.v22i1.843>
- Limbong, J. J. A., Sembiring, I., Hartomo, K. D., Kristen, U., Wacana, S., & Korespondensi, P. (2022). ANALISIS KLASIFIKASI SENTIMEN ULASAN PADA E-COMMERCE SHOPEE BERBASIS WORD CLOUD DENGAN METODE NAIVE BAYES DAN K-NEAREST ANALYSIS OF REVIEW SENTIMENT CLASSIFICATION ON E-COMMERCE SHOPEE WORD CLOUD BASED WITH NAÏVE BAYES AND K-NEAREST NEIGHBOR METHODS. 9(2), 347–356. <https://doi.org/10.25126/jtiik.202294960>
- Maulana, F., Orisa, M., & Zulfia Zahro', H. (2021). Klasifikasi Data Produk Mebel Aneka Jaya Menggunakan Metode K-Nearest Neighbor Berbasis Web. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 5(2), 460–466. <https://doi.org/10.36040/jati.v5i2.3782>
- Nugroho, A., & Religia, Y. (2021). Analisis Optimasi Algoritma Klasifikasi Naive Bayes menggunakan Genetic Algorithm dan Bagging. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(3), 504–510. <https://doi.org/10.29207/resti.v5i3.3067>
- Putra, M. Y., & Putri, D. I. (2022). Pemanfaatan Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Kelas XI. *Jurnal Tekno Kompak*, 16(2), 176–187.
- Saadah, P., Nurdiawan, O., Kurnia, D. A., Amalia, D. R., & Ali, I. (2021). Klasifikasi Pemberian Sanksi Pajak Dengan Metode K-Nearest Neighbour Di KPP Pratama Cirebon Dua. *JURNAL DATA SCIENCE & INFORMATIKA (JDSI)*, 1(2), 41–45.
- Sandi, D., Utami, E., & Fatkhurohman, A. (2022). Klasifikasi Opini Dengan Menggunakan Algoritma K- Nearest Neighbor Pada Berita Vaksinasi Di Twitter. *JURNAL NUANSA INFORMATIKA*, 16.

- Setiawan, R., & Triayudi, A. (2022). Klasifikasi Status Gizi Balita Menggunakan Naïve Bayes dan K-Nearest Neighbor Berbasis Web. *Jurnal Media Informatika Budidarma*, 6(2), 777. <https://doi.org/10.30865/mib.v6i2.3566>
- Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner. *JBASE - Journal of Business and Audit Information Systems*, 4(1), 13–21. <https://doi.org/10.30813/jbase.v4i1.2729>
- Supriatman, A. (2021). Pembobotan TF-IDF pada Judul Penelitian Dosen Sebagai Dasar Klasifikasi Menggunakan Algoritma K-NN (Studi Kasus : Universitas Siliwangi). *VI*(1), 1573–1579.

