

KLASIFIKASI JUDUL TUGAS AKHIR MAHASISWA MENGUNAKAN K-NEAREST NEIGHBOR

Oleh

AHMAD NUSI

T3117107

SKRIPSI

Untuk Memenuhi Salah Satu Syarat Ujian
Guna Memperoleh Gelar Sarjana



**PROGRAM SARJANA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS ICHSAN GORONTALO
GORONTALO
2024**

PERSETUJUAN SKRIPSI

KLASIFIKASI JUDUL TUGAS AKHIR MAHASISWA MENGUNAKAN K-NEAREST NEIGHBOR

Oleh

AHMAD NUSI

T3117107

SKRIPSI

Untuk memenuhi salah satu syarat ujian
guna memperoleh gelar Sarjana
Program Studi Teknik Informatika,
ini telah disetujui oleh Tim Pembimbing

Gorontalo Juni 2024

Pembimbing I


Sudirman S. Panna, M.Kom
NIDN: 0924038205

Pembimbing II


Yusrianto Malago, M.Kom
NIDN: 0909108901

PENGESAHAN SKRIPSI

**KLASIFIKASI JUDUL TUGAS AKHIR MAHASISWA
MENGUNAKAN K-NEAREST NEIGHBOR**

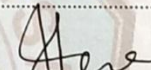
OLEH

AHMAD NUSI

T3117107

Diperiksa Oleh Panitia Ujian Strata (S1)
Universitas Ichsan Gorontalo

1. Ketua Penguji
Amiruddin, M.Kom, MCF
2. Anggota
Hastuti Dalai, M.Kom
3. Anggota
Serwin, M.Kom
4. Anggota
Sudirman S. Panna, M.Kom
5. Anggota
Yusrianto Malago, M.Kom



Mengetahui,

Dekan Fakultas Ilmu Komputer



Irvan A. Salhi, M.Kom
NIDN: 0928028101

Ketua Program Studi



Sudirman S. Panna, M.Kom
NIDN: 0924038205

PERNYATAAN SKRIPSI

Dengan ini saya menyatakan bahwa :

1. Skripsi ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar Akademik (Sarjana) baik di universitas Ichsan Gorontalo maupun perguruan Tinggi lainnya.
2. Skripsi ini adalah murni gagasan, rumusan, dan penelitian saya sendiri, tanpa Bantuan pihak lain, kecuali arahan tim pembimbing.
3. Dalam skripsi ini tidak terdapat karya yang dipublikasikan orang lain, Kecuali secara tertulis dicantumkan sebagai acuan dalam naskah dengan Disebutkan nama pengarang dan dicantumkan dalam daftar pustaka
4. Pernyataan ini saya buat dengan sesungguhnya dan apa bila kemudian hari Terdapat penyimpangan dan ketidak benaran dalam pernyataan ini, maka Saya bersedia menerima sanksi akademik berupa pencabutan gelar yang Telah diperoleh karena skripsi ini, serta sanksi lainnya sesuai dengan norma Yang berlaku diperguruan tinggi ini.

Gorontalo,

2024

pernyataan

Ahmad Nusi

KATA PENGANTAR

Alhamdulillah Robbilalamin penulis mampu menyelesaikan Skripsi dengan judul “**Klasifikasi Judul Tugas Akhir Mahasiswa Menggunakan K-Nearst Neighbor**”, untuk memenuhi salah satu syarat penyusunan tugas akhir Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo.

Penulis memahami bahwa penyusunan Skripsi ini tidaklah muda tanpa adanya suport dari ibu/bapak dosen serta pihak lainnya. Untuk itu penulis haturkan banyak terima kasih dan penghargaan yang setinggi-tingginya kepada :

1. Ibu Dr. Dra. Juriko Abdussamad, M.Si, selaku Ketua Yayasan Pengembangan Ilmu Pengetahuan dan Teknologi (YPIPT) Ichsan Gorontalo;
2. Bapak Dr. Abdul Gaffar La Tjokke M.Si, selaku Rektor Universitas Ichsan Gorontalo;
3. Pak Irvan Abraham Salihi, M.Kom, selaku Dekan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo
4. Bapak Sudirman Melangi, M.Kom, selaku Wakil Dekan I Bidang Akademik Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
5. Ibu Irma Surya Kumala Idris, M.Kom, selaku Wakil Dekan II Bidang Administrasi Umum dan Keuangan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo
6. Bapak Sudirman S. Panna, M.Kom, selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo dan juga sebagai pembimbing utama penulis.
7. Bapak Yusrianto Malago, M.Kom, selaku Pembimbing II yang telah banyak membimbing penulis selama ini;
8. Bapak dan Ibu Dosen Universitas Ichsan Gorontalo yang sudah mendidik dan mengajarkan berbagai disiplin ilmu kepada penulis;
9. Kedua Orang Tua saya yang tercinta, atas segala kasih sayang, jerih payah dan doa restunya dalam membesarkan dan mendidik penulis

10. Rekan-rekan seperjuangan yang telah banyak memberikan bantuan dan moril yang sangat besar kepada penulis
11. Kepada semua pihak yang ikut membantu dalam penyelesaian Skripsi ini yang tak sempat penulis sebutkan satu-persatu.

Semoga Allah SWT, melimpahkan balasan atas jasa-jasa mereka kepada kami. Penulis memahami sepenuhnya bahwa dalam Skripsi ini masih masih terdapat kekurangan. Sebab itu, penulis sangat mengharapkan kritik dan saran yang konstruktif. Akhirnya penulis berharap semoga hasil yang telah dicapai ini dapat bermanfaat bagi kita semua. Aamiinn.

Gorontalo, Februari 2024

Penulis

ABSTRACT

AHMADNUSI.T3117107.CLASSIFICATIONOFSTUDENTS'FINALAS SIGNMENTTITLESUSING KNN

This study discusses the application of the K-Nearest Neighbor (KNN) algorithm in classifying students' final assignment titles. The dataset consists of 1000 tweets obtained through a crawling process using PHP. The preprocessing process includes case folding and stop words removal. After preprocessing, word weighting is carried out using the Term Frequency-Inverse Document Frequency (TF-IDF) method. The data is divided into training and test data with a ratio of 80:20. The KNN algorithm is applied to classify data based on the calculated word weights. The evaluation results show a Fairly High level of accuracy, proving that KNN effectively classifies tweet sentiment. This study contributes to the field of text mining and sentiment analysis, especially in the use of the KNN algorithm for text data in the Indonesian language.



Keywords: sentiment classification, KNN algorithm, text mining, TF-IDF, preprocessing, sentiment analysis, tweets, Indonesian language

ABSTRAK

AHMADNUSI. T3117107 .KLASIFIKASI JUDUL TUGAS AKHIR
MAHASISWA MENGGUNAKAN KNN

Penelitian ini membahas penerapan algoritma *K-Nearest Neighbor* (KNN) dalam klasifikasi judul tugas akhir mahasiswa. Dataset yang digunakan terdiri dari 1000 tweet yang diperoleh melalui proses *crawling* menggunakan PHP. Proses *reprocessing* meliputi *case folding* dan *stop words removal*. Setelah *preprocessing*, dilakukan pembobotan kata menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Data kemudian dibagi menjadi data latih dan data uji dengan perbandingan 80:20. Algoritma KNN diterapkan untuk mengklasifikasi data berdasarkan bobot kata yang telah dihitung. Hasil evaluasi menunjukkan tingkat akurasi yang cukup tinggi, membuktikan bahwa KNN efektif dalam klasifikasi *sentiment tweet*. Penelitian ini memberikan kontribusi dalam bidang *text mining* dan analisis sentimen, khususnya dalam penggunaan algoritma KNN untuk data teks dalam bahasa Indonesia.

Kata kunci: klasifikasi sentimen, algoritma KNN, *text mining*, TF-IDF, *preprocessing*, analisis sentimen, *tweet*, bahasa Indonesia



DAFTAR ISI

PERSETUJUAN SKRIPSI.....	Error! Bookmark not defined.
PENGESAHAN SKRIPSI.....	iii
PERNYATAAN SKRIPSI.....	iv
KATA PENGANTAR.....	v
ABSTRACT.....	vii
ABSTRAK.....	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xi
DAFTAR TABEL.....	xiii
BAB I PENDAHULUAN.....	1
1.1Latar Belakang.....	1
1.2Identifikasi Masalah	4
1.3Rumusan Masalah	4
1.4Tujuan Penelitian	4
1.5Manfaat Penelitian.....	4
1.5.1Manfaat Teoritis.....	5
1.5.2Manfaat Praktis.....	5
BAB II LANDASAN TEORI	6
1.1Tinjauan Studi	6
1.2Tinjauan Pustaka	6
2.2.4Text Mining	8
2.2.5Algoritma KNN.....	8
2.2.6Penerapan Algoritma KNN	8
ProsesKNN.....	11
7.2.4Perangkat Lunak.....	13

7.3Kerangka Pikir	14
BAB IIIMETODE PENELITIAN	18
3.1 Jenis, Metode, Subjek, Objek, Waktu dan Lokasi Penelitian	18
3.2Pengumpulan Data.....	18
3.3Pemodelan / Abstraksi	19
BAB IV HASIL PENELITIAN.....	20
4.1 Hasil Pengumpulan Data.....	20
4.2 Pra processing.....	21
4.2.1 Import Library.....	21
4.2.2 Mengimport Dataset.....	22
4.2.3 Case Folding	22
4.2.4 Tokenisasi	24
4.2.5 Stopword Removal	26
4.2.6 Stemming	28
4.3 Pembobotan TF-IDF	30
4.4 Klasifikasi	32
4.4.1 Pembagian Data Training dan Testing	32
BAB V PEMBAHASAN.....	34
5.1 Evaluasi.....	34
5.2 Pembahasan Hasil Pemodelan	34
5.3 Aplikasi Hasil Pemodelan	34
5.3.1 Halaman Data View Dataset.....	34
5.3.2 Halaman Casefolding.....	35
5.3.3 Halaman Tokenisasi.....	36
5.3.4 Halaman Stopword Removal.....	36
5.3.5 Halaman Steamiming.....	37
5.3.6 Halaman TF-IDF.....	37
5.3.7 Halaman Model Pengujian.....	38
BAB VI PENUTUP.....	39
6.1 Kesimpulan.....	39
6.2 Saran.....	39

DAFTAR PUSTAKA.....	40
----------------------------	-----------

DAFTAR GAMBAR

Gambar 2. 1 Proses <i>Knowledge Discovery</i> in Database (KDD) [6]	7
Gambar 2. 4 Kerangka Pikir.....	14
Gambar 3. 1 Model yang Diusulkan.....	19

DAFTAR TABEL

Tabel 1. 1 Data brita	3
Tabel 2. 1 Referensi Penelitian.....	6
Tabel 2. 11 Perangkat Lunak Pendukung	13

BAB I

PENDAHULUAN

1.1 Latar Belakang

Tugas akhir (Skripsi) merupakan syarat seorang mahasiswa sebagai tahapan dalam penyelesaian program sarjana (S1), yaitu berupa karya ilmiah yang dibuat oleh mahasiswa. Karya ilmiah yang disusun biasanya mengacu pada penulisan akademik dan analisis, fokus pada bidang studi tertentu. Penelitian yang dilakukan mahasiswa yaitu menemukan permasalahan atau hipotesa pada kasus tertentu kemudian menerapkan sebuah metode yang cocok untuk menyelesaikan permasalahan yang ditemukan atau sebagai pembuktian diterima atau ditolaknya hipotesis yang diajukan. Di tiap tahunnya penelitian atau tugas akhir kian bertambah dari sekian banyaknya judul skripsi memungkinkan penelitian hamper serupa ataupun sama, maka dari itu diperlukan adanya sebuah pengelompokan judul untuk mengetahui sebaran judul agar mudah mengetahui kelas-kelas setiap judul.[1]

Masalah Klasifikasi judul tugas akhir menerangkan tentang permasalahan yang terjadi dalam klasifikasi judul tugas akhir. Masalah tersebut bisa berupa kekurangan sistem klasifikasi judul tugas akhir yang ada, keterbatasan dalam mengelompokkan judul tugas akhir ke dalam kategori yang cocok, atau bahkan masalah-masalah lain yang berkaitan dengan klasifikasi judul tugas akhir seluruh Indonesia. Penjelasan mengenai masalah tersebut akan membantu dalam menyusun tujuan dan manfaat dari penelitian ini. Misalnya, masalah klasifikasi judul tugas akhir di Indonesia yang sering terjadi adalah kesulitan dalam mengkategorikan judul tugas akhir ke dalam kategori yang tepat karena banyaknya judul yang bermunculan di setiap tahunnya. Hal ini dapat menyebabkan salah klasifikasi dan mengurangi keakuratan dari sistem klasifikasi judul. Oleh karena itu, diperlukan metode yang lebih efektif dalam mengelompokkan judul tugas akhir ke dalam kategori yang sesuai.

Tujuan dari penelitian ini adalah untuk mengelompokkan judul tugas akhir khususnya di fakultas ilmu komputer berdasarkan kategori menggunakan

metode KNN (K-Nearest Neighbors). Metode ini digunakan karena kemampuannya dalam mengelompokkan judul tugas akhir ke dalam kategori yang sesuai dengan akurasi yang tinggi. Dengan menggunakan metode KNN, diharapkan dapat meningkatkan keakuratan dalam klasifikasi judul tugas akhir. Selain itu, tujuan dari penelitian ini juga adalah untuk mengidentifikasi kelas-kelas yang sering diangkat dalam penulisan tugas akhir. Hasil dari penelitian ini diharapkan dapat memberikan informasi yang bermanfaat bagi para akademis khususnya peneliti dalam menentukan judul tugasakhir sebagai klompok refrensi yang diinginkan.. Keunggulan dari metode KNN dibuktikan dari peneliti sebelumnya yaitu penelitian yang dilakukan oleh Nengah Widya Utami,2022 dengan judul text mining dalam analisis sentimen pembelajaran daring di masa pandemi covid 19 menggunakan algoritma k-nearest neighbor hasil penelitiannya dengan Nilai akurasi KNN sebesar 82.48% yang berarti KNN dapat dikatakan akurat dalam melakukan analisis sentimen.[2]penrlitian yang dilakukan oleh Andi Ahmad Irfa 2018, dengan judul penelitiannya Klasifikasi Topik Judul tugas akhir Berbahasa Indonesia Menggunakan *k-Nearest Neighbor* dengan menghasilkan performa micro average f1-measure sebesar 69,9%.[3], dan penelitian yang dilakukan oleh Retno Sari, 2020 Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi menggunakan Algoritma K-Nearest Neighbor (K-NN), Eksperimen telah dilakukan dengan mengubah nilai k dan didapat akurasi yang terbesar dengan nilai k=7 dengan akurasi 77.01, precission 92.38, recall 61.56 dan nilai AUC 0.894.[4]

Tabel 1. 1Data brita

No	Judul Tugas Akhir	Kelas
1	Analisis dan Perancangan Sistem Informasi Geografis untuk Pemetaan Jalan Rusak di Kota XYZ	GIS
2	Implementasi Algoritma Genetika pada Optimasi Penjadwalan Produksi di Perusahaan ABC	SPK
3	Pengembangan Aplikasi Pemilihan Jurusan Berbasis Sistem Pakar	SISTEM PAKAR
4	Analisis dan Perancangan Sistem Informasi Penjualan Online pada Toko Komputer XYZ	SISFO
5	Implementasi Metode Collaborative Filtering pada Sistem Rekomendasi Film	SPK
6	Pengembangan Aplikasi Deteksi Emosi Berbasis Pengolahan Citra untuk Mendukung Penyakit Mental	Image procesing
7	Analisis Pengaruh Teknologi Blockchain pada Keamanan Data pada Sistem Informasi Perbankan	Networking
8	Penerapan Teknologi Virtual Reality pada Sistem Pembelajaran Berbasis Gamifikasi	Gaming
9	Implementasi Metode Support Vector Machine pada Klasifikasi Citra Digital untuk Deteksi Penyakit Mata	Image procesing
10	Analisis Pengaruh Penggunaan Internet of Things pada Pengembangan Smart Home	Networking
11	Penerapan Teknologi Big Data pada Prediksi Harga Saham	Datamining
12	Analisis dan Perancangan Sistem Informasi Pengelolaan Inventaris pada Gudang Perusahaan ABC	SISFO
13	Implementasi Metode Naive Bayes pada Sistem Deteksi Spam Email	Datamining
14	Pengembangan Aplikasi Pendeteksi Kecurangan pada Ujian Online Berbasis Citra	Image procesing
15	Penerapan Teknologi Artificial Intelligence pada Prediksi Cuaca	Datamining
16	Implementasi Metode K-Means pada Sistem Pengelompokan Data Pelanggan pada Toko Online	Datamining
17	Pengembangan Aplikasi Pencarian Rute Tercepat pada Transportasi Umum Berbasis Geolokasi	GIS
18	Penerapan Teknologi Image Processing pada Pengenalan Wajah untuk Sistem Keamanan	Image procesing
19	Implementasi Metode Decision Tree pada Sistem Identifikasi Penyakit Pada Tanaman Padi	SISTEM PAKAR
20	Pengembangan Aplikasi Pendeteksi Kecelakaan Lalu Lintas Berbasis Kamera CCTV	Networking
21	Penerapan Teknologi Artificial Intelligence pada Sistem Identifikasi Kendaraan Bermotor	Image procesing
22	Implementasi Metode Fuzzy Logic pada Sistem Pengambilan Keputusan Perizinan Mendirikan Bangunan	SPK
23	Implementasi Metode Support Vector Regression pada Prediksi Konsumsi Listrik di Perkantoran	Datamining
24	Implementasi Metode Apriori pada Sistem Rekomendasi Produk pada Marketplace Online	SPK
25	Pengembangan Aplikasi Pendeteksi Serangan Phishing pada Email Berbasis Machine Learning	Networking

<https://rivierapublishing.id/blog/contoh-judul-skripsi> (20/3/2024)

Variabel yang digunakan pada penelitian ini yaitu judul tugas akhir (Dependent Variable), yakni data-data judul tugas akhir yang akan dikelompokkan ke dalam kategori yang sesuai dan Kategori (Independent Variable), yaitu kategori yang akan diterapkan pada tiap judul tugas akhir untuk mengelompokkannya.

Data kumpulan tugas akhir merupakan data yang dibutuhkan dalam penelitian ini. Data-data tersebut akan dipreprocessing sebelum diproses dengan metode KNN untuk proses klasifikasi judul tugas akhir. Disamping itu peneliti juga membutuhkan data kategori asli dari setiap judul tugas akhir yang nantinya akan digunakan sebagai patokan dalam mengevaluasi akurasi dari hasil klasifikasi judul tugas akhir menggunakan metode KNN. Jenis data yang digunakan adalah data kuantitatif yang akan diolah secara numerik dan dianalisis menggunakan teknik statistik.

Berdasarkan penjelasan latar belakang maka peneliti akan melakukan penelitian dengan judul **“Klasifikasi Judul Tugas Akhir Mahasiswa Menggunakan K-Nearest Neighbor”**.

1.2 Identifikasi Masalah

Judul tugas akhir kian tahun kian bertambah. Masalah tersebut bisa berupa kekurangan sistem klasifikasi judul tugas akhir yang ada, keterbatasan dalam mengkategorikan judul tugas akhir ke dalam kategori yang sesuai.

1.3 Rumusan Masalah

Berdasarkan identifikasi masalah diatas, maka permasalahannya dirumuskan yaitu bagaimana hasil analisa pengelompokan judul tugas akhir dengan menerapkan metode KNN?

1.4 Tujuan Penelitian

Berdasarkan Rumusan permasalahannya diatas, maka tujuan dari penelitian ini yaitu untuk mengetahui hasil analisa pengelompokan judul tugas akhir dengan menerapkan metode KNN.

1.5 Manfaat Penelitian

Penelitian ini diharapkan mempunyai manfaat yaitu:

1.5.1 Manfaat Teoritis

Penelitian ini diharapkan dapat memberikan sumbangan pengembangan ilmu bidang kajian data mining tentang kemampuan metode KNN dalam penerapan analisis teks mining.

1.5.2 Manfaat Praktis

Hasil dari penelitian ini diharapkan dapat digunakan sebagai *alternative* atau masukan kepada semua pihak yang berkepentingan khususnya dalam menganalisa kelompok judul tugas akhir.

BAB II

LANDASAN TEORI

1.1 Tinjauan Studi

Adapun penelitian terdahulu yang berhubungan dengan tema peneliti yang dijadikan sebagai bahan referensi dalam menentukan metode yang akan digunakan.

Tabel 2. 1Referensi Penelitian

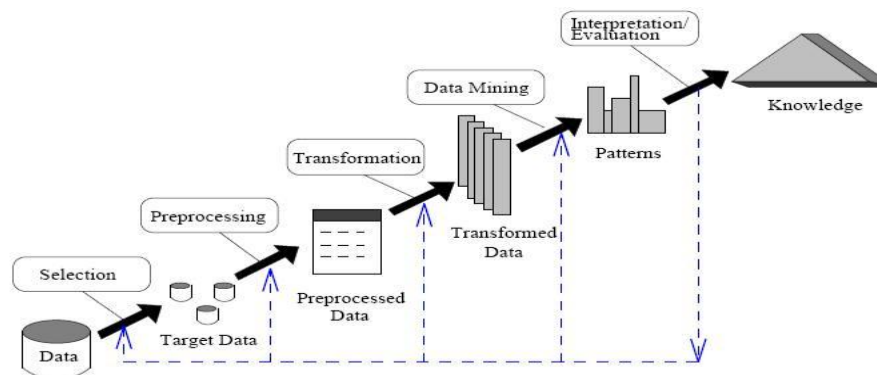
No	Nama Peneliti	Judul Penelitian	Deskripsi Singkat
1.	Nengah Widya Utami,2022[2]	Text Mining Dalam Analisis Sentimen Pembelajaran Daring Di Masa Pandemi	Hasil penelitiannya dengan Nilai akurasi KNN sebesar 82.48% yang berarti KNN dapat dikatakan akurat dalam
2	Andi Ahmad Irfa 2018[3]	Klasifikasi Topik Judul tugas akhir Berbahasa Indonesia Menggunakan k-	menghasilkan performa micro average f1-measure sebesar 69,9%
3.	Retno Sari, 2020[4]	Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi menggunakan Algoritma K-Nearest Neighbor (K-NN)	Eksperimen telah dilakukan dengan mengubah nilai k dan didapat akurasi yang terbesar dengan nilai k=7 dengan akurasi 77.01, precission 92.38, recall 61.56 dan nilai AUC 0.894.

1.2 Tinjauan Pustaka

1.2.1 Data Mining

Data Mining diartikan sebagai proses untuk memperoleh informasi yang bermanfaat dari gudang basis data yang besar. Data mining juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar

yang membantu dalam pengambilan keputusan. Istilah data mining kadang disebut juga *Knowledge Discovery* (Tan dalam Eko Prasetyo). Menurut Daryl Pregibon menyatakan bahwa “Data mining adalah campuran statistik, kecerdasan buatan, dan riset basis data” [5].



Gambar 2. 1Proses *Knowledge Discovery* in Database (KDD)[6]

Data mining merupakan inti dari proses *Knowledge Discovery* in Database (KDD). KDD adalah proses terorganisir untuk mengidentifikasi pola yang valid, baru, berguna, dan dapat dimengerti dari sebuah dataset yang besar dan kompleks [5].

1.2.2 Fungsi Data Mining

Penambangan data sebenarnya mempunyai banyak sekali manfaat atau fungsi. Namun, secara umum, terdapat 2 fungsi data mining yaitu: [7]

1. Deskriptif: Kapasitas yang berbeda mengacu pada kapasitas dalam memahami informasi yang terkonsentrasi lebih lanjut. Melalui siklus ini, adalah normal bahwa perilaku informasi dapat diketahui. Informasi ini nantinya dapat dimanfaatkan untuk mengungkap kualitas informasi yang ideal. Dengan menggunakan kapasitas ilustratif dari penggalian informasi, contoh tertentu yang awalnya tersembunyi dalam suatu informasi dapat ditemukan. Dengan demikian, dengan asumsi disadari bahwa ada contoh yang suram dan memiliki harga diri, itu menyiratkan bahwa atribut informasi dapat diketahui.
2. Prediktif: Fungsi Prediktif dari data mining adalah elemen interaksi yang pada akhirnya akan mengungkap contoh informasi tertentu. Contoh ini

harus terlihat dari beberapa faktor yang terkandung dalam informasi. Setiap kali sebuah contoh ditemukan, contoh tersebut dapat digunakan untuk mengukur berbagai faktor yang nilai atau jenisnya pada saat ini belum jelas. Oleh karena itu, fungsi prediktif dipandang sebagai fungsi prediktif yang sebanding dengan analisis prediktif. Kemampuan ini juga dapat digunakan untuk menilai variabel luar biasa yang tidak ada dalam informasi. Selanjutnya, kapasitas ini dapat bekerja dengan dan menguntungkan setiap individu yang membutuhkan perkiraan yang tepat untuk meningkatkan hal-hal yang signifikan.

2.2.4 Text Mining

Text mining ialah metode yang dapat digunakan untuk melakukan klasifikasi, clustering, information extraction, dan information retrieval[8]. Text mining adalah upaya dalam menemukan informasi agar pengguna dapat berinteraksi dengan sekumpulan dokumen dari masa ke masa dengan menggunakan seperangkat alat analisis [9]. Proses utama dari text mining ini yaitu untuk mendapatkan kata yang dapat menggantikan isi dari dokumen dan kemudian akan dilakukan proses pengkajian untuk melihat keterkaitan tiap dokumen.

2.2.5 Algoritma KNN

Algoritma KNN adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasi sebelumnya. Termasuk dalam supervised learning, dimana hasil query instance yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam KNN.[10]

2.2.6 Penerapan Algoritma KNN

Berikut adalah contoh penerapan Algoritma FP-Growth pada data yang penulis dapatkan[11].

Dataset

Dataset yang digunakan pada penelitian ini adalah tweet sebanyak 1000 tweet. Dataset didapatkan melalui proses crawling dengan menggunakan php. Dataset

diperlukan untuk membangun sistem klasifikasi menggunakan algoritma KNN. Berikut ini adalah sample dari dataset yang digunakan.

@jokowi DPR kerjanya apa sih selain nyinyir kinerja pemerintah tiap hari! #bubarkandpr
RT @ebithEBITH: Semoga Bapak Presiden bisa evaluasi kinerja aparat keamanan negeri ini... biar rakyat aman
RT @DiaZKillaZ: Disaat situasi kaya gni... Hari ini masi aja ada yg nyinyir2in pemerintahan, kaya yang udah plg tau aja dalem2an kinerja
RT @kaka_Diwan: Yth : Bpk Presiden Jokowi, terkait Kasus Mako Brimob dan Bom Surabaya serta bom sidoarjo yang baru terjadi malam ini. Ini
@SBYudhoyono Hanya orang paham dan cerdas bisa menilai kinerja mantan atau masih menjabat presiden
@jokowi Saya dukung 100% dgn dikeluarkannya Perpu ini jika memang dr DPR menghambat kinerja POLRI dlm memberantas

Preprocessing Dataset

Preprocessing adalah proses untuk mengolah data yang belum sesuai dengan bentuk data yang di perlukan untuk proses klasifikasi ini. Untuk tugas akhir ini preprocessing yang dipakai adalah case folding dan stopwords removal. Case folding untuk mengubah semua upper case menjadi lower case dan stopwords removal untuk menghapus kata, space dan karakter yang tidak dipakai. Berikut ini adalah data yang telah di preprocessing.

jokowi dpr kerjanya sih nyinyir kinerja pemerintah bubarkandpr
ebithebith semoga presiden evaluasi kinerja aparat keamanan negeri biar rakyat aman
segenap rakyat bersatu untuk Indonesia yang lebih baik
kaka_diwan yth bpk presiden jokowi terkait mako brimob bom surabaya bom sidoarjo malam ini...
sbyudhoyono orang paham cerdas menilai kinerja mantan menjabat presiden
jokowi dukung dgn dikeluarkannya perpu dr dpr menghambat kinerja polri dlm memberantas t... kredvfgv

Split Dataset

Tujuan dari proses data split adalah untuk mendapatkan data training dan data testing. Untuk pembagian data training dan testing penulis menggunakan 3 skenario yaitu 80%:20% , 70%:30% dan 60%:40%.

Feature Extraction

Pada tahapan feature extraction tweet yang telah di preprocessing diubah menjadi nilai. Untuk menjadikan tweet menjadi nilai maka feature extraction yang penulis gunakan adalah TF binary. Di dalam proses ini juga tweet yang telah preprocessing akan dipisah menjadi perkata untuk menjadi atribut di data training maupun data testing. Yang harus diperhatikan adalah atribut data training dan testing harus sama agar dapat dilakukan perhitungan jarak pada proses Distance KNN. Berikut adalah langkah dalam proses Feature Extraction. 1. Langkah pertama yang harus dilakukan adalah membuat atribut pada data training yang akan digunakan untuk proses TF binary. Atribut didapatkan dari kata-kata yang terdapat pada semua tweet data training. Tetapi tidak semua kata yang ada pada tweet dijadikan atribut, kata yang diambil adalah kata yang berhubungan dengan kinerja pemerintah dan juga kata yang memiliki makna diambil salah satu. Dibawah ini adalah contoh dari kata yang terdapat pada tweet yang dijadikan atribut.

Tabel sample atribut

No	Kata Unigram
1	Agama
2	maaf
3	Bahwa
4	Belum
5	Aman
6	Awal
7	Segenap
8	Rakyat
9	Indonesia
10	Gagal

2. Langkah kedua adalah menyamakan atribut antara data training dan data testing sehingga tidak terjadi error pada saat perhitungan jarak.

3. Langkah ketiga atau langkah terakhir adalah proses TF Binary. Di tahap ini record yang terdapat pada data training dan testing di cek satu persatu apakah sama dengan atribut yang telah dibuat. Dari Tabel data tweet dilihat apakah kata yang ada pada tweet tersebut ada di dalam atribut yang telah dibuat pada tabel sample atribut.

Tabel Data Tweet

Nomor	Tweet
1	mbahuyok maaf bahwa belum aman kepada segenap rakyat indonesia apapun suku agamanya
2	segenap rakyat bersatu untuk Indonesia yang lebih baik
3	sbyudhoyono orang paham cerdas menilai kinerja mantan menjabat presiden

Maka hasil yang didapatkan setelah dilakukan proses feature extraction dengan menggunakan TF Binary yang bila kata yang terdapat pada tweet ada maka bernilai 1 sedangkan kalau tidak ada bernilai 0. Hasil pada Tabel TF binary hanya sebagai contoh saja, sebenarnya atribut yang digunakan sebanyak 258 atribut.

Contoh TF binary

No	bahwa	belum	aman	segenap	rakyat	indonesia	apapun	suku	agamanya	akan	gagal
1	1	1	1	1	1	1	1	1	1	0	0
2	0	0	0	1	1	1	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0

ProsesKNN

Untuk proses ini data training yang telah melalui proses *preprocessing* dan *feature extraction* digunakan harus sudah mempunyai kelas. Selanjutnya setelah data *training* dan data *testing* telah siap, maka proses *Distance KNN* bisa dilakukan. Prosesnya adalah sebagai berikut:

- Proses pertama adalah mencari jarak antara *record* data *testing* ke *record* data *training* yang sudah berbentuk *vector* dengan menggunakan rumus *Euclidean Distance*. Rumus *Euclidean Distance* pada masalah ini adalah sebagaiberikut:

$$dist = \sqrt{(test(i,k) - train(j,k))^2}$$

Dimana $test(i,k)$ adalah seluruh *record* dan atribut yang ada pada data *testing* sedangkan $train(j,k)$ adalah seluruh *record* dan atribut yang ada pada data *testing*. **Tabel hasil proses 1-3** adalah tabel jarak dari *record tweet* pada tabel TF Binary ke setiap *record tweet* data *training* yang digunakan penulis.

- Yang kedua setelah mendapatkan jarak dengan menggunakan rumus *Euclidean Distance* selanjutnya adalah menentukan jumlah k atau

tetangga berdasarkan jarak pada **tabel Hasil dari proses 1-3** . K yang digunakan untuk contoh kali ini adalah 7, jadi dari **tabel Hasil dari proses 1-3** diambil jarak 7 terdekat berdasarkan jumlah k yang ditentukan.

5. Yang ketiga adalah mengurutkan hasil dari nomer 2 secara *ascending* yaitu mengurutkan jarak dari paling kecil ke jarak paling 7 berdasarkan tetangga yang telah ditentukan.

Hasil dari proses 1-3

0	0	0
0	0	0
0	0	0
0	0	2.645751
0	0	2.645751
0	0	2.645751
0	0	2.828427

6. Yang keempat setelah mendapatkan K yang telah diurutkan secara *ascending* selanjutnya yaitu melakukan klasifikasi. Tabel dibawah ini adalah tabel *tweet* dari **tabel Hasil dari proses 1-3**. Jadi untuk data *testing record tweet* 1 dan 2 jarak terdekatnya dengan data *training* adalah *record tweet* 1,2,4,6,7,8 pada data *training* sedangkan untuk data *testing record tweet* 3 jarak terdekatnya adalah dengan 3, 203, 403, 68, 268, 486,5.

Tabel *tweet* berdasarkan jarak

1	1	3
2	2	203
4	4	403
6	6	68
7	7	268
8	8	468
10	10	5

Kelas data *testing*

<i>Record Tweet</i>	Kelas
1	Postif
2	Positif
3	Positif

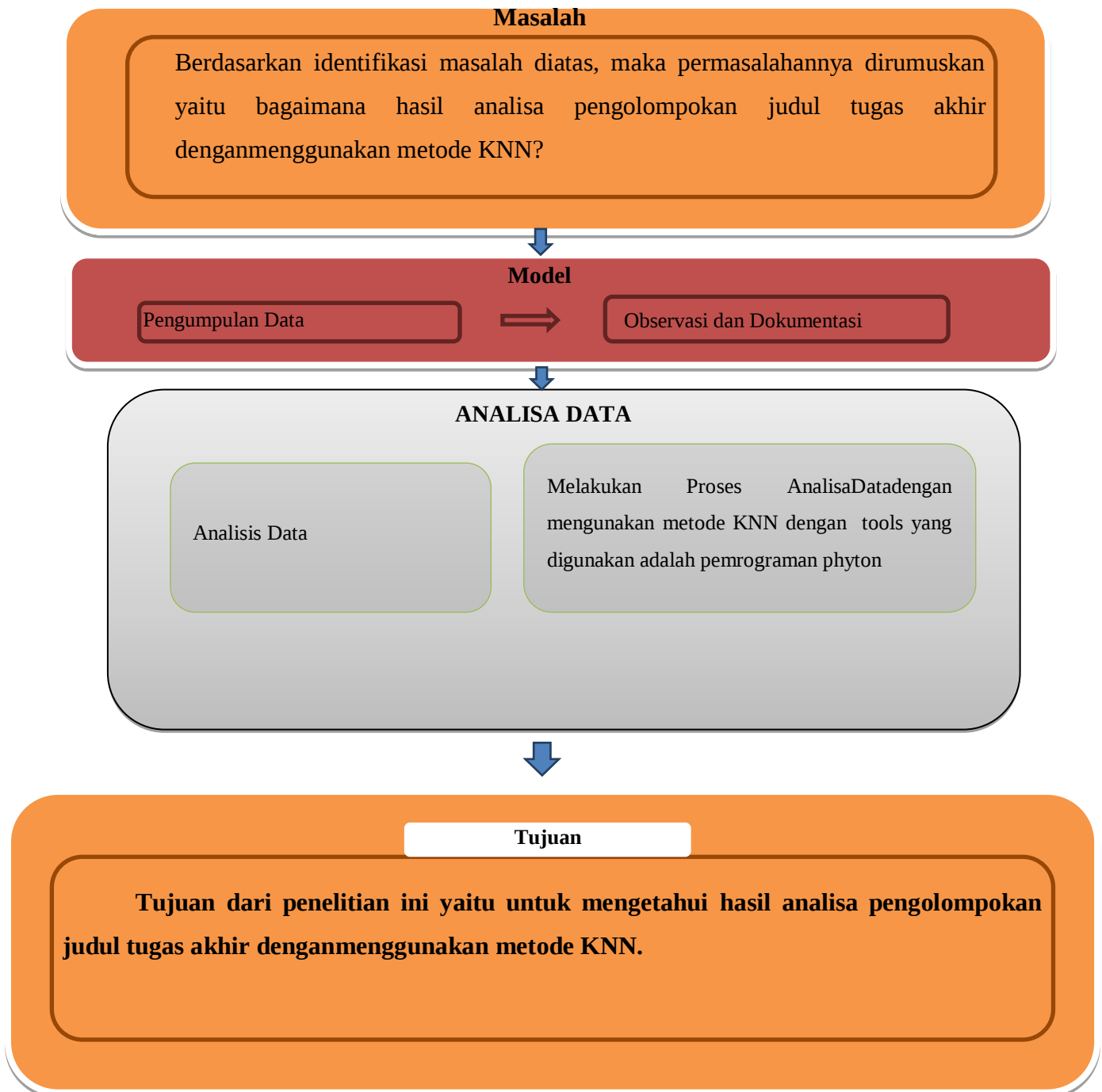
7. Setelah mendapatkan *record tweet testing* berdekatan dengan *record tweet training* selanjutnya yaitu melakukan klasifikasi. Proses klasifikasinya yaitu dengan mengikuti kelas mayoritas dari K sebelumnya. Tabel **Kelas data testing** adalah tabel setelah dilakukan pengujian dengan 1 yang telah ditentukan sebelumnya yaitu 7 didapatkan untuk *record tweet* 1, 2, 3 data *testing* kelasnya adalah kelas bagus karena dari 7 *tweet* pada Tabel **Tweet berdasarkan jarak** mayoritasnya kelas bagus.

7.2.4 Perangkat Lunak

Tabel 2. 2Perangkat Lunak Pendukung

No	Perangkat Lunak Pendukung	Berfungsi
1	Phyton	Suatu Bahasa Pemograman yang digunakan untuk melakukan olah dan analisa data

7.3 Kerangka Pikir



Gambar 2.2Kerangka Pikir

BAB III

METODE PENELITIAN

3.1 Jenis, Metode, Subjek, Objek, Waktu dan Lokasi Penelitian

Penelitian ini menggunakan metode penelitian studi kasus media judul tugas akhir. Dengan demikian jenis penelitian ini adalah kuantitatif. Subjek penelitian ini adalah Analisa pengelompokan judul tugas akhir Menggunakan Metode KNN. Penelitian ini dimulai dari Agustus 2023 – November 2023 mencakup pengumpulan dan pengolahan data beserta bimbingan.

3.2 Pengumpulan Data

1. Data Sekunder

Data Sekunder adalah data yang diambil dari studi literatur yang dilakukan. Sumber dari studi literatur yaitu jurnal. Makalah ilmiah atau buku yang membahas tentang Metode Algoritma KNN.

2. Data Primer

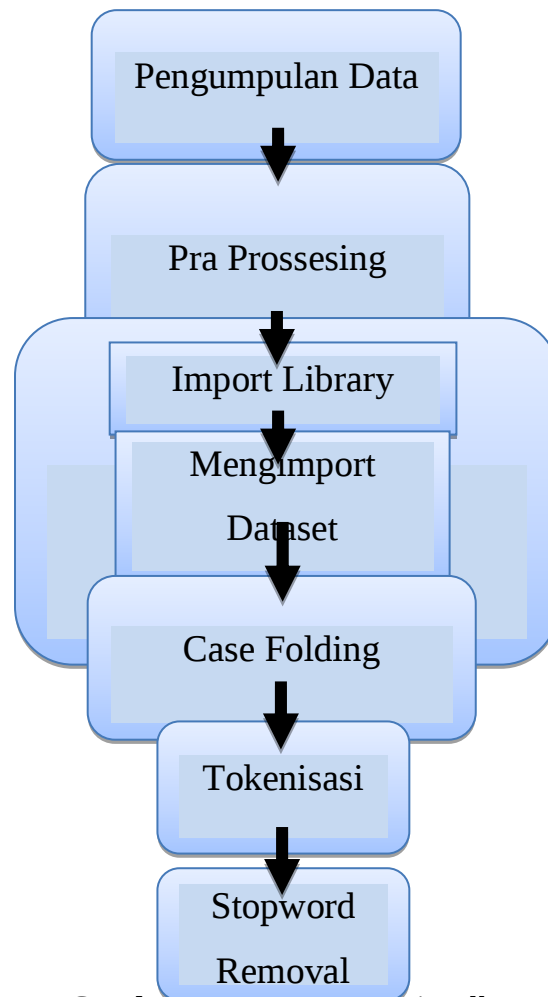
Untuk memperoleh data primer yang merupakan data lapangan atau langsung pada objek penelitian di media pem judul tugas akhiran.

Teknik yang digunakan :

- a. Observasi adalah meninjau atau mengamati secara langsung. Adapun pada penelitian ini dilakukan dengan cara mengumpulkan data judul tugas akhir melalui beberapa media pem judul tugas akhiran.

3.3 Pemodelan / Abstraksi

Berikut adalah suatu permodelan:



Gambar 3.1 Model yang Diusulkan

BAB IV HASIL PENELITIAN

4.1 Hasil Pengumpulan Data

Pada penelitian ini, dilakukan pengumpulan data dari Fakultas Ilmu Komputer Prodi Teknik Informatika Universitas Ichsan Gorontalo yang diakses melalui siakun, guna mendukung terlaksananya penelitian yang dilakukan. Adapun data judul yang saya dapat berjumlah 118 data judul, yang terdiri dari Berikut ini data yang ditunjukkan pada Tabel 4.1

Tabel 4.1 Hasil Pengumpulan Data

No	Judul	Kategori
1	Sistem pendukung keputusan penentuan keikutsertaan prioritas	SPK

No	Judul	Kategori
	program kb menggunakan metode analytical herarchy process	
2	Analisis sentimen opini public pengguna twitter terhadap kenaikan harga bbm menggunakan algoritma naïve bayes	DATA MINING
3	Rancang bangun web service sistem informasi berkas administrasi tugas akhir teknik informatika (SIABTI) berbasis progressive web app (android)	SISFO
4	Aplikasi E-learning pembelajaran sekolah menengah pertama negeri 2 satu atap bulango utara	SISFO
5	Klasifikasi malware menggunakan teknik machine learning	DATA MINING
6	Penerapan metode K-Means untuk clustering penjualan suku cadang kendaraan viar (STUDI KASUS CV. GOTAMA VIAR GORONTALO)	DATA MINING
7	Mobile administrasi desa di kantor desa kuala utara berbasis android	SISFO
8	Implementasi augmented reality dalam pengenalan jenis jenis perangkat keras computer berbasis android	IOT
9	Game pembelajaran bacaan AL-QUR'AN dengan menggunakan metode dirosa	GAME
10	Sistem pakar diagnosa kerusakan printer canon ip 2770 menggunakan metode case base reasoning	SISTEM PAKAR
...		
118	prediksi produksi udang vaname menggunakan metode regresi linier berganda (STUDI KASUS: DINAS PERIKANAN DAN KELAUTAN PROVINSI GORONTALO)	DATA MINING

Sumber : Fakultas Ilmu Komputer

4.2 Pra Prosessing

Preprocessing dilakukan setelah dataset siap untuk diolah. Pada tahapan preprocessing terdapat beberapa proses yang akan dilakukan sehingga data menjadi terstruktur dengan tahapan Import Library, Import Dataset, Case folding, Tokenizing, Stopword removal, Stemming, dan Normalization.

4.2.1 Import Library

Sebelum dilakukan proses pra prosessing, langkah pertama yang dilakukan adalah memanggil library yang akan digunakan seperti skrip berikut:

```
!pip install gradio
```

```

!pip install Sastrawi
!pip install xlswriter
!pip install imblearn
import string
import gradio as gr
from sklearn.pipeline import Pipeline # Fixed typo here
from imblearn.over_sampling import SMOTE
import pandas as pd
import numpy as np
import re
import nltk
from nltk.corpus import stopwords

from nltk.tokenize import word_tokenize
from nltk.stem import PorterStemmer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.neighbors import KNeighborsClassifier

import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
from sklearn.manifold import TSNE

```

4.2.2 Mengimport Dataset

Selanjutnya akan melakukan proses import dataset yang akan digunakan seperti skrip berikut

```

# Load the dataset
df = pd.read_csv('dataset.csv', sep=';', encoding='latin-1')
print("Dataset awal:")
# print(df.head(100))
df.head(118)

```

4.2.3 Case Folding

Pada proses ini dilakukan pengubahan huruf besar menjadi huruf kecil pada semua judul pada data yang dikumpulkan serta membersihkan karakter-karakter seperti ^a – zA – Z0 – 9\s. Berikut adalah skrip yang digunakan untuk case folding.

```

# Define the case folding function
def casefolding(Judul):
    Judul = Judul.lower()
    Judul = Judul.strip()
    Judul = re.sub(r'[^a-zA-Z0-9\s]', '', Judul)
    return Judul

# Apply case folding to the 'Judul' column
df['casefolded_Judul'] = df['Judul'].apply(casefolding)

```

```
# print("\nDataset setelah case folding:")
# print(df[['Judul', 'Hasil Case Folding']].head(118))
df.head(118)
```

Setelah dijalankan skrip yang digunakan untuk proses case folding, maka akan menampilkan data seperti berikut ini :

Tabel 4.2 Hasil Proses Case Folding

No	Judul	Hasil Case Folding
1	Sistem pendukung keputusan penentuan keikutsertaan prioritas program kb menggunakan metode analytical herarchy process	sistem pendukung keputusan penentuan keikutsertaan prioritas program kb menggunakan metode analytical herarchy process
2	Analisis sentimen opini public pengguna twitter terhadap kenaikan harga bbm menggunakan algoritma naïve bayes	analisis sentimen opini public pengguna twitter terhadap kenaikan harga bbm menggunakan algoritma nave bayes
3	Rancang bangun web service sistem informasi berkas administrasi tugas akhir teknik informatika (SIABTI) berbasis progressive web app (android)	rancang bangun web service sistem informasi berkas administrasi tugas akhir teknik informatika siabti berbasis progressive web app android
4	Aplikasi E-learning pembelajaran sekolah menengah pertama negeri 2 satu atap bulango utara	aplikasi elearning pembelajaran sekolah menengah pertama negeri 2 satu atap bulango utara
5	Klasifikasi malware menggunakan teknik machine learning	klasifikasi malware menggunakan teknik machine learning
6	Penerapan metode K-Means untuk clustering penjualan suku cadang kendaraan viar (STUDI KASUS CV. GOTAMA VIAR GORONTALO)	penerapan metode kmeans untuk clustering penjualan suku cadang kendaraan viar studi kasus cv gotama viar gorontalo
7	Mobile administrasi desa di kantor desa kuala utara berbasis android	mobile administrasi desa di kantor desa kuala utara berbasis android
8	Implementasi augmented reality dalam pengenalan jenis jenis perangkat keras computer berbasis android	implementasi augmented reality dalam pengenalan jenis jenis perangkat keras computer berbasis android
9	Game pembelajaran bacaan AL-QUR'AN dengan menggunakan metode dirosa	game pembelajaran bacaan alquran dengan menggunakan metode dirosa

No	Judul	Hasil Case Folding
10	Sistem pakar diagnosa kerusakan printer canon ip 2770 menggunakan metode case base reasoning	sistem pakar diagnosa kerusakan printer canon ip 2770 menggunakan metode case base reasoning
...	...	...
118	prediksi produksi udang vaname menggunakan metode regresi linier berganda (STUDI KASUS: DINAS PERIKANAN DAN KELAUTAN PROVINSI GORONTALO)	prediksi produksi udang vaname menggunakan metode regresi linier berganda studi kasus dinas perikanan dan kelautan provinsi gorontalo

4.2.4 Tokenisasi

Selanjutnya adalah proses pemisahan kalimat (string) yang setiap kalimatnya disusun menjadi potongan tunggal. Berikut ini adalah sourcecode python pada google colab untuk proses Tokenisasi,

```
def tokenization(Judul):
    tokens = word_tokenize(Judul)
    return tokens
df['tokens'] = df['casefolded_Judul'].apply(tokenization)
df.head(118)
```

Setelah dijalankan skrip yang digunakan untuk proses case tokenisasi, maka akan menampilkan data seperti berikut ini

Tabel 4.3 Hasil Proses Case Tokenisasi

No	Hasil Case Folding	Tokenisasi
1	sistem pendukung keputusan penentuan keikutsertaan prioritas program kb menggunakan metode analytical herarchy process	[sistem,pendukung,keputusan,penentuan,keikutsertaan,prioritas,program,kb,menggunakan,metode,analytical,herarchy,process]
2	analisis sentimen opini public pengguna twitter terhadap kenaikan harga BBM menggunakan algoritma nave bayes	[analisis,sentimen,opini,public,pengguna,twitter,terhadap,kenaikan,harga,bbm,menggunakan,algoritma,nave,bayes]

No	Hasil Case Folding	Tokenisasi
3	rancang bangun web service sistem informasi berkas administrasi tugas akhir teknik informatika siabti berbasis progressive web app android	[rancang,bangun,web,service,sistem,informasi,berkas,administrasi,tugas,akhir,teknik,informatika,siabti,berbasis,progressive,web,app,android]
4	aplikasi elearning pembelajaran sekolah menengah pertama negeri 2 satu atap bulango utara	[aplikasi,elearning,pembelajaran,sekolah,menengah,pertama,negeri,2,satu,atap,bulango,utara]
5	klasifikasi malware menggunakan teknik machine learning	[klasifikasi,malware,menggunakan,teknik,machine,learning]
6	penerapan metode kmeans untuk clustering penjualan suku cadang kendaraan viar studi kasus cv gotama viar gorontalo	[penerapan,metode,kmeans,untuk,clustering,penjualan,suku,cadang,kendaraan,viar,studi,kasus,cv,gotama,viar,gorontalo]
7	mobile administrasi desa di kantor desa kuala utara berbasis android	[mobile,administrasi,desa,di,kantor,desa,kuala,utara,berbasis,android]
8	implementasi augmented reality dalam pengenalan jenis jenis perangkat keras computer berbasis android	[implementasi,augmented,reality,dalam,penge-nalan,jenis,jenis,perangkat,keras,computer,berbasis,android]
9	game pembelajaran bacaan alquran dengan menggunakan metode dirosa	[game,pembelajaran,bacaan,alquran,dengan,menggunakan,metode,dirosa]
10	sistem pakar diagnosa kerusakan printer canon ip 2770 menggunakan metode case base reasoning	[sistem,pakar,diagnosa,kerusakan,printer,canon,ip,2770,menggunakan,metode,case,base,reasoning]
...	...	...
118	prediksi produksi udang vaname	[prediksi,produksi,udang,vaname,menggunakan,me

No	Hasil Case Folding	Tokenisasi
	menggunakan metode regresi linier berganda studi kasus dinas perikanan dan kelautan provinsi gorontalo	tode,regresi,linier,berganda,studi,kasus,dinas,perikanan,dan,kelautan,provinsi,gorontalo]

4.2.5 Stopword Removal

Pada proses ini terjadi penghilangan kata yang tidak berkaitan dengan topik dokumen. Berikut ini adalah sourcecode python pada google colab untuk proses Stopword removal

#Stemming

```
# Define stemming function
def stemming(removal):
    porter = PorterStemmer()
    stemmed_tokens = [porter.stem(token) for token in removal]
    return stemmed_tokens

# Apply stemming to the 'tokens' column
df['stemmed_tokens'] = df['removal'].apply(stemming)

# Display the final table with relevant columns
print("\nTabel Hasil Case Folding, Tokenisasi, removal, dan Stemming:")
result_df = df[['Judul', 'casfolding', 'tokens', 'removal', 'stemmed_tokens']]
result_df.head(118)
```

Dari proses stopword diatas diperoleh hasil data yang telah dihilangkan karena tidak berkaitan dengan topik. Berikut adalah tabel hasil stopword dari dari beberapa data yang disimpan.

Tabel 4.4 Hasil Proses Case Stopwords Removal

No	Tokenisasi	Removal
1	[sistem,pendukung,keputusan,penentuan,k eikutsertaan,prioritas,program,kb,menggun akan,metode,analytical,herarchy,process]	[sistem,pendukung,keputusan,penentuan,keik utsertaan,prioritas,program,kb,metode,analyt ical,herarchy,process]
2	[analisis,sentimen,opini,public,pengguna,tw itter,terhadap,kenaikan,harga,bbm,menggu nakan,algoritma,nave,bayes]	[analisis,sentimen,opini,public,pengguna,twitt er,kenaikan,harga,bbm,algoritma,nave,bayes]
3	[rancang,bangun,web,service,sistem,inform	[rancang,bangun,web,service,sistem,informas

No	Tokenisasi	Removal
	asi,berkas,administrasi,tugas,akhir,teknik,informatika,siabti,berbasis,progressive,web,app,android]	i,berkas,administrasi,tugas,teknik,informatika,siabti,berbasis,progressive,web,app,android]
4	[aplikasi,elearning,pembelajaran,sekolah,menengah,pertama,negeri,2,satu,atap,bulango,utara]	[aplikasi,elearning,pembelajaran,sekolah,menengah,negeri,2,atap,bulango,utara]
5	[klasifikasi,malware,menggunakan,teknik,machine,learning]	[klasifikasi,malware,teknik,machine,learning]
6	[penerapan,metode,kmeans,untuk,clustering,penjualan,suku,cadang,kendaraan,viar,studi,kasus,cv,gotama,viar,gorontalo]	[penerapan,metode,kmeans,clustering,penjualan,suku,cadang,kendaraan,viar,studi,cv,gotama,viar,gorontalo]
7	[mobile,administrasi,desa,di,kantor,desa,kuala,utara,berbasis,android]	[mobile,administrasi,desa,kantor,desa,kuala,utara,berbasis,android]
8	[implementasi,augmented,reality,dalam,pengenalan,jenis,jenis,perangkat,keras,computer,berbasis,android]	[implementasi,augmented,reality,pengenalan,jenis,jenis,perangkat,keras,computer,berbasis,android]
9	[game,pembelajaran,bacaan,alquran,dengan,menggunakan,metode,dirosa]	[game,pembelajaran,bacaan,alquran,metode,dirosa]
10	[sistem,pakar,diagnosa,kerusakan,printer,canon,ip,2770,menggunakan,metode,case,base,reasoning]	[sistem,pakar,diagnosa,kerusakan,printer,canon,ip,2770,metode,case,base,reasoning]
...	...	...
118	[prediksi,produksi,udang,vaname,menggunakan,metode,regresi,linier,berganda,studi,kasus,dinas,perikanan,dan,kelautan,provinsi,gorontalo]	[prediksi,produksi,udang,vaname,metode,regresi,linier,berganda,studi,dinas,perikanan,kelautan,provinsi,gorontalo]

4.2.6 Stemming

Proses ini dilakukan penghapusan imbuhan dari setiap kata. Sehingga mendapatkan kata dasarnya. Berikut ini adalah sourcecode python pada google colab untuk proses Stemming

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
def stemming(removal):
    factory = StemmerFactory()
```

```

stemmer = factory.create_stemmer()
do = [] # Fixed indentation here
for w in removal:
    dr = stemmer.stem(w)
    do.append(dr)

d_cleane = []
d_cleane=" ".join(do)
print(d_cleane)
return d_cleane
df['stemming'] = df['removal'].apply(stemming)

# Assuming data_to_csv is defined elsewhere
df.to_csv('data_clean.csv', index=False)
data_clean=pd.read_csv('data_clean.csv', encoding='latin1')
data_clean.head(10)

```

Setelah proses steamming diatas diperoleh hasil data yang berbentuk kata dasar. Berikut adalah tabel hasil steamming dari dari beberapa data yang disimpan.

Tabel 4.5Hasil Proses Stemming

No	Removal	Stemming
1	[sistem,pendukung,keputusan,penentua n,keikutsertaan,prioritas,program,kb,met ode,analytical,herarchy,process]	[sistem, dukung, putus, tentu, keikutsertaan, prioritas, program, kb, metode, analytical, herarchy, process]
2	[analisis,sentimen,opini,public,pengguna, twitter,kenaikan,harga,bbm,algoritma,na ve,bayes]	analisis sentimen opini public guna twitter naik harga bbm algoritma nave bayes
3	[rancang,bangun,web,service,sistem,info rmasi,berkas,administrasi,tugas,teknik,in formatika,siabti,berbasis,progressive,we b,app,android]	rancang bangun web service sistem informasi berkas administrasi tugas teknik informatika siabti bas progressive web app android
4	[aplikasi,elearning,pembelajaran,sekolah, menengah,negeri,2,atap,bulango,utara]	aplikasi elearning ajar sekolah tengah negeri 2 atap bulango utara
5	[klasifikasi,malware,teknik,machine,learn ing]	klasifikasi malware teknik machine learning
6	[penerapan,metode,kmeans,clustering,p enjualan,suku,cadang,kendaraan,viar,stu di,cv,gotama,viar,gorontalo]	terap metode kmeans clustering jual suku cadang kendara viar studi cv gotama viar gorontalo

No	Removal	Stemming
7	[mobile,administrasi,desa,kantor,desa,kuala,utara,berbasis,android]	mobile administrasi desa kantor desa kuala utara bas android
8	[implementasi,augmented,realiti,penge-nalan,jenis,jenis,perangkat,keras,computer,berbasis,android]	implementasi augmented reality kenal jenis jenis perangkat keras computer bas android
9	[game,pembelajaran,bacaan,alquran,metode,dirosa]	game ajar baca alquran metode dirosa
10	[sistem,pakar,diagnosa,kerusakan,printer,canon,ip,2770,metode,case,base,reasoning]	sistem pakar diagnosa rusa printer canon ip 2770 metode case base reasoning
...	...	...
118	[prediksi,produksi,udang,vaname,metode,regresi,linier,berganda,studi,dinas,perikanan,kelautan,provinsi,gorontalo]	prediksi produksi udang vaname metode regresi linier ganda studi dinas ikan laut provinsi gorontalo

4.3 Pembobotan TF-IDF

Tahapan selanjutnya setelah preprocessing dan pelabelan dilakukan, kemudian peneliti memberikan bobot pada setiap kata dengan menggunakan pembobotan TF-IDF. Pada tahap ini dilakukan pembobotan kata dari hasil stemming dengan metode *Term Inverse Document Frequency* (TF-IDF).

```

from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer

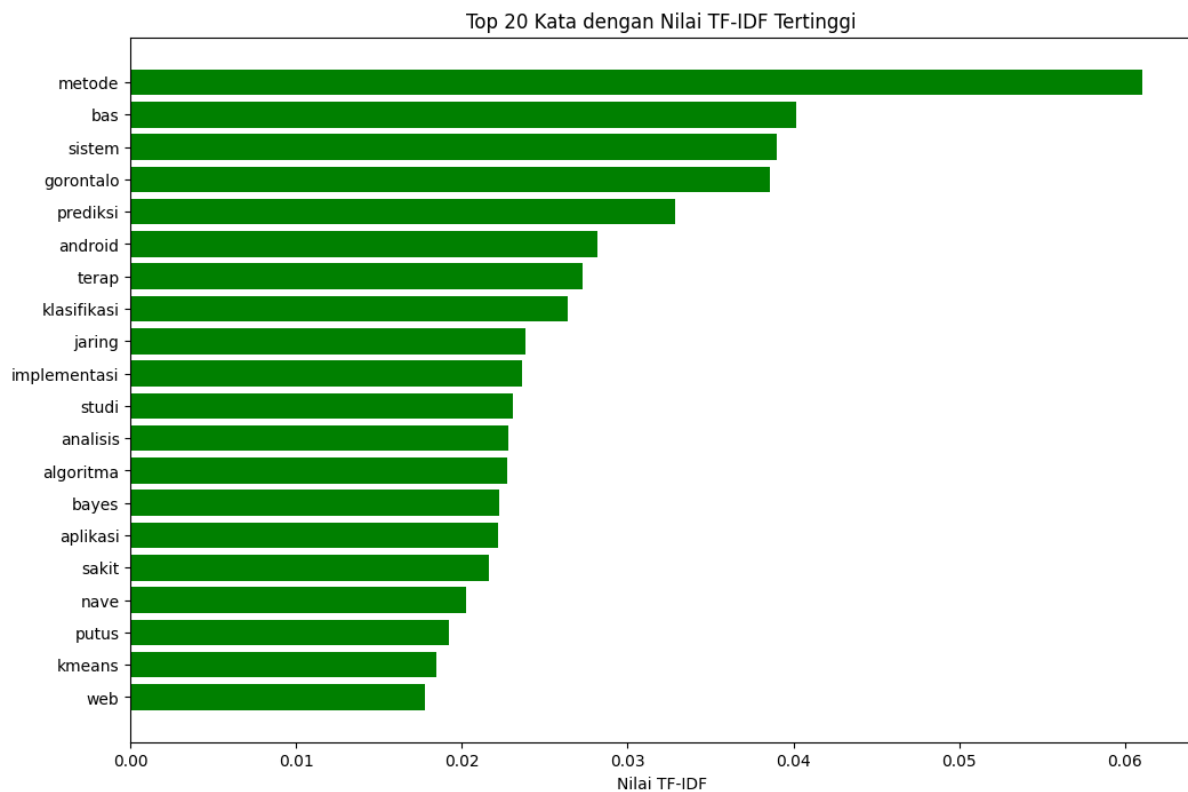
tf = TfidfVectorizer()
text_tf = tf.fit_transform(data_clean['Judul'].astype('U'))

# Visualisasi
feature_names = tf.get_feature_names_out()
tfidf_mean = np.mean(text_tf.toarray(), axis=0)
tfidf_df = pd.DataFrame({'word': feature_names, 'tfidf': tfidf_mean})
top_n = 20
top_tfidf_df = tfidf_df.nlargest(top_n, 'tfidf')
plt.figure(figsize=(12, 8))
plt.barh(top_tfidf_df['word'], top_tfidf_df['tfidf'], color='green')
plt.xlabel('Nilai TF-IDF')
plt.title(f'Top {top_n} Kata dengan Nilai TF-IDF Tertinggi')
plt.gca().invert_yaxis()

```

```
plt.show()
```

Pembobotan kata bertujuan untuk menentukan seberapa penting setiap kata dalam dokumen atau baris data. Tiap kata memiliki bobot berbeda karena dipengaruhi oleh frekuensi kemunculan kata tersebut pada tiap dokumen dan banyaknya dokumen yang terdapat kata tersebut.



Gambar 4.11 : Visualisasi Top 20 Term

4.4 Klasifikasi

Pada tahapan ini akan dilakukan proses klasifikasi dengan menggunakan. Proses Klasifikasi ini dilakukan setelah hasil dokumen-dokumen tersebut diberikan bobot menggunakan TF-IDF. Pada tahapan ini peneliti melakukan klasifikasi menggunakan algoritma K-Nearest Neighbor.

4.4.1. Pembagian Data Training dan Testing

Sebelum tahap klasifikasi dilakukan terlebih dahulu akan dilakukan pembagian dataset menjadi dua yaitu data latih dan data uji. Data latih akan digunakan untuk melatih model sedangkan data uji akan digunakan untuk menguji model yang telah dilatih agar dapat diketahui seberapa baik model yang telah dibuat. Pembagian data untuk data training dan data testing menggunakan perbandingan 80% : 20%.

Pembagian dataset ini dapat dilakukan dengan menggunakan modul atau fungsi dari library python yaitu modul `train_test_split` dari library Scikit-Learn. Adapun script python dari pembagian dataset ini adalah sebagai berikut :

```
# from tkinter.constants import X
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(text_tf, data_clean['Kategori'],
test_size=0.2, random_state=42)
print('Number Transaction X_Train Dataset :', x_train.shape)
print('Number Transaction X_Test Dataset :', x_test.shape)
print('Number Transaction Y_Train Dataset :', y_train.shape)
print('Number Transaction Y_Test Dataset :', y_test.shape)
```

Setelah pembagian dataset dilakukan maka data siap diterapkan pada model. Berikut script python dari model KNN :

```
# Melatih model KNN
from sklearn.neighbors import KNeighborsClassifier
knn_model = KNeighborsClassifier(n_neighbors=5) # Contoh dengan K=5, bisa
disesuaikan
knn_model.fit(x_train, y_train)

# Evaluasi model dengan data uji
accuracy = knn_model.score(x_test, y_test)
print(f'Accuracy on test set: {accuracy:.2f}')

# Contoh prediksi dengan data uji
y_pred = knn_model.predict(x_test)
```

```
# Prediksi pada data baru
def predict_new_data(new_data, model, vectorizer):
    new_data_clean = casefolding(new_data)
    new_data_tokens = tokenization(new_data_clean)
    new_data_filtered = remove_stopwords(new_data_tokens)
    new_data_stemmed = stemming(new_data_filtered)
    # No change is needed here as the vectorizer passed in
    # should already be fitted
    new_data_vectorized = vectorizer.transform([new_data_stemmed])
    prediction = model.predict(new_data_vectorized)
    return prediction[0]
```

```
def predict_judul(new_data):
    prediction = predict_new_data(new_data, knn_model, tf) # Pass in the model and
vectorizer
    return {'prediction': prediction}
```

```
# Gradio UI
# Use gr.Textbox instead of gr.inputs.Textbox
iface = gr.Interface(
    fn=predict_judul,
    inputs=gr.Textbox(label="Masukkan Judul"),
    outputs="text",
    title="Prediksi Kategori Berita dengan Model KNN",
    description="Masukkan judul berita untuk diprediksi kategori beritanya."
)
iface.launch()
```

BAB V

PEMBAHASAN

5.1 Evaluasi

Data yang digunakan dalam peneitian ini bersumber dari Fakultas Ilmu komputer Prodi Teknik Informatika, yang terdiri dari 9 kategori judul, yaitu SPK, Text Minning, Sistem Pakar, Data Minning, Game Edukasi, GIS, Web dan Mobile, Jaringan, Prototipe

5.2 Pembahasan Hasil Pemodelan

Penerapan metode K-Nearest Neighbor dalam Klasifikasi Tugas Judul Tugas Akhir dengan proses pemodelan dengan tahapan sebagai berikut :

1. Preprocessing

Pada tahapan ini beberapa proses dilakukan diantaranya dataset Judul Tugas Akhir dalam bentuk file excel dikonversi menjadi file csv agar dapat diolah di python lebih cepat dan lebih mudah. Setelah itu dilakuan proses Case folding, Tokenizing, Stopword removal, Stemming, dan Normalization.

2. Data yang berhasil dikumpulkan diperoleh data 118 yang terdiri atas 9 Kategori Judul

3. Pembagian Data Training dan Data Testing

Pembagian Data Training dan Data Testing untuk pemodelan K-Nearest Neighbor dalam Klasifikasi Judul Tugas Akhir baru yaitu 80:20 sehingga didapatkan masing-masing Data Training dan Data Testing dan mendapatkan Akurasi sebesar 0.75%

5.3 Aplikasi Hasil Pemodelan

5.3.1. Halaman Data View Dataset

Prediksi Kategori Berita dengan Model KNN

View Dataset	Case Folding	Tokenisasi	StopWords Removal	stemming	TF-IDF	Model KNN	Prediksi
Dataset Judul							
Judul							
Sistem pendukung keputusan penentuan keikutsertaan prioritas program kb menggunakan metode analytical herarchy process							
Analisis sentimen opini public pengguna twitter terhadap kenaikan harga bbm menggunakan algoritma naive bayes							
Rancang bangun web service sistem informasi berkas administrasi tugas akhir teknik informatika (SIABTI) berbasis progressive web app (android)							
Aplikasi E-learning pembelajaran sekolah menengah pertama negeri 2 satu atap bulango utara							
Klasifikasi malware menggunakan teknik machine learning							
Penerapan metode K-Means untuk clustering penjualan suku cadang kendaraan viar (STUDI KASUS CV. GOTAMA VIAR GORONTALO)							
Mobile administrasi desa di kantor desa kuala utara berbasis android							
Implementasi augmented reality dalam pengenalan jenis jenis perangkat keras computer berbasis android							
Game pembelajaran bacaan AL-QUR'DAN dengan menggunakan metode dirosa							
Sistem pakar diagnosa kerusakan printer canon ip 2770 menggunakan metode case base reasoning							
Klasifikasi mutu greenbean coffee menggunakan metode convolutional neural network(CNN)							
Penerapan algoritma K-means untuk clustering daerah penyebaran penyakit demam berdarah dengue(DBD)							

Use via API  Built with Gradio 

Gambar 5.1 : Halaman Dataset

5.3.2. Halaman Casefolding

Prediksi Kategori Berita dengan Model KNN

View Dataset	Case Folding	Tokenisasi	StopWords Removal	stemming	TF-IDF	Model KNN	Prediksi
Hasil Case Folding							
Judul							
sistem pendukung keputusan penentuan keikutsertaan prioritas program kb menggunakan metode analytical herarchy process							
analisis sentimen opini public pengguna twitter terhadap kenaikan harga bbm menggunakan algoritma nave bayes							
rancang bangun web service sistem informasi berkas administrasi tugas akhir teknik informatika siabti berbasis progressive web app android							
aplikasi elearning pembelajaran sekolah menengah pertama negeri 2 satu atap bulango utara							
klasifikasi malware menggunakan teknik machine learning							
penerapan metode kmeans untuk clustering penjualan suku cadang kendaraan viar studi kasus cv gotama viar gorontalo							
mobile administrasi desa di kantor desa kuala utara berbasis android							
implementasi augmented reality dalam pengenalan jenis jenis perangkat keras computer berbasis android							
game pembelajaran bacaan alquran dengan menggunakan metode dirosa							
sistem pakar diagnosa kerusakan printer canon ip 2770 menggunakan metode case base reasoning							
klasifikasi mutu greenbean coffee menggunakan metode convolutional neural networkknn							
penerapan algoritma kmeans untuk clustering daerah penyebaran penyakit demam berdarah denguedbd							

Use via API  Built with Gradio 

Gambar 5.1 : Halaman Dataset

5.3.3. Halaman Tokenisasi

Prediksi Kategori Berita dengan Model KNN

View Dataset Case Folding **Tokenisasi** StopWords Removal stemming TF-IDF Model KNN Prediksi

Hasil Tokenisasi

Judul
sistem,pendukung,keputusan,penentuan,keikutsertaan,prioritas,program,kb,menggunakan,metode,analytical,herarchy,process
analisis,sentimen,opini,public,pengguna,twitter,terhadap,kenaikan,harga,bbm,menggunakan,algoritma,nave,bayes
rancang,bangun,web,service,sistem,informasi,berkas,administrasi,tugas,akhir,teknik,informatika,siabti,berbasis,progressive,web,app,android
aplikasi,elearning,pembelajaran,sekolah,menengah,pertama,negeri,2,satu,atap,bulango,utara
klasifikasi,malware,menggunakan,teknik,machine,learning
penetapan,metode,kmeans,untuk,clustering,penjualan,suku,cadang,kendaraan,viar,studi,kasus,cv,gotama,viar,gorontalo
mobile,administrasi,desa,di,kantor,desa,kuala,utara,berbasis,android
implementasi,augmented,reality,dalam,pengenalan,jenis,jenis,perangkat,keras,computer,berbasis,android
game,pembelajaran,bacaan,alquran,dengan,menggunakan,metode,dirosa
sistem,pakar,diagnosa,kerusakan,printer,canon,ip,2770,menggunakan,metode,case,base,reasoning
klasifikasi,mutu,greenbean,coffee,menggunakan,metode,convolutional,neural,workcnn
penetapan,algoritma,kmeans,untuk,clustering,daerah,penyebaran,penyakit,demam,berdarah,denguedbd

Use via API  Built with Gradio 

Gambar 5.1 : Halaman Dataset

5.3.4. Halaman Stopword Removal

Prediksi Kategori Berita dengan Model KNN

View Dataset Case Folding Tokenisasi **StopWords Removal** stemming TF-IDF Model KNN Prediksi

Hasil StopWords Removal

Judul
sistem,pendukung,keputusan,penentuan,keikutsertaan,prioritas,program,kb,metode,analytical,herarchy,process
analisis,sentimen,opini,public,pengguna,twitter,kenaikan,harga,bbm,algoritma,nave,bayes
rancang,bangun,web,service,sistem,informasi,berkas,administrasi,tugas,teknik,informatika,siabti,berbasis,progressive,web,app,android
aplikasi,elearning,pembelajaran,sekolah,menengah,negeri,2,atap,bulango,utara
klasifikasi,malware,teknik,machine,learning
penerapan,metode,kmeans,clustering,penjualan,suku,cadang,kendaraan,viar,studi,cv,gotama,viar,gorontalo
mobile,administrasi,desa,kantor,desa,kuala,utara,berbasis,android
implementasi,augmented,reality,pengenalan,jenis,jenis,perangkat,keras,computer,berbasis,android
game,pembelajaran,bacaan,alquran,metode,dirosa
sistem,pakar,diagnosa,kerusakan,printer,canon,ip,2770,metode,case,base,reasoning
klasifikasi,mutu,greenbean,coffee,metode,convolutional,neural,workcnn
penerapan,algoritma,kmeans,clustering,daerah,penyebaran,penyakit,demam,berdarah,denguedbd

Use via API  Built with Gradio 

Gambar 5.1 : Halaman Dataset

5.3.5. Halaman Steamining

Prediksi Kategori Berita dengan Model KNN

View Dataset Case Folding Tokenisasi StopWords Removal **stemming** TF-IDF Model KNN Prediksi

Hasil Stemming

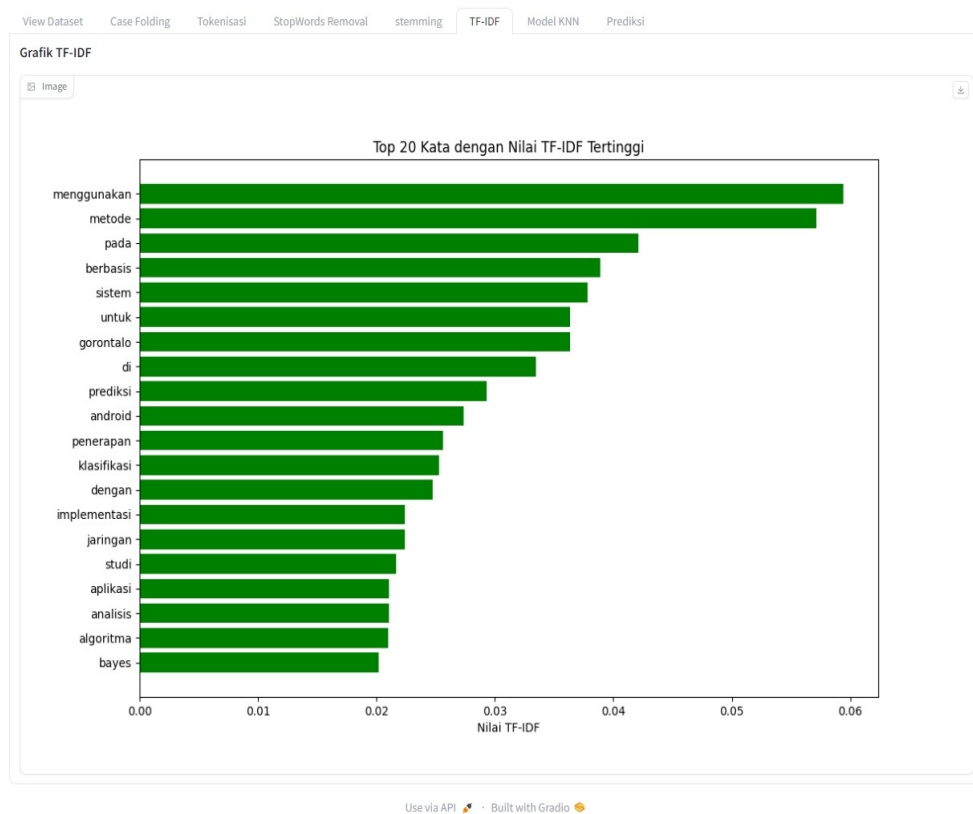
Judul	Kategori
sistem dukung putus tentu keikutsertaan prioritas program kb metode analytical herarchy process	SPK
analisis sentimen opini public guna twitter naik harga BBM algoritma nave bayes	Teks Mining
rancang bangun web service sistem informasi berkas administrasi tugas teknik informatika siabti bas progressive web app android	Web dan Mob:
aplikasi elearning ajar sekolah tengah negeri 2 atap bulango utara	Web dan Mob:
klasifikasi malware teknik machine learning	Data Mining
terap metode kmeans clustering jual suku cadang kendaraan viar studi cv gotama viar gorontalo	Data Mining
mobile administrasi desa kantor desa kuala utara bas android	Web dan Mob:
implementasi augmented reality kenal jenis jenis perangkat keras computer bas android	Web dan Mob:
game ajar baca alquran metode dirosa	Game Edukas:
sistem pakar diagnosa rusa printer canon ip 2770 metode case base reasoning	Sistem Paka:
klasifikasi mutu greenbean coffee metode convolutional neural networkcnn	Data Mining
terap algoritma kmeans clustering daerah sebar sakit demam darah denguebd	GIS

Use via API Built with Gradio

Gambar 5.1 : Halaman Dataset

5.3.6. Halaman TF-IDF

Prediksi Kategori Berita dengan Model KNN



Gambar 5.1 : Halaman Dataset

5.3.7. Halaman Model Pengujian

Prediksi Kategori Berita dengan Model KNN

View Dataset
Case Folding
Tokenisasi
StopWords Removal
stemming
TF-IDF
Model KNN
Prediksi

Evaluasi Model KNN
Accuracy on test set: 0.75
Contoh Prediksi

Actual	Predicted
Web dan Mobile	Web dan Mobile
Jaringan	Jaringan
Data Mining	Data Mining
Web dan Mobile	Game Edukasi
Data Mining	Data Mining
Sistem Pakar	Sistem Pakar
Prototipe	Prototipe
Data Mining	Data Mining
Prototipe	GIS
SPK	SPK
GIS	GIS
GIS	Data Mining

Use via API  Built with Gradio 

Gambar 5.1 : Halaman Dataset

BAB VI

PENUTUP

6.1 Kesimpulan

Berdasarkan hasil penelitian yang telah diuraikan sebelumnya maka ditarik suatu kesimpulan sebagai berikut :

1. Penerapan metode K-Nearest Neighbor dalam mengklasifikasi Judul Tugas Akhir dapat diterapkan secara efektif karena metode KNN memberikan prediksi yang mendekati nilai.
2. Penerapan metode K-Nearest Neighbor (KNN) dalam Klasifikasi Judul Tugas Akhir telah terbukti efektif. Kinerja yang mendekati tercapai saat menggunakan nilai $K=3$, dengan Nilai Akurasi Sebesar 0.75

6.2 Saran

Penelitian ini masih memiliki banyak kekurangan, oleh sebab itu penulis memberikan saran untuk penelitian selanjutnya, yaitu:

1. Mempertimbangkan metode pelabelan dataset secara manual dan dengan begitu dapat membandingkan hasil keakuratan dari pelabelan dataset secara manual dan otomatis.
2. Menggunakan metode analisis sentimen lain seperti Support Vector Machine (SVM), K-NN, Decision Tree dan dengan begitu dapat membandingkan hasil performa yang dilakukan untuk mencari metode klasifikasi yang terbaik

DAFTAR PUSTAKA

- [1] MochammadIlman Asnada¹, Bayu Rahayudi², Achmad Ridok³Program StudiTeknikInformatika, FakultasIlmuKomputer,PengelompokanTopikSkripsiMahasiswaFakultasIlmuKomputerUniversitasBrawijayaberdasarkanJudulpadaPeriode 2015-2019menggunakan Metode Semi Supervised K-Means 2022
- [2] 2021 Herfandi, Arman Diansyah, Eri Sasmita Susanto, “JINTEKS (Jurnal Informatika Teknologi dan Sains) ISSN 2686-3359 (Online) JINTEKS (Jurnal Informatika Teknologi dan Sains) ISSN 2686-3359 (Online),” (*Jurnal Inform. Teknol. dan Sains*, vol. 3, no. 1, pp. 301–307, 2021).
- [3] G. Tika, A. Adiwijaya, and N. D. W. Cahyani, “Klasifikasi Topik Berita Berbahasa Indonesia Menggunakan Multilayer Perceptron,” *eProceedings Eng.*, vol. 6, no. 1, pp. 2137–2143, 2019.
- [4] R. Sari, “Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (K-Nn),” *EVOLUSI J. Sains dan Manaj.*, vol. 8, no. 1, pp. 10–17, 2020.
- [5] R. Feldman and J. Sanger, *The Text Mining Handbook*. 2006.
- [6] A. Bengnga and R. Ishak, “Prediksi Jumlah Mahasiswa Registrasi Per Semester Menggunakan Linier Regresi Pada Universitas Ichsan Gorontalo,” *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 136–143, 2018.
- [7] I. A. Ahmad Sabri, M. Man, W. A. W. Abu Bakar, and A. N. Mohd Rose, “Web Data Extraction Approach for Deep Web using WEIDJ,” *Procedia Comput. Sci.*, vol. 163, pp. 417–426, 2019.
- [8] “Data Mining: Pengertian, Fungsi, Penerapan, dan Metode Pengambilannya - Cermati.com.” .
- [9] A. Firdaus and W. I. Firdaus, “Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan),” *J. JUPITER*, vol. 13, no. 1, p. 66, 2021.
- [10] M. Learning, “Learning basic concept,” pp. 1–39, 2019.
- [11] O. Sakti *et al.*, “Klasifikasi Teks Dengan Menggunakan Algoritma K-Nearest,” vol. 5, no. 3, pp. 8237–8248, 2018.

LAMPIRAN CODING

```
!pip install gradio
!pip install Sastrawi
!pip install xlswriter
!pip install imblearn
import string
import gradio as gr
from sklearn.pipeline import Pipeline # Fixed typo here
from imblearn.over_sampling import SMOTE
import pandas as pd
import numpy as np
import re
import nltk
from nltk.corpus import stopwords

from nltk.tokenize import word_tokenize
from nltk.stem import PorterStemmer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.neighbors import KNeighborsClassifier

import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
from sklearn.manifold import TSNE
```

```
# Load the dataset
df = pd.read_csv('dataset.csv', sep=';', encoding='latin-1')
print("Dataset awal:")
# print(df.head(100))
df.head(118)
```

```
# Define the case folding function
def casefolding(Judul):
    Judul = Judul.lower()
    Judul = Judul.strip()
    Judul = re.sub(r'[^a-zA-Z0-9¥s]', '', Judul)
    return Judul

# Apply case folding to the 'Judul' column
df['casefolded_Judul'] = df['Judul'].apply(casefolding)
# print("\nDataset setelah case folding:")
# print(df[['Judul', 'Hasil Case Folding']].head(118))
df.head(118)
```

```
def tokenization(Judul):
    tokens = word_tokenize(Judul)
    return tokens
df['tokens'] = df['casefolded_Judul'].apply(tokenization)
df.head(118)
```

```
# Define stemming function
def stemming(removal):
    porter = PorterStemmer()
    stemmed_tokens = [porter.stem(token) for token in removal]
    return stemmed_tokens

# Apply stemming to the 'tokens' column
df['stemmed_tokens'] = df['removal'].apply(stemming)

# Display the final table with relevant columns
print("\nTabel Hasil Case Folding, Tokenisasi, removal, dan Stemming:")
result_df = df[['Judul', 'casfolding', 'tokens', 'removal', 'stemmed_tokens']]
result_df.head(118)
```



**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS ICHSAN GORONTALO
LEMBAGA PENELITIAN**

Kampus Unisan Gorontalo Lt.3 - Jln. Achmad Nadjamuddin No. 17 Kota Gorontalo
Telp: (0435) 8724466, 829975 E-Mail: lembagapenelitian@unisan.ac.id

SURAT KETERANGAN

Nomor : 5019/SK/LEMLIT-UNISAN/GTO/VI/2024

Yang bertanda tangan di bawah ini :

Nama : Dr. Rahmisyari, ST.,SE.,MM
NIDN : 0929117202
Jabatan : Ketua Lembaga Penelitian

Menerangkan bahwa :

Nama Mahasiswa : Ahmad Nusi
NIM : T3117107
Fakultas : Fakultas Ilmu Komputer
Program Studi : Teknik Informatika
Judul Penelitian : KLASIFIKASI JUDUL TUGAS AKHIR MAHASISWA
MENGUNAKAN K-NEAREST NEIGHBOR

Adalah benar telah melakukan pengambilan data penelitian dalam rangka Penyusunan Proposal/Skripsi pada KETUA PROGRAM STUDI TEKNIK INFORMATIKA UNISAN GORONTALO.

Gorontalo, 05 Juni 2024
Ketua

Dr. Rahmisyari, ST.,SE.,MM
NIDN 0929117202

+



**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS ICHSAN GORONTALO
LEMBAGA PENELITIAN**

Kampus Unisan Gorontalo Lt.3 - Jln. Achmad Nadjamuddin No. 17 Kota Gorontalo
Telp: (0435) 8724466, 829975 E-Mail: lembagapenelitian@unisan.ac.id

Nomor : 5019/PIP/LEMLIT-UNISAN/GTO/I/2024

Lampiran : -

Hal : Permohonan Izin Penelitian

Kepada Yth,

Ketua Program Studi Teknik Informatika Unisan Gorontalo

di,-

Tempat

Yang bertanda tangan di bawah ini :

Nama : Dr. Rahmisyari, ST.,SE.,MM

NIDN : 0929117202

Jabatan : Ketua Lembaga Penelitian

Meminta kesediannya untuk memberikan izin pengambilan data dalam rangka penyusunan **Proposal / Skripsi**, kepada :

Nama Mahasiswa : Ahmad Nusi

NIM : T3117107

Fakultas : Fakultas Ilmu Komputer

Program Studi : Teknik Informatika

Lokasi Penelitian : PROGRAM STUDI TEKNIK INFORMATIKA UNIVERSITAS ICHSAN GORONTALO

Judul Penelitian : ANALISA TINGKAT KEPUASAN MAHASISWA TERHADAP KINERJA SISTEM INFORMASI AKADEMIK (SIKUN) DI PROGRAM STUDI TEKNIK INFORMATIKA MENGGUNAKAN KNN

Atas kebijakan dan kerja samanya diucapkan banyak terima kasih.

Gorontalo, 26 Januari 2024

Dr. Rahmisyari, ST.,SE.,MM
 NIDN 0929117202

+



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS ICHSAN GORONTALO

FAKULTAS ILMU KOMPUTER

SURAT KEPUTUSAN MENDIKNAS RI NOMOR 84/D/O/2001

Jl. Achmad Najamuddin No. 17 Telp. (0435) 829975 Fax (0435) 829976 Gorontalo

SURAT REKOMENDASI BEBAS PLAGIASI

No. 189/FIKOM-UIG/R/VI/2024

Yang bertanda tangan di bawah ini :

Nama : Irvan Abraham Salihi, M.Kom
 NIDN : 0928028101
 Jabatan : Dekan Fakultas Ilmu Komputer

Dengan ini menerangkan bahwa :

Nama Mahasiswa : Ahmad Nusi
 NIM : T3117107
 Program Studi : Teknik Informatika (S1)
 Fakultas : Fakultas Ilmu Komputer
 Judul Skripsi : Klasifikasi Judul Tugas Akhir Mahasiswa
 Menggunakan Metode K-Nearest Neighbor

Sesuai hasil pengecekan tingkat kemiripan skripsi melalui aplikasi Turnitin untuk judul skripsi di atas diperoleh hasil *Similarity* sebesar 22%, berdasarkan Peraturan Rektor No. 32 Tahun 2019 tentang Pendeteksian Plagiat pada Setiap Karya Ilmiah di Lingkungan Universitas Ichsan Gorontalo dan persyaratan pemberian surat rekomendasi verifikasi calon wisudawan dari LLDIKTI Wil. XVI, bahwa batas kemiripan skripsi maksimal 30%, untuk itu skripsi tersebut di atas dinyatakan **BEBAS PLAGIASI** dan layak untuk diujikan.

Demikian surat rekomendasi ini dibuat untuk digunakan sebagaimana mestinya.

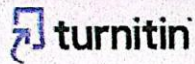
Mengetahui
 Dekan,

Irvan Abraham Salihi, M.Kom
 NIDN. 0928028101

Gorontalo, 25 Juni 2024
 Tim Verifikasi,

Zulfrianto Y. Lamasigi, M.Kom
 NIDN. 0914089101

Terlampir :
 Hasil Pengecekan Turnitin



Similarity Report ID: oid:25211:62021192

PAPER NAME

SKRIPSI T3117107 AHMAD NUSI.pdf

AUTHOR

AHMAD NUSI ahmadnusi006@gmail.com

WORD COUNT

6683 Words

CHARACTER COUNT

43637 Characters

**PAGE
COUNT**

43 Pages

FILE SIZE

1.9MB

SUBMISSION DATE

Jun 26, 2024 6:17 PM GMT+8

REPORT DATE

Jun 26, 2024 6:18 PM GMT+8

22% Overall Similarity

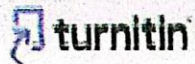
The combined total of all matches, including overlapping sources, for each database.

- 22% Internet database
- 3% Publications database
- Crossref database
- Crossref Posted Content database
- 1% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Quoted material
- Cited material
- Small Matches (Less than 30 words)

Summary



Similarity Report ID: oid:25211:62021192

22% Overall Similarity

Top sources found in the following databases:

- 22% Internet database
- 3% Publications database
- Crossref database
- Crossref Posted Content database
- 1% Submitted Works database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	core.ac.uk	14%
	Internet	
2	dolanindonesiaku.com	2%
	Internet	
3	neliti.com	1%
	Internet	
4	ojs.stmik-banjarbaru.ac.id	<1%
	Internet	
5	ejournal.uin-suska.ac.id	<1%
	Internet	
6	etd.umy.ac.id	<1%
	Internet	
7	scribd.com	<1%
	Internet	
8	coursehero.com	<1%
	Internet	

[Sources overview](#)

RIWAYAT HIDUP



Nama : Ahmad Nusi
 TTL : Limboto, 12 Januari 1998
 Jenis Kelamin : Laki-Laki
 Status Kawin : Belum kawin
 Agama : Islam
 Email : ahmadnusi006@gmail.com

ORANG TUA

Ayah : Udin Nusi
 Ibu : Hadidjah Abubakar

PENDIDIKAN

1. Tahun 2010, Menyelesaikan Pendidikan di Sekolah Dasar Negeri 20 Limboto
2. Tahun 2013, Menyelesaikan Pendidikan di Sekolah Menengah Pertama (SMP) 2 Limboto
3. Tahun 2016, Menyelesaikan Pendidikan di Sekolah Menengah Kejuruan (SMK) 2 Limboto