

**KLASIFIKASI BIDANG ILMU PADA PUBLIKASI TERINDEKS
GOOGLE SCHOLAR MENGGUNAKAN METODE K- NEAREST
NEIGHBOR**

LAPORAN TUGAS AKHIR

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



DISUSUN OLEH:

RIJI IRAWAN

NIM 32601800021

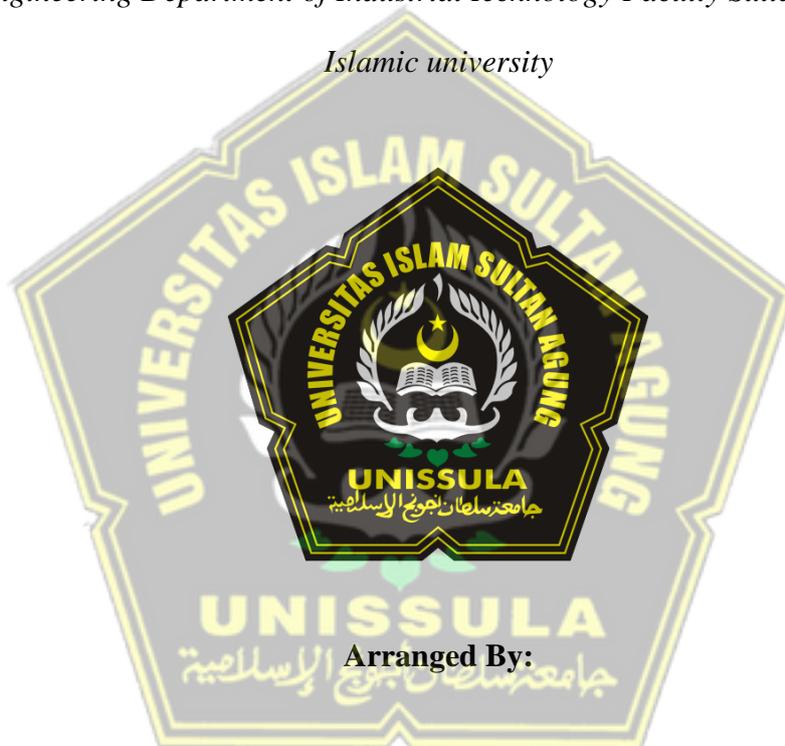
**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM SULTAN AGUNG
SEMARANG
JULI 2023**

FINAL PROJECT

**CLASSIFICATION OF FIELDS OF SCIENCE IN GOOGLE SCHOLAR
INDEXD PUBLICATIONS USING THE K-NEAREST NEIGHBOR
METHOD**

*Proposed to complete the requirement to obtain a bachelor's (S1) at informatics
Engineering Department of Industrial Iechnology Faculty Sultan Agung*

Islamic university



Arranged By:

RIJI IRAWAN

NIM 32601800021

**MAJORING OF INFORMATICS ENGINEERING
INDUSTRIAL TECHNOLOGY FACULTY
SULTAN AGUNG ISLAMIC UNIVERSITY
SEMARANG
JULI 2023**

LEMBAR PENGESAHAN PEMBIMBING

Laporan Tugas Akhir dengan judul “**Klasifikasi Bidang Ilmu pada Publikasi Terindeks Google Scholar Menggunakan Metode K-Nearest Neighbor**” ini disusun oleh :

Nama : Riji Irawan

NIM : 32601800021

Program Studi : Teknik Informatika

Telah disahkan oleh dosen pembimbing pada :

Hari : Senin

Tanggal : 24 Juli 2023

Mengesahkan,

Pembimbing I



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

Pembimbing II



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

Mengetahui,

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



Ir. Sri Mulyono, M.Eng
NIDN. 0626066601

LEMBAR PENGESAHAN PENGUJI

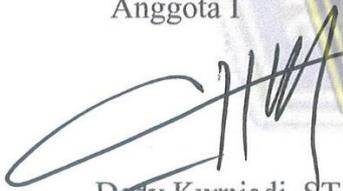
Laporan tugas akhir dengan judul “Klasifikasi Bidang Ilmu pada Publikasi Terindeks Google Scholar Menggunakan Metode K-Nearest Neighbor” ini telah dipertahankan di depan tim penguji Tugas Akhir pada :

Hari : Selasa

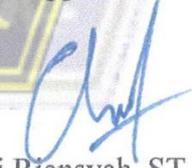
Tanggal : 25 Juli 2023

TIM PENGUJI

Anggota I


Dedy Kurniadi, ST,M.Kom.
NIDN. 0622588002

Anggota II


Andi Riansyah, ST, M.Kom.
NIDN.0609108802

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Riji Irawan

NIM : 32601800021

Judul Tugas Akhir :Klasifikasi Bidang Ilmu Pada Publikasi Terindeks Google Scholar menggunakan Metode K-Nerest

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, Juli 2023

Yang Menyatakan,



Riji Irawan

SURAT PERSETUJUAN UNGGAH KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Riji Irawan
NIM : 32601800021
Program Studi : Teknik Informatika
Fakultas : Teknologi industri

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul :

Klasifikasi Bidang Ilmu pada Publikasi Terindeks Google Scholar Menggunakan Metode K-Nearest Neighbor 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, Juli 2023

Yang menyatakan,



Riji Irawan

KATA PENGANTAR

Dengan mengucapkan syukur alhamdulillah atas kehadiran Allah SWT yang telah memberikan rahmat dan karunianya kepada penulis, sehingga dapat menyelesaikan Tugas Akhir dengan judul “klasifikasi Bidang Ilmu Pada Publikasi Terindeks Google Scholar Menggunakan Metode K-Nearest Neighbor” 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 disusun dan dibuat dengan adanya bantuan dari berbagai pihak, materi maupun teknis, oleh karena itu saya selaku penulis mengucapkan trimakasih kepada:

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

Semarang, Juli 2023



Riji Irawan

DAFTAR ISI

LAPORAN TUGAS AKHIR	i
FINAL PROJECT	ii
LEMBAR PENGESAHAN PEMBIMBING	iii
LEMBAR PENGESAHAN PENGUJI	iv
SURAT PERNYATAAN KEASLIAN TUGAS AKHIR	v
SURAT PERSETUJUAN UNGGAH KARYA ILMIAH	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
ABSTRAK	xiii
BAB I PENDEHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah	2
1.4. Tujuan Tugas Akhir	2
1.5. Sistematik Penulis	2
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI	4
2.1 Tinjauan Pustaka	4
2.2 Dasar Teori	8
2.2.1 Data Mining	8
2.2.2 KNN	9
2.2.3 Klasifikasi	10

2.2.4	PHP	10
2.2.5	VSCode	10
2.2.6	SINTA	11
2.2.7	Google scholar	11
2.2.8	Evaluasi Mesin Klasifikasi.....	12
BAB III METODE PENELITIAN.....		13
3.1	Tahap penelitian	13
3.1.1	Koleksi data.....	15
3.1.2	Preprocessing	15
3.1.3	Penerapan Metode.....	17
3.1.4	Evaluasi.....	18
3.2	Metode Perancangan Alur Sistem	18
3.2.1	Analisis Kebutuhan	18
3.2.2	Implementasi Sistem	19
3.2.3	Rancangan Alur Sistem.....	20
3.2.4	Perancangan Interface	23
BAB IV HASIL DAN ANALISIS PENELITIAN.....		26
4.1	Implementasi <i>User Interface</i>	26
4.2	Analisa Dan Pengujian	29
4.3	Analisa Akurasi	30
4.4	Pembahasan Hal yang Mempengaruhi Hasil Akurasi	33
BAB V KESIMPULAN DAN SARAN.....		36
5.1	Kesimpulan.....	36
5.2	Saran.....	36
DAFTAR PUSTAKA		37

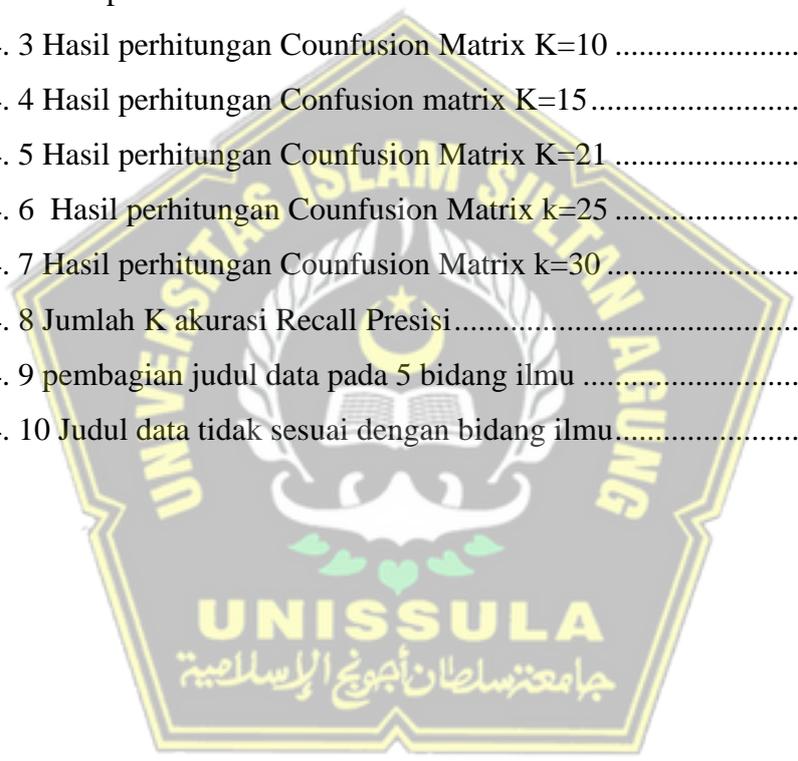


DAFTAR GAMBAR

Gambar 1. 1 Ilustrasi Klasifikasi KNN	9
Gambar 3. 1 Skema proses tahapan penelitian	13
Gambar 3. 2 Tahapan preprocessing	14
Gambar 3. 3 Proses penerapan metode di dalam coding	17
Gambar 3. 4 Proses evaluasi dengan ConfusionMatrix di dalam coding	18
Gambar 3. 5 Alur perancangan sistem	21
Gambar 3. 6 Flowchart sistem	22
Gambar 3. 7 Gambar halaman menu.....	23
Gambar 3. 8 Halaman home	23
Gambar 3. 9 Halaman dataset	24
Gambar 3. 10 Halaman predict	24
Gambar 3. 11 Halaman hasil predict.....	25
Gambar 4. 1 Halaman awal UI.....	26
Gambar 4. 2 Rancangan halaman home UI	27
Gambar 4. 3 Rancangan halaman UI dataset	27
Gambar 4. 4 Rancangan Halaman UI predict	28
Gambar 4. 5 halaman hasil predict UI	28

DAFTAR TABEL

Tabel 3. 1 Contoh hasil proses Case Folding pada kolom artikel.....	15
Tabel 3. 2 Contoh hasil Tokenezing	16
Tabel 3. 3 Contoh hasil Stopword.....	16
Tabel 3. 4 Contoh hasil Stemming.....	17
Tabel 4. 1 Pengujian alur pada sistem.....	29
Tabel 4. 2 Tahapan analisa akurasi	30
Tabel 4. 3 Hasil perhitungan Counfusion Matrix K=10	30
Tabel 4. 4 Hasil perhitungan Confusion matrix K=15.....	31
Tabel 4. 5 Hasil perhitungan Counfusion Matrix K=21	31
Tabel 4. 6 Hasil perhitungan Counfusion Matrix k=25	31
Tabel 4. 7 Hasil perhitungan Counfusion Matrix k=30.....	32
Tabel 4. 8 Jumlah K akurasi Recall Presisi.....	32
Tabel 4. 9 pembagian judul data pada 5 bidang ilmu	33
Tabel 4. 10 Judul data tidak sesuai dengan bidang ilmu.....	34



ABSTRAK

SINTA (*science and Technology indexs*) adalah sebuah platfrom layanan informasi yang di luncurkan pada tahun 2016, SINTA didirikan oleh kementerian riset teknologi dan Pendidikan tinggi republik Indonesia untuk mengevaluasi atau memantau hasil kinerja penelitian dan pengabdian masyarakat pada perguruan tinggi riset di Indonesia. SINTA mengindeks beberapa basis data seperti *Google scholar*, *DOAJ*, *Crossref* dan *Scopus*, dari beberapa basis data yang diindeks belum ada pembagian atau *klasifikasi* terhadap bidang ilmu tersebut, untuk itu perlu adanya *klasifikasi* otomatis pada semua judul data, kemudian dilakukan *preprocessing* bobot pada sebuah judul data dengan menggunakan Trem *FrekuensiInvers*. Data *Frekuensi* menentukan kesamaan dari data *training* dan data *testing* dengan *cosinus similrity*. berikutnya menerapkan sebuah *klasifikasi* yang bertujuan untuk menghasilkan data apakah *positive* atau *negative*. Proses *klasifikasi* yang di terapkan pada penelitian ini akan mengklasifikasikan setiap artikel menjadi lima bidang ilmu dengan menggunakan metode KNN. KNN merupakan sebuah metode memprediksi sebuah data baru didasarkan pada label-label pada data yang terdekat dengan data tersebut. Tujuan pada penelitian kali ini mengimplementasikan sebuah *system* yang bisa *mengklasifikasikan* jurnal ilmiah yang belum berdasarkan lima bidang ilmu. Peneliti melakukan beberapa kali percobaan menggunakan K yang telah ditentukan dan K=25 memberikan hasil terbaik dari beberapa kali percobaan tersebut dengan tingkat akurasi 0.5378.

Kata kunci: Klasifikasi, 5 Bidang Ilmu, *web Of Science*, *K-Nearest Neighbor*, SINTA

ABSTRACT

SINTA (*Science and Technology Index*) is an information service platform that was launched in 2016. It was established by the Ministry of Research, Technology, and Higher Education of the Republic of Indonesia to evaluate and monitor the research and community service performance of research universities in Indonesia. SINTA indexes several databases such as *Google Scholar*, *DOAJ*, *Crossref*, and *Scopus*. However, these databases do not have a predefined division or classification of scientific disciplines. Therefore, there is a need for automatic classification of all data titles, followed by preprocessing and weighting of the titles using the Term Frequency-Inverse Document Frequency (TF-IDF) method to determine the similarity between training and testing data using cosine similarity. Subsequently, a classification method is applied to determine whether the data is positive or negative. In this research, the classification process aims to classify each article into five scientific fields using the K-nearest neighbors (KNN) method. KNN is a method that predicts a new data point based on the labels of the nearest neighboring data points. The objective of this research is to implement a system that can classify scientific journals into five fields of study. The researchers conducted several experiments using a predefined value of K, and K=25 yielded the best results with an accuracy rate of 0.5378.

Key words: classification, 5 Fields of Science, *web Of Science*, *K-Nearest Neighbor*, SINTA

BAB I

PENDEHULUAN

1.1. Latar Belakang

Google Scholar adalah media tempat melakukan pencarian artikel-artikel ilmiah merupakan dalam bentuk kata yang telah dipublikasikan 21 tahun yang lalu, *google scholar* terindeks mencakup dokumen-dokumen yang diperoleh dari publikasi ilmiah. *Google scholar* mempermudah pengguna untuk pencarian literatur akademik secara menyeluruh, pengguna bisa melakukan pencarian jurnal-jurnal diseluruh bidang ilmu akademik seperti makalah, jurnal, dan teks ilmiah lainnya yang telah di unggah oleh penerbit karangan ilmiah tersebut. *Google scholar* sangat berkontribusi bagi penggunanya, menjadikan wadah untuk mencari referensi menyelesaikan penelitian pada tugas akhir mahasiswa maupun mahasiswi, dan jurnal yang sesuai pencarian akan muncul di halaman pertama. (Fadhilaturrahmi dkk., 2020)

Dapat kita ketahui bahwa *Google scholar* sumber pencarian teks ilmiah yang lengkap dan tidak sama dengan sumber-sumber lainnya yang kurang cukup memadai. *Google scholar* banyak sekali sumber yang bisa dikutif oleh penggunanya untuk mencari karangan ilmiah yang lagi booming pada saat ini, dan sering kali digunakan menjadi sumber pembelajaran bagi mahasiswa yang sedang melakukan penelitian tugas akhir. *Google scholar* sangat mudah untuk diakses bagi para penggunanya dan akan menampilkan karangan ilmiah sesuai judul pencarian dan menampilkan jurnal sesuai kebutuhan (Rachman, 2018)

Google scholar belum dikategorikan berdasarkan 5 bidang karena tidak ada pembersihan atau pengindeksan manual yang dilakukan dan hanya bergantung pada pengindeksan otomatis. Masih banyak data *google scholar* yang tidak sinkron dan perlu untuk diperbaiki.

Metode KNN dapat didefinisikan sebagai algoritma *machine learning* yang bertujuan untuk mengklasifikasi sebuah data pada jurnal. Algoritma KNN salah satu merupakan algoritma yang sederhana yang bisa digunakan untuk mengolah sebuah

data. Meskipun algoritma KNN merupakan algoritma yang sederhana, tetapi algoritma tersebut memiliki kelebihan dibandingkan dengan algoritma lainnya, dan dapat mengklasifikasi sebuah data secara menyeluruh.(Admojo & Ahsanawati, 2020)

Berdasarkan latar belakang di atas, penelitian ini mengusulkan metode KNN untuk *mengklasifikasi* bidang ilmu pada publikasi terindeks *google scholar*. Penelitian ini mengambil penelitian kasus belum dikategorikan berdasarkan 5 bidang ilmu di dalam *google scholar* .

1.2. Rumusan Masalah

berikut dari permasalahan yang melatar belakangi pada penelitian ini telah didapatkan rumusan masalahnya yaitu adalah bagaimana cara mengimplementasikan metode KNN pada *klasifikasi* bidang ilmu berdasarkan judul artikel menjadi 5 bidang ilmu.

1.3. Batasan Masalah

Berikut dibawah memiliki beberapa pembatasan masalah pada penelitian tugas akhir ini:

- a) *Klasifikasi* data akan dilakukan pembagian di dalam *google scholar* yang terindeks di web sinta.
- b) Penelitian ini menggunakan data yang didapatkan dari judul publikasi.

1.4. Tujuan Tugas Akhir

Tugas akhir berikut ini bertujuan untuk mendapatkan hasil terindeksi data yang ada di dalam *google scholar* dan menerapkan algoritma K-NN sebagai metode *klasifikasinya*.

1.5. Sistematik Penulis

Sistem pada penulisan penelitian kali ini untuk memenuhi atau proses menyelesaikan membuat rancangan tulisan pada tugas akhir yaitu berikut dibawah ini:

BAB 1: PENDEHULUAN

pada bab 1 peneliti menulis sebuah pendahuluan yang bertujuan untuk menjelaskan penjudulan, melatar belakngi pada tugas akhir ini, perumusan masalahnya, pembatasan masalahnya, tujuan dari tugas akhir ini, dan sistem penulisannya.

BAB 2: TINJAUAN PUSTAKA DAN DASAR TEORI

Selanjutnya pada bab 2 ini penulis memaparkan penelitiann yang sudah pernah di buat dan dasar teori yang berfungsi untuk mempermudah penulis agar paham dengan alur metode yang akan di terapkan dengan menggunakan metode K-Nearest neighbor.

BAB 3: METODE PENELITIAN

Berikutnya pada bab 3 ini mejelaskan alur metode pada peletian dan memiliki tahapan yaitu, koleksi data, preprocessing, penerapan metode, dan evaluasi.

BAB 4: HASIL PENELITIAN

Selanjutnya pada bab 4 peneliti menuliskan atau memaparkan tahapan hasil penelitian kali ini dengan mengaplikasikan system pada data yang ada di dalam google scholar

BAB 5: KESIMPULAN DAN SARAN

Pada bab 5 peneliti meringkas hasil kesimpulan pada tugas akhir kali ini dengan menjelaskan dari bab 1 samapai bab 5 dan saran dari peneliti untuk peliti selanjutnya yang ingin mengembangkan tahapan selanjutnya.

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Dalam penelitian tugas akhir kali ini diperlukan bahan tinjauan pustaka yang digunakan sebagai bahan acuan referensi dan pertimbangan, peneliti melakukan pencarian referensi pada penelitian sebelumnya yang terhubung terhadap judul penelitian ini, sebagai berikut:

Pada penelitian sebelumnya tahun 2019 Sebastian melakukan penelitian untuk membuat sistem yang dapat mengklasifikasi sebuah produk *E-marketplace* menggunakan metode KNN. TF-IDF metode yang di gunakan untuk pembobotan, kesamaan untuk menghitung jarak kesamaan produk dan metode K-NN. Berdasarkan pendapatan pengujian 150 data barang KNN menggunakan K=5 mengklasifikasi 146 data dan 4 data diklasifikasikan ke dalam golongan yang salah. Nilai K=5 memberikan hasil terbaik untuk kasus ini dengan akurasi 97,33%. Selanjutnya peneliti menguji menggunakan 150 dengan 5 data yang diklasifikasikan ke dalam kelas yang salah, dan menghasilkan akurasi 96,67%. Kesimpulannya penelitian ini memberikan hasil yang baik berdasarkan pada percobaan(Sebastian, 2019)

Dalam penelitian sebelumnya oleh Athallah pada tahun 2020 melakukan penelitian dengan menggunakan algoritma K-NN. Penelitian ini bertujuan mengklasifikasi data pada siswa sekolah SMKN 1 Balik Papan untuk memenuhi syarat penelitian tugas akhir mahasiswa, di mana peneliti ini akan menggunakan metode K-NN, peneliti menggunakan pemrograman PHP dan VCD. Untuk selanjutnya peneliti melakukan percobaan dengan 4 data lalu pakai 30 data sesuai dengan bidang ilmu masing-masing. Selanjutnya menggunakan perhitungan 250 data coba dan 100 latih. untuk hasil dengan percobaan klasifikasi pada penelitian ini, data siswa SMKN 1 balik papan dengan hasil yang baik dengan acurasi 71.65% sesuai dengan apa yang diharpkan. Lalu selanjutnya pada nilai K=20 berikut dengan

menggunakan nilai $K=25$ data bidang penguasaan pembelajaran. (Athaillah dkk., 2020).

Pada penelitian sebelumnya oleh akromunnisa & hidayat tahun 2019 melakukan penelitian tentang klasifikasi dokumen tugas akhir, tujuan dari penelitian kali ini adalah mengkalsifikasi sebuah dokumen abstark tugas akhir mahasiswa agama islam dengan mnggunakan metode yang di terapkan adalah K-NN. Berdasarkan perhitungan total perdataan yang akan di gunakan 60 dokumen mata kuliah kewarganegaraan, menggunakan 445 dokumen matkul agama islam dan 503 data judulnya . Berikut telah dilakakukan pembagian antara data training dan data testing. Data testing melakukan klasifikasi secara seponan dengan kelas yang sudah dirancang. Berikut dari percobaan mengklasifikasi matkul kewarganegaraan telah medapatkan hasil accuracy yang cukup tinggi dengan hasil 9:2 yaitu 99,0% lalu percobaan dengan menggunakan 8:3 yaitu 91,0% .(Akromunnisa & Hidayat, 2019)

Dalam penelitian sebelumnya oleh Homipage pada tahun 2021 Pengadaan dana mahasiswa akan dilakukan klasifikasi pada data yang sudah di tentukan sebelumnya (data traning) dan mengamibl data calon mahasiswa yang telah mendapatkan dana mahasiswa (data testing). Pada penelitian berikut bertujuan untuk mempermudah untuk menyeleksi calon dana mahasiswa menggunakan metode K-NN agar seleksi dana mahasiswa tepat sasaran kepada calon mahasiswa. Menggunakan metode K-NN telah terbukti menghasilkan akurasi yang baik dalam mengklasifikasi sebuah data. Dari beberapa kali percobaan dengan menggunakan 40 data dari 100 yang telah dikelasifikasi,. Lalu diuji deng menggunakan *confusion matrixs* telah menghasilkan uji coba sebesar 90.6% . berikut telah memaparkan bahwasannya menggunakan metode K-NN bisa dilakukan klasifikasi untuk seleksi calon dana mahasiswa.(Homepage dkk., 2021)

Pada penelitian sebelumnya oleh yadi saputra tahun 2018 siswa multimedia mengalami tidak stabilnya kelulusan, bertambahnya siswa baru menyebabkan banyaknya data mahasiswa yang menumpuk, prediksi yang dapat diolah sebagai pendukung program studi yang mengarahkan siswa selesai sekolah tepat waktu.

Pada tugas akhir ini prediksi kelulusann siswa dengan metode KNN. Peneliti akan mengambil sampel data semester V SMK 1 Sukamara menggunakan 20 data latih dan 2 data uji dengan menggunakan $k=5$ yang akan di-implementasikan dengan metode KNN untuk menentukan kelulusan siswa.(Yandi Saputra & Primadasa, 2018)

Dalam penelitian sebelumnya oleh ding tahun 2022 Dengan munculnya berbagai serangan jaringan yang terus-menerus, semakin penting untuk memastikan keamanan jaringan. Deteksi intrusi, sebagai salah satu teknologi penting untuk menjamin keamanan jaringan, telah banyak diteliti. Namun ketidakseimbangan kelas menyebabkan masalah yang menantang, yaitu data normal jauh lebih banyak daripada data serangan. Dalam menghadapi data yang tidak seimbang, bagaimana membuat model klasifikasi bekerja secara lebih efektif disebut sebagai masalah pembelajaran yang tidak seimbang. Dalam penelitian ini, kami mengusulkan metode pengambilan sampel data tabular untuk menyelesaikan masalah pembelajaran yang tidak seimbang, yang bertujuan untuk menyeimbangkan sampel normal dan sampel serangan. Pertama, untuk sampel normal, dengan meminimalkan kehilangan informasi sampel, kami menggunakan metode K-nearest neighbor untuk undersampling yang efektif. Kemudian kami merancang model generative adversarial networks klasifikasi bantu tabular (TACGAN) untuk oversampling sampel serangan.Kami telah melakukan percobaan verifikasi pada tiga set data deteksi intrusi yang nyata. Hasil eksperimen menunjukkan bahwa metode yang diusulkan mencapai hasil yang sangat baik dalam Akurasi, F1, AUC, dan Recall.(Ding dkk., 2022)

Dalam penelitian sebelumnya oleh nur ghaniviyanto tahun 2021 Menurut hasil survei, terdapat banyak judul berita dari berbagai topik yang tersebar di internet. Hal ini tentu membuat pembaca berita kesulitan ketika ingin mencari topik berita yang diinginkan untuk dibaca. Masalah ini dapat diatasi dengan pengelompokan atau yang disebut klasifikasi. Proses klasifikasi dilakukan tentunya dengan menggunakan proses komputerisasi. Penelitian ini bertujuan untuk mengklasifikasikan beberapa topik berita dalam bahasa Indonesia menggunakan

model klasifikasi KNN untuk mengonversi kata-kata menjadi vektor yang bertujuan untuk memudahkan proses klasifikasi. Penggunaan KNN dalam penelitian ini juga menentukan nilai K yang optimal yang akan digunakan. Selain menggunakan model klasifikasi. Hasil yang diperoleh menggunakan model KNN memiliki akurasi sebesar 89,2% dengan nilai $K=7$. Model word2vec dan KNN juga unggul dibandingkan dengan model klasifikasi support vector machine, logistic regression, dan random forest. (Nur Ghaniaviyanto Ramadhan, 2021)

Dalam penelitian sebelumnya oleh chen tahun 2022 Perkembangan teknologi web semantik mendukung perkembangan yang berkelanjutan dari data terhubung dan aplikasinya. Untuk dapat memanfaatkan secara efektif data terhubung yang semakin berkembang di web, data dari berbagai sumber perlu digabungkan, yang merupakan langkah kunci dalam analisis dan manajemen data berskala besar dari multi-sumber. Saat ini, penggabungan data ini dapat sulit dilakukan karena berbagai sumber biasanya menyediakan deskripsi konflik yang berbeda untuk entitas di dunia nyata yang sama. Untuk menyelesaikan penggabungan ini, kita perlu memecahkan masalah konflik objek. Penelitian ini mengusulkan metode penggabungan data terhubung berdasarkan perhitungan kesamaan dan k-nearest neighbor. Metode ini memiliki dua kontribusi. Pertama, metode perhitungan kesamaan data terhubung yang diusulkan dapat secara efektif menggabungkan node URI dan node blank dalam data terhubung. Kedua, diusulkan strategi penggabungan node tipe literal berdasarkan metode k-nearest neighbor classification, yang mewujudkan penggabungan otomatis dan memiliki kemandirian dari sumber data. Hasil penelitian menunjukkan bahwa dibandingkan dengan metode lainnya, metode ini dapat meningkatkan kekompakan, konsistensi, dan ketepatan hingga 12,9%, 30,6%, dan 12%. (Chen, 2022)

Dalam penelitian sebelumnya oleh ribadas tahun 2022 tentang Dalam penelitian ini, kami memperkenalkan pendekatan pembelajaran malas (lazy learning) multi-label untuk mengatasi indeks semantik otomatis pada koleksi dokumen besar dengan adanya kosakata label yang kompleks dan terstruktur dengan korelasi antar label yang tinggi. Metode yang diusulkan adalah

perkembangan dari algoritma k-Nearest Neighbors yang tradisional, yang menggunakan autoencoder besar yang dilatih untuk memetakan ruang label yang besar ke ruang laten ukuran yang lebih kecil, serta meregenerasi label yang diprediksi dari ruang laten ini. Dalam penelitian ini, kami mengusulkan dan mengevaluasi beberapa pendekatan representasi dokumen dan konfigurasi autoencoder label yang berbeda. (Ribadas-Pena dkk., 2022)

Pada penelitian sebelumnya oleh Swetha pada tahun 2022 tentang mengevaluasi keefektifan dan akurasi dari dua pendekatan yang berbeda, yaitu algoritma K-Nearest Neighbors (KNN) dan Novel Linear Regression (LR), dalam sistem prediksi harga rumah. Penelitian ini mengeksplorasi ketepatan dan akurasi metode-metode ini dalam memprediksi harga rumah. Algoritma KNN dievaluasi dengan memvariasikan parameter KNN, sedangkan teknik Novel Linear Regression dinilai menggunakan ukuran sampel sebesar 20. Hasil penelitian menunjukkan bahwa KNN mengungguli Novel Linear Regression dalam hal akurasi klasifikasi, mencapai akurasi sebesar 67,92% dibandingkan dengan 53,46% untuk Novel Linear Regression. Sebagai kesimpulan, algoritma K-Nearest Neighbors menghasilkan hasil prediksi yang lebih baik dalam pemantauan harga rumah dan meningkatkan akurasi prediksi harga dibandingkan dengan Novel Linear Regression. (Swetha & Dhanalakshmi, 2022)

2.2 Dasar Teori

2.2.1 Data Mining

Data mining berikut adalah Langkah-langkah dimana identik dengan menggunakan metode statistik matematika atau teknologi *artificial intelligence* yang bertujuan untuk mengekstrak atau mengidentifikasi sebuah pembelajaran yang bermanfaat bagi penggunaannya. Data mining bisa diartikan sebagai pengumpulan sebuah data dan pengolahan data yang berukuran besar. Data mining memiliki tugas-tugas yang telah dikelompokkan yaitu klasifikasi, kelustering, asosiasi, regresi atau estimasi. Tahapan data mining dibagi menjadi lima bagian yang pertama tahapan pembersihan data, integrasi data transformation, data mining, evaluasi pola, setelah melakukan proses data mining pola yang telah dihasilkan dari proses tersebut selanjutnya perlu

untuk melakukan evaluasi kepada pola tersebut agar pengguna mengetahui hasil dari hipotesis awalnya. metode menggunakan Data mining dilakukan untuk menyaring sebuah data agar bisa diidentifikasi. (Cahyanti dkk., 2020)

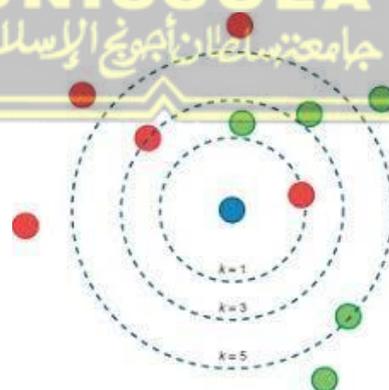
2.2.2 KNN

K-Nearest Neighbor adalah sebuah bilangan algoritma yang menggunakan metode klasifikasi yang bisa dilakukan untuk membuat sebuah data berdasarkan data pengetahuan. tahapan mengolah data menggunakan algoritma K-Nearest Neighbor, berikut mempunyai beberapa Langkah menentukan nilai K dari tetangga paling dekat, penghitungan jarak, menentukan tetangga dengan jarak dengan label kelas K-Nearest Neighbor, berikutnya dilakukan pelabelan pada kelas menggunakan tahapan pengujian terhadap nilai K dari tetangga yang paling dekat dengan data *testing* dan data *training*, jarak dihitung sesuai dengan apa yang telah ditetapkan, berikut di bawah rumusan menghitung jarak antara *Euclidean Distance*.

a. Rumusan perhitungan jarak antara *Euclidean Distance*

$$dist = \sqrt{\sum_{i=1}^n (x_{i2} - x_{i1})^2} \quad (1)$$

Dimana *dist* merupakan jarak antara X_{i2} merupakan biodata *testing* X_{i1} merupakan biodata biasa, I merupakan atribut lalu N yaitu total atributnya. (Yuliansyah dkk., 2017)



Gambar 1. 1 Ilustrasi Klasifikasi KNN

Berikut merupakan gambaran metode K-Nearest Neighbor yang mengklasifikasi sebuah objek dari yang terkecil sampai yang terbesar.

2.2.3 Klasifikasi

Klasifikasi adalah mengelompokkan sesuatu *Class* data secara sistematis yang sebelumnya tidak diketahui datanya, *klasifikasi* merupakan algoritma dalam penelitian kali ini dan klasifikasi bertujuan untuk menyusun suatu data secara berkelompok dan tertata sesuai bidangnya masing masing, klasifikasi mempunyai 4 dasar komponen yaitu ::(Ayudhitama & Pujiyanto, 2020)

- 1) Kelas adalah atribut yang berganti ke label dan bisa dibilang objeknya.
- 2) *Prediktor* adalah atribut data yang akan dilakukan Klasifikasi.
- 3) *Training dataset* berikut data mempunyai label yang sebelumnya
- 4) *Testing dataset* adalah data yang baru dibuat lalu akan diklasifikasikan

Pada *klasifikasi* ini digunakan algoritma K-NN. K-Nearest Neighbor berikut melakukan perbandingan pada data testing dan data latih. Untuk tahapan awal yang di proses mencari nilai K.

2.2.4 PHP

PHP adalah bahasa skrip serbaguna untuk pengembangan web. Sejak awal kelahirannya, PHP telah dianggap sebagai bahasa pemrograman terbaik di dunia oleh sebagian besar pengembang. Saat ini, PHP masih mendominasi bahasa pemrograman sisi server web. Menurut laporan W3TECHS, PHP menduduki peringkat teratas dalam daftar bahasa pemrograman yang digunakan oleh aplikasi web September 2022, dengan pasar sebesar 77,3%. Berkat fitur-fitur gratis dan dapat diperluas, ekosistem sumber terbuka PHP telah mengalami kemajuan yang luas dan cepat, dan saat ini terdapat lebih dari 138 ribu repositori PHP yang luar biasa di GitHub.(Lin dkk., 2023)

2.2.5 VSCode

Visual Studio Code adalah IDE (Integrated Development Environment) yang kuat, khusus untuk membangun aplikasi web dan program cloud. Alat ini adalah program ringan yang mirip dengan fungsionalitas Visual Studio, namun dilengkapi dengan utilitas revisi dan kompilasi yang komprehensif, yang terpadu dalam antarmuka yang modern. Program yang ringan dan memiliki alat

pemrograman yang handal. Salah satu fitur menonjol dari Visual Studio Code adalah kemampuan kontrol Git yang ditawarkan oleh program ini. Ini memfasilitasi pengujian, pembuatan, pengemasan, dan bahkan implementasi berbagai jenis perangkat lunak. (Macedo & Pierce, 2022)

2.2.6 SINTA

SINTA (Science and Technology Index) adalah sebuah database data ilmiah di Indonesia. SINTA memberikan akses terhadap sitasi dan keahlian di Indonesia. Sistem informasi penelitian berbasis web ini menawarkan sistem akses yang cepat, mudah, dan komprehensif untuk mengukur kinerja para peneliti, institusi, dan jurnal di Indonesia. SINTA adalah program pengindeksan (jurnal, artikel, dan penulis) yang berasal dari Indonesia dan dirancang oleh Kementerian Riset, Teknologi, dan Pendidikan Tinggi. (Kemenristekdikti). (Ahmadi, 2019)

2.2.7 Google scholar

Google scholar merupakan platform yang dirancang untuk mencari karangan ilmiah termasuk jurnal, artikel, buku, tesis dan makalah. *Google scholar* diluncurkan pada tahun 2004 oleh konsorsium *google* yang bertujuan untuk mengindeks sebuah artikel ilmiah, ataupun karya ilmiah lainnya yang sedang viral, semua orang dapat mengakses *google scholar* termasuk dosen ataupun mahasiswa yang dapat mencari sebuah karya ilmiah sesuai judul yang telah ditentukan. *Google scholar* ikut serta membantu pengguna untuk menyediakan artikel ilmiah khusus yang telah disediakan oleh institusi-institusi. *Google scholar* telah menyediakan jurnal yang banyak sekali dikutip oleh penggunanya menyesuaikan pencarian dan menampilkan jurnal yang paling tinggi. Karena mencari artikel di situs *google scholar* yang terbilang mudah, kini situs *google scholar* menjadi tempat pembelajaran bagi semua orang. (Statistika dkk., 2017)

2.2.8 Evaluasi Mesin Klasifikasi

Evaluasi klasifikasi diperlukan sebagai pengukuran pekerjaan dan *system* digunakan *confusion matrix* untuk mengetahui hasil dari akurasi, dan *precision*, *recall* adalah bertujuan untuk pengukuran kelengkapan klasifikasi. Evaluasi klasifikasi di bagi menjadi tiga bagian beriku di bawah ini: (Athaillah dkk., 2020)

- a. Akurasi merupakan nilai yang dibandingkan pada kasus yang diidentifikasi benar pada semua kelas seperti di bawah ini

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

- b. *precision* adalah proporsi kasus dengan hasil positif persamaan

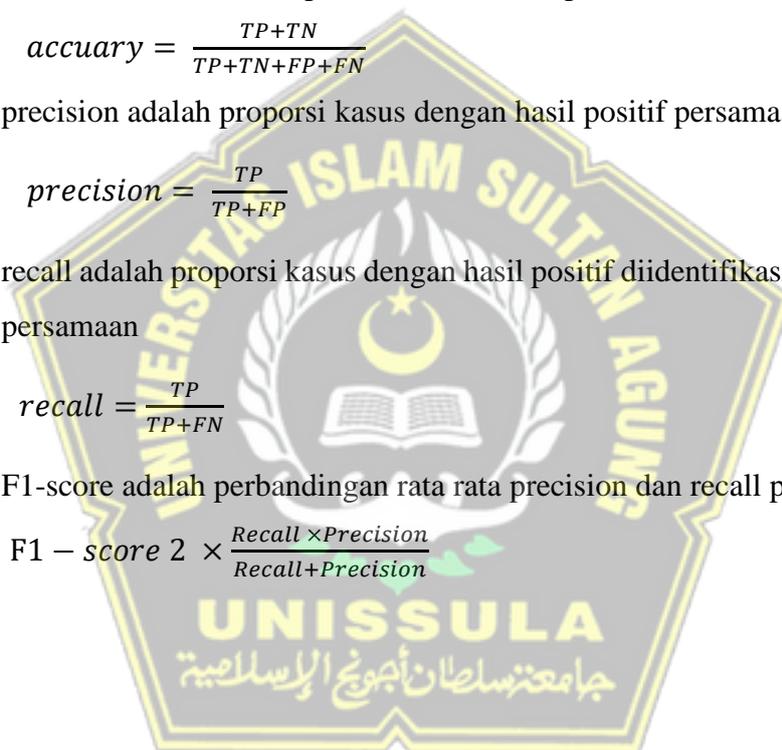
$$precision = \frac{TP}{TP+FP} \quad (3)$$

- c. *recall* adalah proporsi kasus dengan hasil positif diidentifikasi dengan benar persamaan

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

- d. F1-score adalah perbandingan rata rata *precision* dan *recall* persamaan

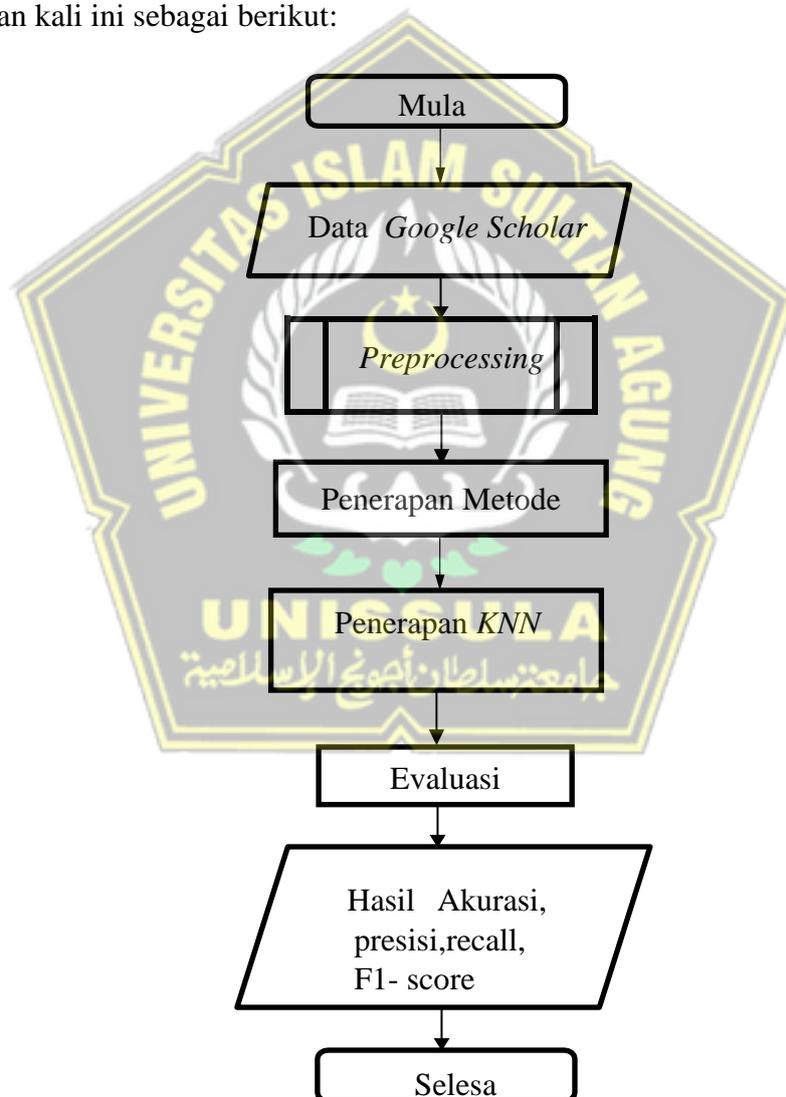
$$F1 - score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (5)$$



BAB III METODE PENELITIAN

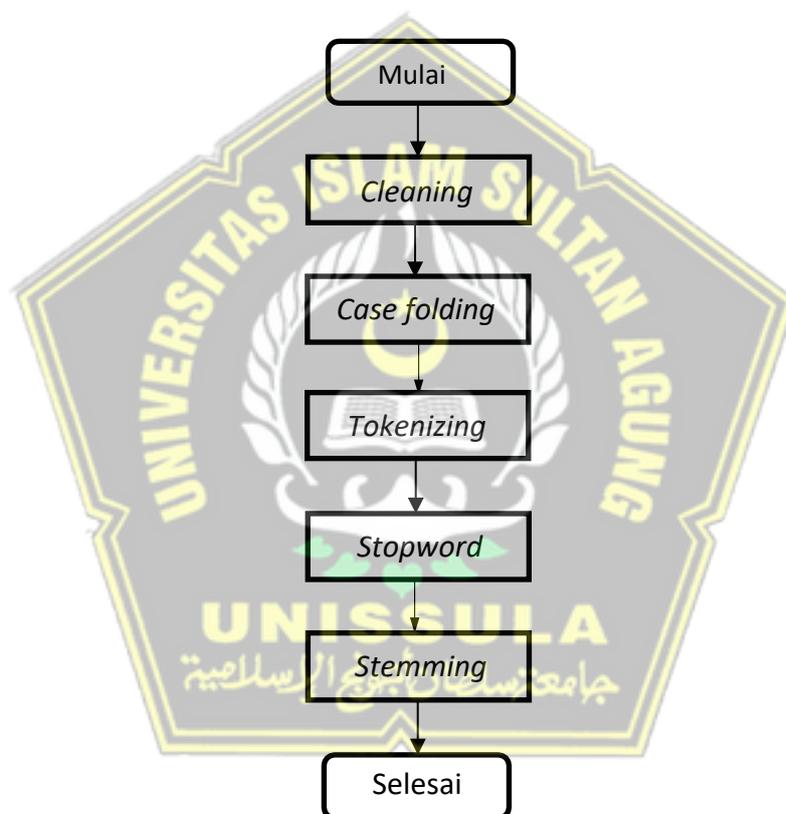
3.1 Tahap penelitian

Pada tahapan penelitian tugas akhir ini, algoritma yang diterapkan untuk klasifikasi yaitu K-NN. Algoritma ini, mengklasifikasi sebuah data WoS di SINTA ke dalam 5 bidang ilmu. Adapun langkah-langkah alur pada tahapan penelitian kali ini sebagai berikut:



Gambar 3. 1 Skema proses alur rancangan penelitian

Pada Gambar diatas merupakan skema alur proses studi dimana dari tahapan penelitian ini yaitu tahap pertama dengan memulai pengambilan data *google scholar*, setelah mendapatkan data tersebut kemudian melakukan proses validasi yang dilakukan oleh tim pengembang website *google scholar*, setelah data tervalidasi kemudian melakukan *preprocessing* setelah melakukan *preprocessing* kemudian mengolah data menggunakan metode algoritma KNN, setelah data diolah kemudian akan dilakukan evaluasi guna mendapatkan hasil berupa *akurasi*, *presisi* dan *recall*



Gambar 3. 2 Tahapan preprocessing

Gambar 3.2 merupakan tahapan *preprocessing* dimana dalam *preprocessing* memiliki lima tahapan yaitu *cleaning* yang merupakan proses menghilangkan karakter tanda baca pada teks, kemudian *case folding* yang merupakan proses untuk merubah kata dalam kalimat sehingga merubah kata halus, kemudian *tokenization* berikut adalah merupakan menghapus huruf yang berada di dalam dokumen untuk mengubah kata besar menjadi kata kecil, kemudian *stopword* merupakan proses

menghapus kata-kata yang tidak relevan dan mengambil kata yang penting contoh kata yang dibuang yaitu “yang”, “dan”, “di”, dan “dari”, kemudian terakhir yaitu *Stemming* berikut bertujuan untuk mengganti huruf yang berimbuhan menjadikan huruf utama.

3.1.1 Koleksi data

Berikut mengkoleksi sebuah data pada tugas akhir kali ini diperoleh dengan melakukan penelusuran di internet khususnya *Web Sinta*. Tugas akhir ini melakukan pencarian informasi berupa data agar bisa dilakukan pemrosesan *klasifikasi* secara merata. Data yang di perlukan pada penelitian ini berjumlah 100 judul publikasi yang di ambil dari web SINTA.

3.1.2 Preprocessing

Pada tahapan *Preprocessing* adalah bertujuan untuk melakukan pemrosesan sebuah data yang sudah valid, *preprocessing* dapat diartikan sebagai tahapan untuk menyiapkan sebuah document mentahan maupun itu document testing atau document training. *preprocessing* memiliki fungsi yang bertujuan untuk memberiskan data yang kurang baik pada document agar dapat diimplementasikan, tahapan tersebut di bagi menjadi empat bagian sebagai berikut:

a) Case Folding

Case Folding berikut merupakan merubah keseluruhan kata yang ada dalam dokumen mejadi kata kecil. Kata “a” sampai kata akhir yaitu “z” yang diterima. Karakter selain kata dihapus dan diperlukan sebagai pemisah.

Tabel 3. 1 Contoh hasil proses *Case Folding* pada kolom artikel

Artikel	Hasil <i>Case Folding</i>
saya membuat sebuah sistem untuk memenuhi tugas akhir saya. sistem yang saya buat bertujuan utuk mengolah data yang akan menentukannya hasil acurasi	saya membuat sebuah sistem untuk memenuhi tugas akhir saya. sistem yang saya buat bertujuan utuk mengolah data yang akan menentukan hasil acurasi

b) Tokenization

Tokenization adalah merupakan penguraian mendeskripsikan sebuah perkataan asli yang semula berupa huruf dan menghapus pembatas, berikut tanda titik [.], koma [,], tanda kutip [“], tanda kurung tutup [[]], jarak lalu karakteristik huruf yang berada pada huruf berikut.

Tabel 3. 2 Contoh hasil Tokenezing

Hasil Remove	Hasil Tokenezing
saya membuat sebuah sistem untuk memenuhi tugas akhir saya. sistem yang saya buat bertujuan untuk mengolah data yang akan menentukan hasil akurasi	(Saya,'membuat','sebuah,' 'sistem','untuk','memenuhi,' tugas','akhir','saya') (sistem,' yang,' saya,' buat,' bertujuan,' untuk,' mengolah,' data,' yang,' akan,'menentukan,' hasil,' akurasi')

c) Stopword

Tahap *filtering* adalah tahap pengambilan kata-kata penting dari hasil token. Bisa menggunakan algoritma *stoplist / stopwords* adalah kata-kata yang tidak deskriptif yang dapat dibuang. Berikut merupakan cara kerja dari *stopword* “ke” , “hanya” , “berikut” , dan selanjutnya lebih banyak lagi.

Tabel 3. 3 Contoh hasil *Stopword*

Hasil Tokenezing	Hasil Stopword
(Saya,'membuat','sebuah,' 'sistem','untuk','memenuhi,' tugas','akhir','saya') (sistem,' yang,' saya,' buat,' bertujuan,' untuk,' mengolah,' data,' yang,' akan,'menentukan,' hasil,' akurasi')	(Saya,'membuat','sebuah,' 'sistem','untuk','memenuhi,' tugas','akhir','saya') (sistem,'saya,' buat,' rtujuan,' untuk,' mengolah,' data,' yang,' akan,'menentukan,' hasil,' akurasi')

d) Stemming

Jika sudah melakukan tahapan stopwords kemudian tahap yang perlu dilakukan adalah tahapan stemming. Stemming ialah penguraian dan pemetaan dari segala *varians* dari kata berubah hingga kata dasar.

Tabel 3. 4 Contoh hasil *Stemming*

Hasil <i>Stopword</i>	Hasil <i>Stemming</i>
saya membuat sebuah sistem untuk memenuhi tugas akhir saya. sistem yang saya buat bertujuan untuk mengolah data yang akan menentukannya hasil acurasi	saya membuat sebuah sistem untuk memenuhi tugas akhir saya. sistem yang saya buat bertujuan untuk mengolah data yang akan menentukan hasil acurasi

3.1.3 Penerapan Metode

pada studi literature kali ini pengkaji menerapkan pembelajaran yang berkaitan pada penelitian, pengkaji mencari referensi ilmiah yang dapat dipercaya untuk dijadikan acuan sebagai penyusunan pada tugas akhir kali ini, selanjutnya peneliti dapat melihat seluruh Pustaka yang telah di jadikan acuan pada halaman daftar Pustaka.

```

70 // pipeline untuk preprocessing dan klasifikasi
71 $estimator = new PersistentModel(
72     new Pipeline([
73         //PREPROCESSING
74         new RegexFilter([
75             RegexFilter::EXTRA_WHITESPACE,
76             RegexFilter::EXTRA_WORDS
77         ]), //text cleaning
78         new MultibyteTextNormalizer(), //case folding - lower case
79         new StopWordFilter($stopwords_en), //stopwords
80         new WordCountVectorizer(), //tokenizer
81
82         // ALGORITMA KLASIFIKASI
83     ], new KNearestNeighbors(3)),
84     // nama file model
85     new Filesystem('./riji-irawan.model')
86 );

```

Gambar 3. 3 Proses penerapan metode di dalam coding

3.1.4 Evaluasi

Tahapan berikut ini merupakan tahapan hasil evaluasi pada *prediction* yang menghasilkan tiga metode klasifikasi dari nilai terdekat pada data sebenarnya. Tahapan evaluasi melakukan perhitungan nilai K sehingga menghasilkan nilai *akurasi*, *precision*, dan *recall*.

```

104 // validasi, confusion matrix, akurasi, dll
105 $report = new AggregateReport([
106     'breakdown' => new MulticlassBreakdown(),
107     'confussion_matrix' => new ConfusionMatrix(),
108 ]);

```

Gambar 3. 4 Proses evaluasi dengan *ConfusionMatrix* di dalam *coding*

3.2 Metode Perancangan Alur Sistem

3.2.1 Analisis Kebutuhan

Analisis kebutuhan melakukan analisis data/objek pada Google Scholar data tersebut didapatkan melalui web SINTA selanjutnya beberapa document latih yang telah ditentukan document test yang sebelumnya belum pernah dikategorikan, berikut ada 2 kategori input pada sistem pada penelitian ini. selanjutnya dari dua kategori berikut akan dilakukan penerapan text preproscsing sebagai berikut *Case Folding*, bertujuan untuk mengganti seluruh kata didalam dokumen merubah ke kata kecil, *tokenization*, berikut mengolah document menjadii frase/term [kata], telah disesuaikan dengan masing-masing document. berikut Setelah melakukan *tokenization*, proses berikutnya melakukan tahap *stopword remov* atau bisa disebut menghapus huruf atau kata yang tidak penting. Selanjutnya dilakukan proses *stemming*, berikut adalah menghapus atau menghilangkan kata agar menjadi kata utama. Sebuah system tersebut menginput sebuah data dan menghasilkan sebuah klsaifikasi dan ada beberapa proses pada system tersebut antara lain adalah sebagai berikut:

a) Input judul

Input judul adalah hal pertama yang harus di lakukan sebuah sistem tersebut, karena sistem tersebut membutuhkan judul data yang akan digunakan untuk

mengkalsifikasi sebuah data tersebut nantinya, Sebelum menerapkan proses melakukan input judul data sistem ini sudah menyimpan dataset publikasi judul.

b) Proses Klasifikasi

Proses pada sistem ini adalah mengkalsifikasi sebuah judul data, di mana sistem ini akan mengolah judul publikasi terideks *Google Scholar* dengan algoritma K-NN.

c) Melakukan Prediksi

berikut fungsi pada prediksi yaitu sistem bertugas untuk melakukan prediksi dari judul publikasi yang telah dikumpulkan. Sistem ini akan melakukan pencarian tetangga terdekat terhadap judul publikasi yang akan diprediksi, dimana dalam mencari tetangga terdekat sistem ini akan menggunakan algoritma K-NN untuk melakukan prediksi.

d) Menampilkan Hasil Prediksi

fungsi terakhir pada sistem berikut menampilkan hasil dari prediksi yang telah dilakukan sebelumnya, fungsi menampilkan hasil prediksi bertujuan untuk membuat user melihat hasil perhitungan yang telah dilakukan sistem ini, pada fungsi ini user akan melihat berapa persentase tetangga terdekat dari setiap kelas yang dimiliki 5bidang ilmu yang diprediksi.

3.2.2 Implementasi Sistem

Pada Tahapan Analisis Sistem, akan dianalisa, *tools* bertujuan untuk mempersiapkan sesuatu dalam pengembangan pada sebuah sistem ini nantinya, berikutnya ada beberapa *tools* yang berfungsi untuk membuat sebuah sytem pada berikut ini:

a) RubixML

RubixML adalah mesin untuk pembelajaran lebih mendalam menggunakan bahasa PHP yang memiliki produksi dengan lebih dari 40 algoritma pembelajaran mesin, sehingga dapat menentukan objek model dalam satu baris, lalu menggunakannya untuk memprediksi nilai

b) Visual Studio Code

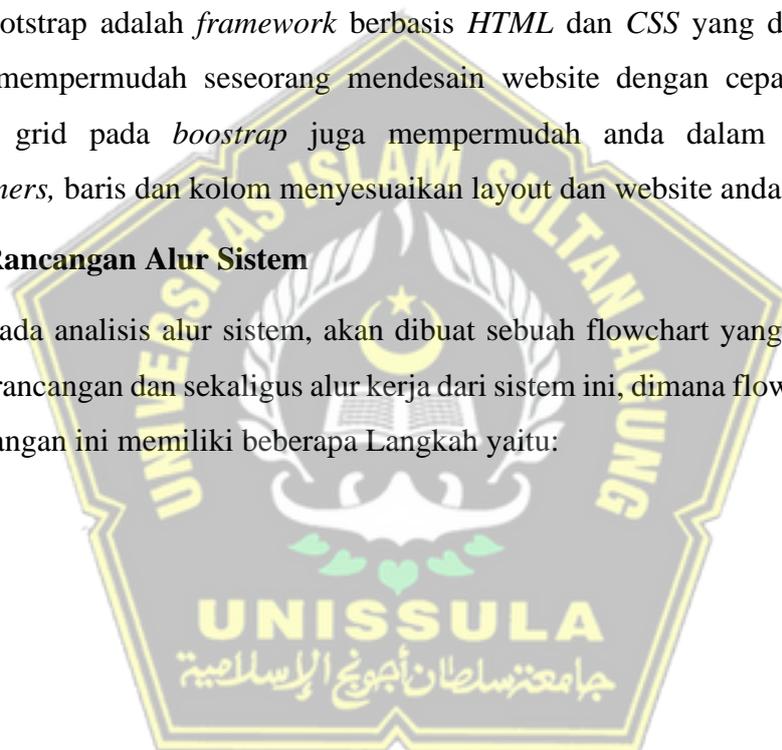
Vscode adalah *code editor* untuk menciptakan berbagai software yaitu Microsoft sebuah perangkat dekstop, vscode merupakan *editor code* yang dikembangkan oleh Microsoft untuk *linux* dan *macOSs*. Vscode menggunakan Bahasa pemograman, yaitu *java*, *javascriptt*, *Go*, *Node.jss*, *Paython*, dan *C++*, Unntuk kembangkan aplikasi *Nodee.JS*. Vscode memiliki berbagai komponen editor pada *azure DevOps*.

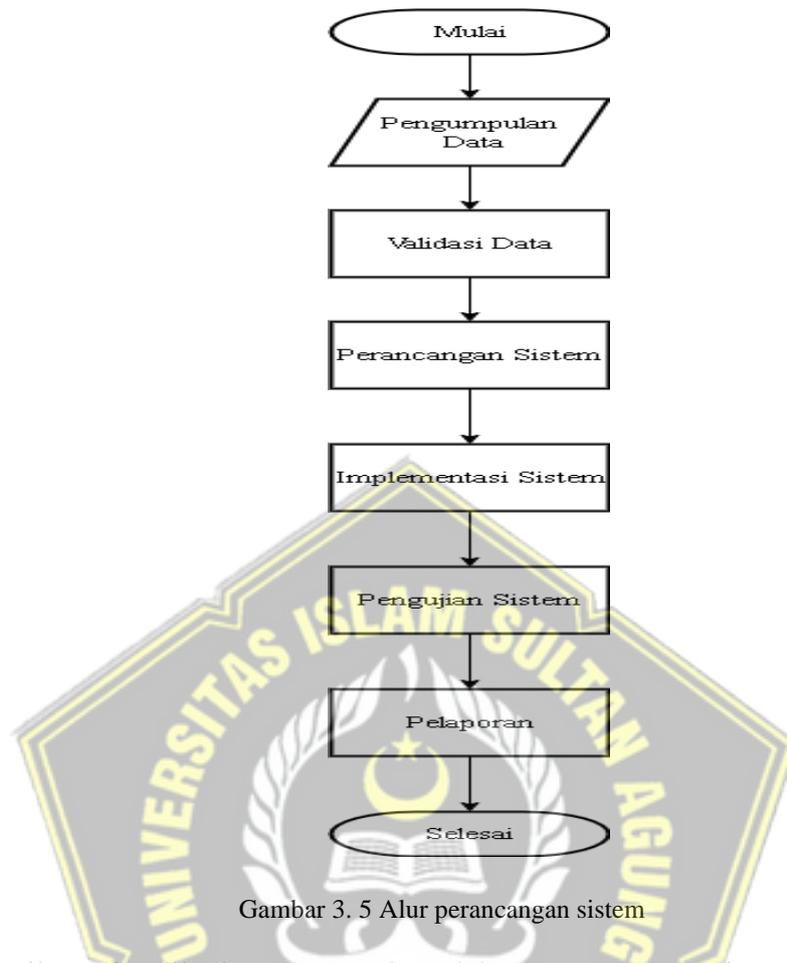
c) Bootstrap

Bootstrap adalah *framework* berbasis *HTML* dan *CSS* yang di kembangkan untuk mempermudah seseorang mendesain website dengan cepat dan mudah. Sistem grid pada *bootstrap* juga mempermudah anda dalam menggunakan *Containers*, baris dan kolom menyesuaikan layout dan website anda.

3.2.3 Rancangan Alur Sistem

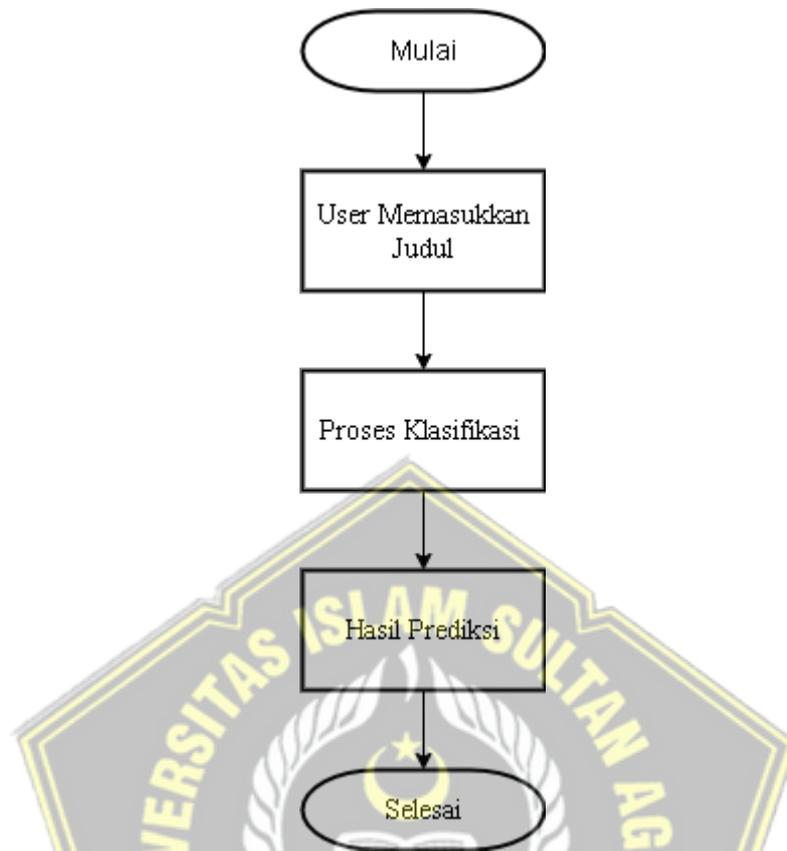
Pada analisis alur sistem, akan dibuat sebuah flowchart yang menunjukkan alur perancangan dan sekaligus alur kerja dari sistem ini, dimana flowchart dari alur perancangan ini memiliki beberapa Langkah yaitu:





Gambar 3. 5 Alur perancangan sistem

berikut Diperlihatkan urutan alur dalam perancangan sistem dimana pada tahap pertama terdapat, pengumpulan data dari database SINTA, tahapan selanjutnya validasi data, kemudian jika data benar langkah selanjutnya yaitu perancangan alur sistem dan dilanjutkan kedalam implementasi *coding* dan selanjutnya adalah penerapan metode. Setelah itu, pengujian sistem dan yang terakhir adalah pelaporan.

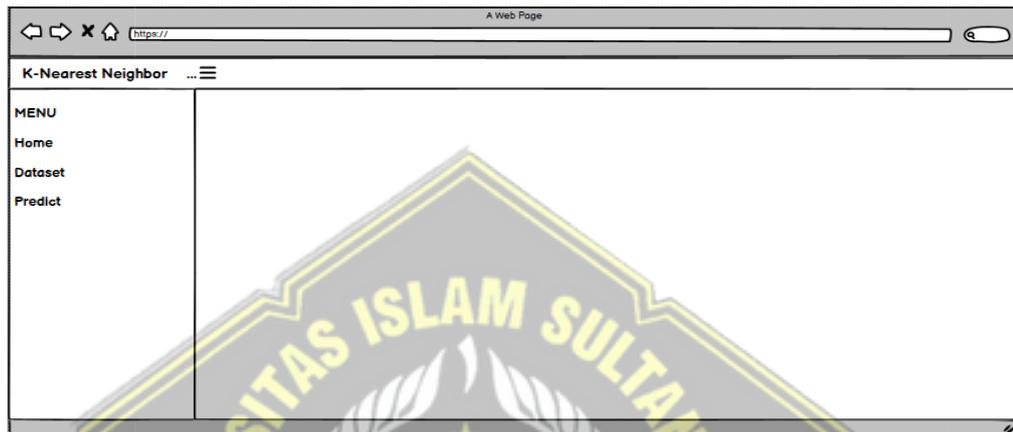


Gambar 3. 6 Flowchart system

berikut gambar 3.6 Adalah flowchart dari sistem ini nantinya, alur penggunaan sistem ini nantinya akan diawali dengan user memulai menggunakan *device* yang akan mereka gunakan, dan dalam penelitian ini, penulis menggunakan laptop sebagai *device* utama, pada bagian awal dari sistem ini, user diharuskan membuka sistem ini, setelah itu user akan melihat beberapa menu seperti Home, halaman *dataset*, dan halaman *predict*. User kemudian memilih halaman *predict*. Dalam halam *predict* user diharuskan memasukkan judul yang akan di prediksi. Setelah user memasukkan judul, dengan menekan tombol prediksi selanjutnya adalah proses *klasifikasi* dimana sistem mengklasifikasi judul dengan metode *K-Nearest Neighbor*. Sistem ini dapat melakukan prediksi judul publikasi dari lima bidang ilmu yang disebutkan, maka alur proses dari sistem ini telah selesai.

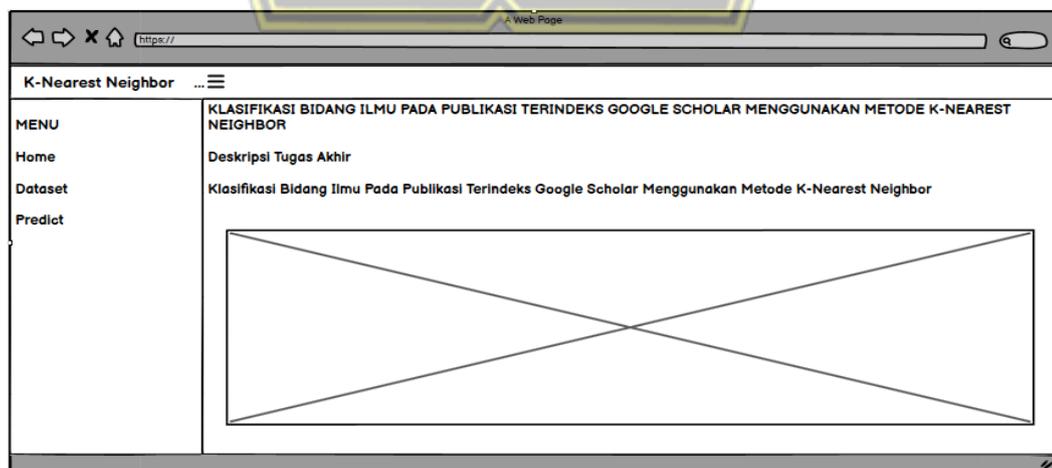
3.2.4 Perancangan Interface

Pada bagian berikut ini merupakan rancangan struktur pada menu *mockup* yang akan dibuat dalam Sistem. Berikut adalah rancangan antarmuka dari Sistem klasifikasi bidang ilmu pada publikasi terindeks *Web of Science* dengan algoritma K-NN sebagai metodenya.



Gambar 3. 7 Gambar halaman menu

Pada gambar 3.7 diatas merupakan tampilan halaman awal saat memasuki halaman website, karena penulis berfokus membuat sistem yang sederhana dan mudah digunakan. Maka halama utama dapat menjalankan 3 fungsi yaitu, halaman *Home* yang berisi tentang deskripsi tugas akhir, halaman *Dataset* berupa 1000 data sampel dan yang terakhir yaitu halaman *predict* yang berfungsi untuk melakukan klasifikasi pada sebuah judul data yang telah diambil dari halaman *datas et*.



Gambar 3. 8 Halaman home

Pada gambar 3.8 diatas iyalah tampilan sebuah halaman home yang menampilkan penjelasan deskripsi metode tugas akhir dan ada beberapa gambar pada tampilan tersebut.

K-Nearest Neighbor ...

MENU
Home
Dataset
Predict

Dataset yang Digunakan K-Nearest Neighbor (KNN)
Sampel Data Yang Digunakan dalam penelitian Ini Menggunakan Data Dari Database Google Scholar.
Data Google Scholar Unduh Dataset

No.	Judul	Bidang Ilmu
1	Reflections on integrated research from community engagement in peatland restoration	Arts & Humanitiest
2	The effect of impedance on interaural azimuth cues derived from a spherical head model	Arts_&_Humanitiest
3	The effectiveness of the socialization of talsi implies in social media	Arts_&_Humanitiest
4	The effect of implementation of asset management and competence of state apparatus	Arts_&_Humanitiest
5	teaching multiculturalism based on Islamic historical relics in Northern Java	Arts & Humanitiest
6	Evaluating the paleoclimatic significance of clay mineral records from a late Pleistocene loess pale	Arts & Humanitiest
7	The use of social media whatsapp among English education students for solving thesis writing problem	Arts_&_Humanitiest
8	Shili on functions of sexual euphemisms in english-Indonesian translation of Duke of Her own by Eloisa James	Arts_&_Humanitiest

Gambar 3. 9 Halaman dataset

Pada gambar 3.9 diatas adalah tampilan sebuah halaman dataset sampel data google scholer yang digunakan untuk melakukan prediksi dan klik tombol unduh dataset untuk mendownload file data tersebut.

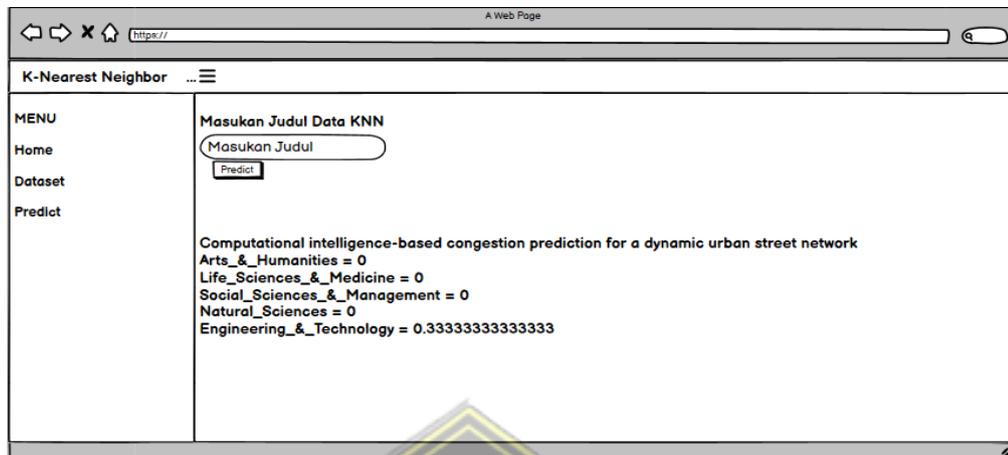
K-Nearest Neighbor ...

MENU
Home
Dataset
Predict

Masukan Judul Data KNN
Masukan Judul
Predict

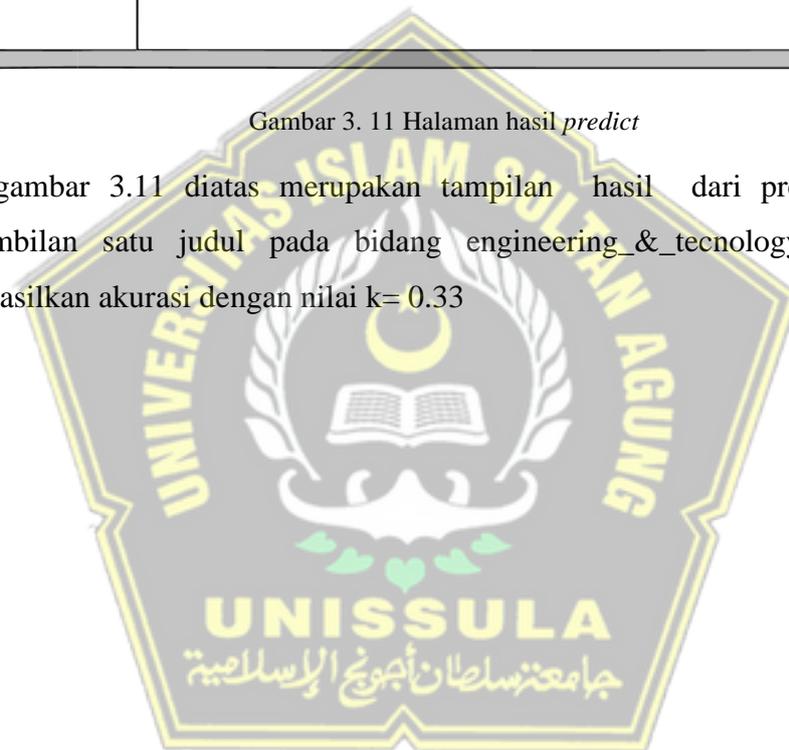
Gambar 3. 10 Halaman *predict*

Pada gambar 3.10 diatas adalah tampilah from *predict*, from ini digunakan memilih judul data pada dataset. Setelah di pilih dan klik tombol *predict* dan akan menghasilkan sebuah *acurasi*



Gambar 3. 11 Halaman hasil *predict*

Pada gambar 3.11 diatas merupakan tampilan hasil dari prediksi dengan pengambilan satu judul pada bidang engineering_&_tecnology yang telah menghasilkan akurasi dengan nilai $k=0.33$

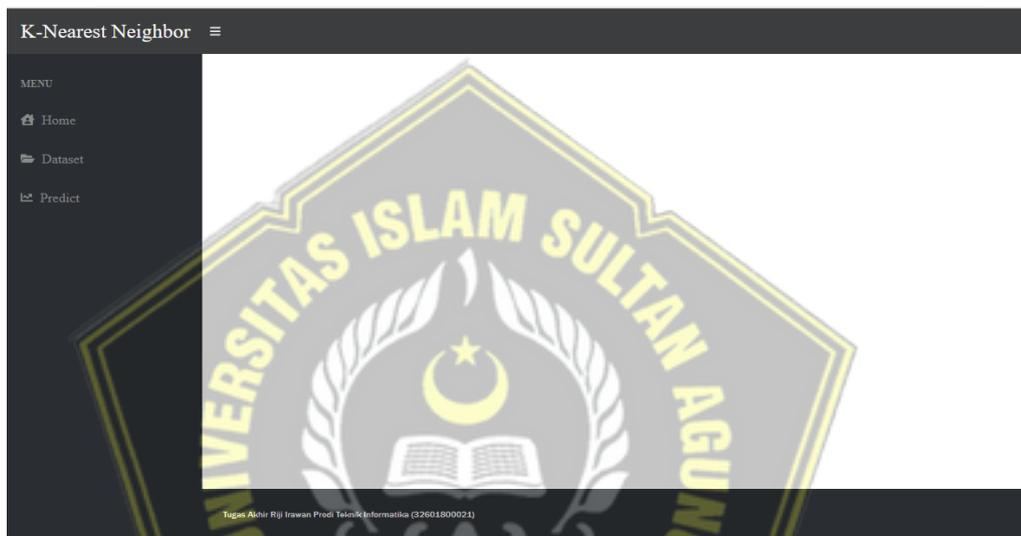


BAB IV

HASIL DAN ANALISIS PENELITIAN

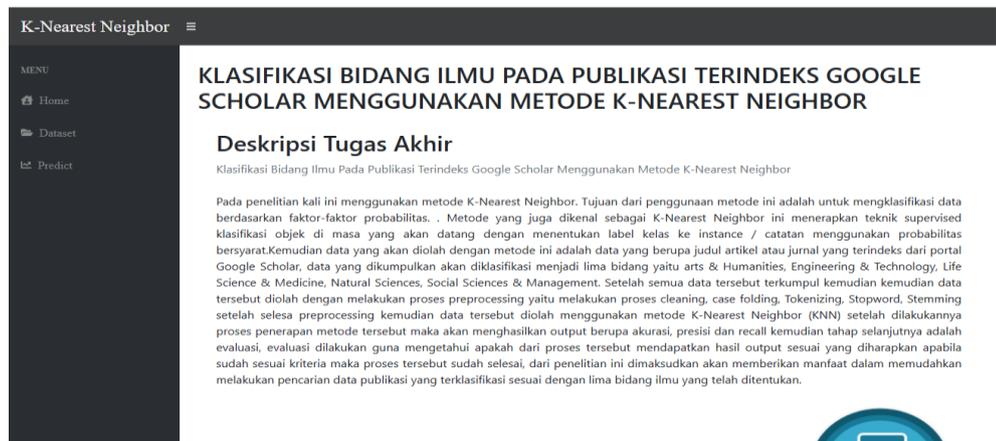
4.1 Implementasi *User Interface*

Implementasi system adalah alur rancangan yang dilakukan setelah rancangan selesai dibuat. Pada sistem tersebut menunjukkan tampilan *user interface*



Gambar 4. 1 Halaman awal *UI*

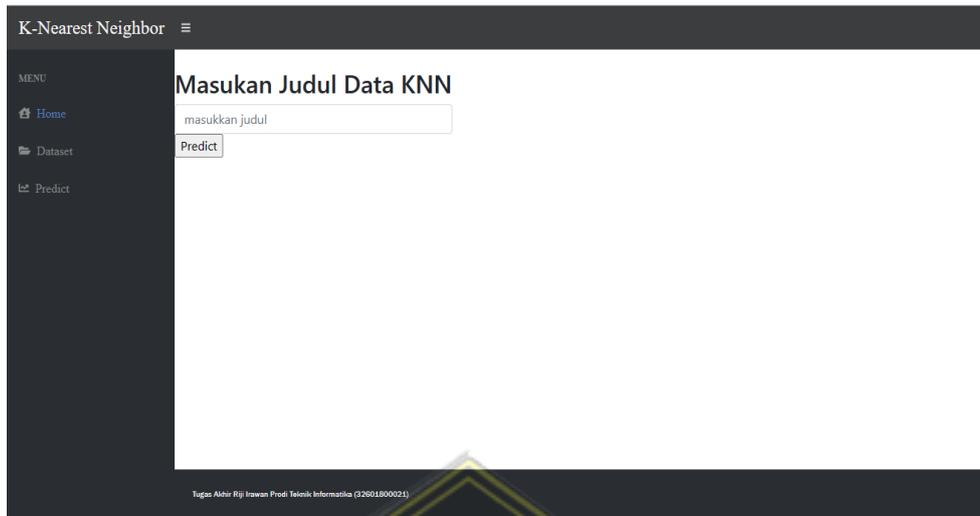
Pada gambar 4.1 diatas adalah rancangan halaman awal pada sistem ini yang berisi Banner, halaman Home yang berisikan tentang penjelasan singkat pada metode K-Nearest neighbor, sedangkan Dataset merupakan halaman yang berisikan data sampel, dan predict merupakan halaman dimana pengimputan judul data sampel dan melakukan predict.

Gambar 4. 2 Rancangan halaman home *UI*

Berikut gambar 4.2 merupakan *UI* tampilan awal pada halaman *Home* yang memaparkan judul tugas akhir, pembahasan singkat pada deskripsi tugas akhir pada penelitian kali ini dan ada beberapa gambar logo pada web *sinta* di bagian bawah pada halaman ini.

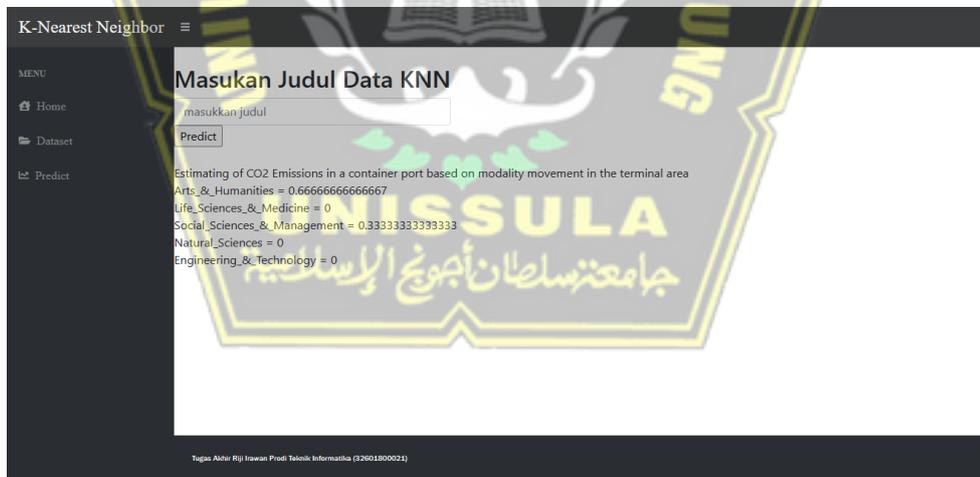
Gambar 4. 3 Rancangan halaman *UI dataset*

Pada gambar 4.3 diatas memaparkan tampilan dari menu dataset yang berupa *fitur download dataset*, jika di klik maka user dapat mengunduh *dataset* tersebut dengan bentuk file *Excel*, dan selanjutnya menu *dataset* menampilkan 1000 judul tabel data yang memiliki lima bidang ilmu.



Gambar 4. 4 Rancangan Halaman UI *predict*

Pada gambar 4.4 diatas merupakan tampilah menu *predict* dimana tujuannya untuk memprediksi sebuah judul *dataset*, pilih judul *dataset* lalu *input* ke kolom “masukan judul” stelah itu kelik tombol *button* maka akan menghasilkan akurasi di setiap bidangnya.



Gambar 4. 5 halaman hasil predict UI

Pada gambar 4.5 diatas sudah sedikit disinggung pada penjelasan gambar 4.4 pada gambar ini memaparkan hasil dari judul data pada lima bidang ilmu Arts & Humanities, Engineering & Technology, Life Sciences & Medicine, Natural Sciences, Social Sciences & Management yang sudah diprediksi dan menghasilkan akurasi di setiap bidangnya.

4.2 Analisa Dan Pengujian

Setelah tahapan implementasi *user interface* telah dilakukan, selanjutnya mealakukan penerapan pada sistem menggunakan metode pengujian *black book testing* dimana pengujian ini bertujuan untuk menguji prangakat lunak pada sistem, dengan tujuan menguji fitur aplikasi tanpa melihatnya, lalu percobaan berikut dibuat sebagai menguji komponen-komponen yang berada didalam system tersebut.

Tabel 4. 1 Pengujian alur pada sistem

Skenario Pengujian	Kasus pengujian	Hasil Pengujian	Kesimpulan
Unduh Dataset	Mengunduh sampel data	Berhasil	Sesuai
Masukan judul	Masukkan judul publikasi	Berhasil	Sesuai
Memprediksi judul	Sitem memprediksi judul		
Hasil prediksi Judul	Output hasil prediksi judul		

Berikut diatas pada Tabel 4.1 merupakan hasil pengujian pada klasifikasi bidang ilmu pada publikasi terindek Google scholar menggunakan metode *K-nearest neighbor* (KNN), sudah sesuai dan fungsinya dapat dijalankan, tahapan testing berjalan dengan baik sesuai dengan fungsinya.

4.3 Analisa Akurasi

Berikutnya melakukan tahapan Analisis Akurasi, dimana tahapan ini untuk mengetahui pengujian tingkat akurasi pada sistem yang telah di buat. Tahapannya menggunakan data testing, selanjutnya dilakukan pembagian dari data testing dan data sampel.

Tabel 4. 2 Tahapan analisa *akurasi*

Nama Data	Jumlah	Jumlah
Data Sampel	90%	900
Data Testing	10%	100

Total Data = 1000

Berikut untuk perhitungan akurasi pada sistem menghasilkan *Confusion Matrix* dimana tujuan dari *Confusion Matrix* ini digunakan untuk mencari *Precision*, *Accurasy*, *recall* dan *F1 Score* dengan K yang berbeda beda yaitu 10, 15, 21,25, 30 mengkalasifikasi judul data publikasi

Tabel 4. 3 Hasil perhitungan *Counfusion Matrix* K=10

	Actually Positive	Actually Negatif
Predict Positivie	77 TP	223 FP
Actually negatif	223 FN	308 TN

Pada tabel 4.3 diatas adalah hasil dari perhitungan *Confusion Matrix* K=10 dengan hasil *True Positive* 77, *Flase negatif* 223, *Flase positif* 223 dan *True Negatif* 308.

Tabel 4. 4 Hasil perhitungan *confusion matrix* K=15

	Actually Positive	Actually Negatif
Predict Positive	86 TP	214 FP
Actually Negatif	214 FN	344 TN

Pada tabel 4.4 Diatas berikut merupakan percobaan kedua dengan perhitungan Confusion matrix K=15 dengan hasil akurasi True Positif 86, False Negatif 214, False 214 dan true positif 344.

Tabel 4. 5 Hasil perhitungan *Confusion Matrix* K=21

	Actually Positive	Actually Negatif
Predict Positive	79 TP	221 FP
Actually Negatif	221 FN	316 TN

Selanjutnya pada tabel 4.5 Adalah percobaan ketiga dengan perhitungan *Confusion Matrix* K=21 dengan akurasi *true positif* 79, *false negatif* 221, *false positif* 221 dan *true negatif* 316.

Tabel 4. 6 Hasil perhitungan *Confusion Matrix* k=25

	Actually Positive	Actually Negatif
Predict Positive	72 TP	228 FP
Actually Negatif	288 FN	228 TN

Berikutnya pada tabel 4.6 Diatas Merupakan percobaan keempat perhitungan menggunakan *Confusion matrix* pada K=25 dengan akurasi *true positif* 72, *false negative* 288, *false positif* 228 dan yang terakhir *true negatif* 288.

Tabel 4. 7 Hasil perhitungan *Counfusion Matrix* k=30

	Actually Positive	Actually Negatif
Predict Positive	92 TP	208 FP
Actually Negatif	208 FN	368 TN

Pada tabel 4.7 Diatas adalah percobaan kelima pada perhitungan menggunakan *Counfusion matrix* pada K=30 dengan hasil akurasi yaitu, *true positif* 92, *flase negatif* 208, *flase positif* 208 dan terakhir *true negatif* 368

Dengan data yang tertera dan dengan pengujian nilai K yang berbeda, dapat diketahui Accuracy, Precision, recall dari setiap K, berdasarkan rumus yang ada mendapatkan hasil seperti pada tabel di bawah ini.

Tabel 4. 8 Jumlah K *akurasi Recall Presisi*

Jumlah K	Akurasi	Recall	Presisi	F1 Score
10	0.4856	0.2566	0.2380	0.1587
15	0.5198	0.2866	0.4672	0.2016
21	0.5029	0.2633	0.5459	0.1849
25	0.5378	0.3066	0.2892	0.2132
30	0.4730	0.2400	0.2331	0.1392

Berdasarkan tabel diatas dapat dilihat hasil dari beberapa kali percobaan, yang bahwa dalam pengujian akurasi berdasarkan nilai K yang telah ditentukan sebelumnya, sistem ini memiliki akurasi yang cukup rendah, yang terendah pada nilai K 24 yaitu bernilai 0.4541, dengan recall dan presisi masing- masing bernilai 0.2266, dan 0.2922. Sedangkan nilai akurasi tertinggi, pada nilai K yang bernilai

24 dengan nilai Accuracy, presisi, recall dan F1 score yaitu 0.5378, 0,3066 ,0,2892, dan 0,2132 skor tersebut merupakan skor tertinggi menurut penulis.

Tabel 4. 9 pembagian judul data pada 5 bidang ilmu

Bidang Ilmu	Jumlah Judul Data
Arts & humanities	20
Engineering & Technology	20
Life Sciences & Medicine	20
Natural Sciences	20
Social Sciences & Management	20

Pada tabel diatas melakukan klasifikasi sebuah judul data pada Arts & humanities, Engineering & Technology, Life Sciences & Medicine, Natural Sciences. Dari 5 bidang ilmu tersebut masing-masing memiliki 20 judul data dari 100 data yang telah dibagi pada bidangnya

4.4 Pembahasan Hal yang Mempengaruhi Hasil Akurasi

Pada penelitian ini telah melakukan pengujian pada 5 bidang ilmu dan menghasilkan akurasi yang cukup renda, yaitu sebesar .53% Maka dari itu dalam pembahasan hal yang mempengaruhi hasil dari akurasi, menjelaskan faktor-faktor apa saja yang mungkin mempengaruhi rendahnya hasil dari akurasi pada pengujian yang telah di terapkan, Adapun faktor faktor yang mempengaruhi hasil dari nilai akurasi sebagai berikut.

1. Terdapat beberapa judul yang membahas tentang topik dan tema yang tidak sesuai dengan bidang ilmunya:

Tabel 4. 10 Judul data tidak sesuai dengan bidang ilmu

No	Judul Artikel	Bidang Ilmu yang Terlabel	Bidang Ilmu yang Seharusnya
1	Dual inhibition of MEK and PI3K pathway in KRAS and BRAF mutated colorectal cancers	Engineering & Technology	Life Sciences & Medicine
2	Success of Management by Objectives (MBO) Extracurricular Programs Based on Participation of School Committees	Natural Sciences	Social Sciences & Management
3	Crowdsourcing as a tool to elicit software requirements	Natural Sciences	Engineering & Technology
4	The international trade foreign direct investment and economic growth: Asean perspective	Arts & humanities	Social Sciences & Management
5	Material flow cost accounting approach for sustainable supply chain management system	Engineering & Technology	Social Sciences & Management
6	Islamic community organization collaboration in rejecting presidential regulation on alcohol investment	Life Sciences & Medicine	Arts & humanities

7	Visual art from the perspectives of Islamic Shariah: A proposed model	Social Sciences & Management	Arts & humanities
8	Advances in adsorptive membrane technology for water treatment and resource recovery applications	Natural Sciences	Engineering & Technology
9	Detection of Mirai Malware Attacks in IoT Environments Using Random Forest Algorithms	Life Sciences & Medicine	Engineering & Technology
10	Design as Formulation: From application to reflection	Arts & humanities	Engineering & Technology
11	Business intelligence and big data analytics for organizational performance management in public sec	Engineering & Technology	Social Sciences & Management
12	Governance disclosures senior management and their influences on tax avoidance	Arts & humanities	Social Sciences & Management

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berikut pada hasil penelitian ini dapat disimpulkan bahwa mengklasifikasi judul artikel yang didapat pada web SINTA trindeks Google Scholar di bagi menjadi lima bidang ilmu menggunakan algoritma K-NN dengan hasil pengujian sistem berupa nilai akurasi ,recall, presisi, F1 Score. Telah didapatkan nilai akurasi 53% Recall 30% Presisi 28% F1 Score 21%.

Pada hasil berikut tergolong yang rendah pada system SINTA, dikarenakan Terdapat beberapa judul yang membahas tentang topik dan tema yang tidak sesuai dengan bidang ilmunya, dari beberapa kali percobaan pengujian data yang menghasilkan akurasi tidak sesuai yang diharapkan, akan tetapi pada sistemnya bekerja sesuai dengan fungsinya yang dapat mengklasifikasi sebuah judul data dengan baik, dan dapat memberikan penjelasan.

5.2 Saran

Saran kepada prangkat sytem :

1. Pada penelitian selanjutnya perlu adanya tambahan algoritma lain untuk menunjang hasil keakuratan suatu hasil yang lebih baik.
2. Karena hasil akurasi keseluruhan dengan nilai K yang berbeda, dengan hasil yang masih rendah dan perlu ditingkatkan bagi peneliti selanjutnya dengan menggunakan metode selain *K-Nearest Neighbor*.
3. Untuk sistem tepat pada desain lebih dirancang sedemikian menarik mungkin.
4. Penegmbangan pada sistem bisa ditambahkan berbagai fitur.

DAFTAR PUSTAKA

- Admojo, F. T., & Ahsanawati. (2020). Klasifikasi Aroma Alkohol Menggunakan Metode KNN. *Indonesian Journal of Data and Science*, 1(2), 34–38. <https://doi.org/10.33096/ijodas.v1i2.12>
- Ahmadi, A. (2019). The use of SINTA (Science and technology index) database to map the development of literature study in Indonesia. *International Journal of Mechanical Engineering and Technology*, 10(2), 918–923.
- 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>
- Athallah, M., Azhar, Y., & Munarko, Y. (2020). Perbandingan Metode Klasifikasi Berita Hoaks Berbahasa Indonesia Berbasis Pembelajaran Mesin. *Jurnal Repositor*, 2(5), 675. <https://doi.org/10.22219/repositor.v2i5.692>
- Ayudhitama, A. P., & Pujiyanto, U. (2020). Analisa 4 Algoritma Dalam Klasifikasi Liver Menggunakan Rapidminer. *Jurnal Informatika Polinema*, 6(2), 1–9. <https://doi.org/10.33795/jip.v6i2.274>
- Cahyanti, D., Rahmayani, A., & Husniar, S. A. (2020). Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara. *Indonesian Journal of Data and Science*, 1(2), 39–43. <https://doi.org/10.33096/ijodas.v1i2.13>
- Chen, Y. (2022). Linked Data Fusion Based on Similarity Calculation and K-Nearest Neighbor. *Journal of Physics: Conference Series*, 2221(1). <https://doi.org/10.1088/1742-6596/2221/1/012043>
- Ding, H., Chen, L., Dong, L., Fu, Z., & Cui, X. (2022). Imbalanced data classification: A KNN and generative adversarial networks-based hybrid approach for intrusion detection. *Future Generation Computer Systems*, 131, 240–254. <https://doi.org/10.1016/j.future.2022.01.026>
- Fadhilaturrahmi, F., Erlinawati, E., & Ananda, R. (2020). Workshop Sinta 2 dan

- Google Scholar di Universitas Pahlawan Tuanku Tambusai. *Jurnal Abdidas*, 1(4), 203–209. <https://doi.org/10.31004/abdidas.v1i4.45>
- Homepage, J., Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). IJCIT (Indonesian Journal on Computer and Information Technology) Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 6(2), 118–127.
- Lin, C., Xu, Y., Fang, Y., & Liu, Z. (2023). VulEye: A Novel Graph Neural Network Vulnerability Detection Approach for PHP Application. *Applied Sciences (Switzerland)*, 13(2). <https://doi.org/10.3390/app13020825>
- Macedo, H. D., & Pierce, K. (2022). *Proceedings of the 20th International Overture Workshop*. <http://arxiv.org/abs/2208.10233>
- Nur Ghaniaviyanto Ramadhan. (2021). Indonesian Online News Topics Classification using Word2Vec and K-Nearest Neighbor. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(6), 1083–1089. <https://doi.org/10.29207/resti.v5i6.3547>
- Rachman, T. (2018). Ji-an-hsien No Title No Title No Title. *Angewandte Chemie International Edition*, 6(11), 951–952., 10–27.
- Ribadas-Pena, F. J., Cao, S., & Darriba Bilbao, V. M. (2022). Improving Large-Scale k-Nearest Neighbor Text Categorization with Label Autoencoders. *Mathematics*, 10(16). <https://doi.org/10.3390/math10162867>
- Sebastian, D. (2019). Implementasi Algoritma K-Nearest Neighbor untuk Melakukan Klasifikasi Produk dari beberapa E-marketplace. *Jurnal Teknik Informatika dan Sistem Informasi*, 5(1), 51–61. <https://doi.org/10.28932/jutisi.v5i1.1581>
- Statistika, D., Matematika, F., Ilmu, D. A. N., & Alam, P. (2017). *Pengelompokan Dosen Its Berbasis*.
- Swetha, Z. S. S., & Dhanalakshmi, R. (2022). *A NOVEL APPROACH FOR*

PREDICTING HOUSE PRICE USING K-NEAREST NEIGHBORS ALGORITHM COMPARING ACCURACY. 12, 4466–4472.

Yandi Saputra, A., & Primadasa, Y. (2018). Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbour Implementation of Classification Method to Predict Student Graduation Using K-Nearest Neighbor Algorithm. *Techno.Com*, 17(4), 9.

Yuliansyah, E., Magdalena, I. R., & Estanto. (2017). Sistem Identifikasi Iris Mata Dengan Metode Independent Component Analysis dan Klasifikasi K-Nearest Neighbor. *e-Proceeding of Engineering*, 4(2), 1810–1815.

