**ANALISIS SENTIMEN MAHASISWA TERHADAP**

**TUGAS-TUGAS KULIAH DENGAN MENGGUNAKAN METODE K-*NEAREST NEIGHBORS***

# PROPOSAL PENELITIAN

****

Oleh

ERVINA TRYASTUTI RAHAYU

NIM : 20050874002

**UNIVERSITAS NEGERI SURABAYA**

**FAKULTAS TEKNIK**

**JURUSAN TEKNIK ELEKTRO**

**PRODI TELEKOMUNIKASI DAN KOMPUTASI CERDAS**

**202****4**

# HALAMAN PERSETUJUAN

|  |  |
| --- | --- |
| Usulan Penelitian Oleh |  |
| Nama : | Ervina Tryastuti Rahayu |
| NIM : | 20050874002 |
| Judul : | Analisis Sentimen Mahasiswa Terhadap Tugas-Tugas Kuliah Dengan Menggunakan Metode  K-*Nearest Neighbors* |

Dengan ini telah disetujui dan dinyatakan memenuhi syarat untuk diseminarkan.

Surabaya, 25 Maret 2024

Pembimbing,



**Prof. Dr. I Gusti Putu Asto Buditjahjanto, S.T., M.T.**

**NIP : 197107061999031001**

# KATA PENGANTAR

Segala puji bagi Allah Yang Maha Esa atas rahmat serta hidayah-Nya, sehingga penulis dapat menyelesaikan proposal penelitian skripsi dengan judul " Analisis Sentimen Mahasiswa Terhadap Tugas-Tugas Kuliah Dengan Menggunakan Metode K- *-Nearest Neighbros*". Proposal Penelitian ini disusun sebagai syarat dalam penulisan skripsi di Program Studi S-1 Teknik Elektro Universitas Negeri Surabaya.

Dalam pengerjaan, penulis mendapatkan bimbingan serta dukungan dari banyak pihak. Maka dari itu, penulis mengucapkan rasa syukur dan terima kasih kepada kedua orang tua dan adik penulis yang senantiasa memberikan harapan dan doanya kepada penulis sehingga penulis mampu menyelesaikan Proposal Penelitian Skripsi ini. Selain itu, penulis juga mengucapkan terima kasih sebesar-besarnya kepada:

1. Ibu Dr. Maspiyah, M.Kes. selaku Dekan Fakultas Teknik Universitas Negeri Surabaya.
2. Bapak Prof. Dr. Agus Wiyono S.Pd., M.T. selaku Wakil Dekan II Fakultas Teknik Universitas Negeri Surabaya.
3. Ibu Dr. Lusia Rakhmawati, S.T., M.T. selaku Koordinator Program Studi S1 Teknik Elektro Universitas Negeri Surabaya.
4. Bapak Prof. Dr. I Gusti Putu Asto Buditjahjanto, S.T., M.T. selaku dosen pembimbing skripsi atas segala arahan dan motivasi dari beliau.
5. Kedua orang tua serta saudara yang telah mendukung dan memberikan motivasi hingga tersusunnya laporan ini.
6. Seluruh dosen pengajar jurusan Teknik Elektro yang telah memberi bimbingan dan ilmu yang bermanfaat.
7. Rekan-rekan S1 Teknik Elektro yang senantiasa memberikan dukungan dan semangat kepada penulis.
8. Serta berbagai pihak yang telah memberikan dukungan dalam skripsi ini yang tidak dapat saya sebutkan satu-persatu.

Harapan Penulis, Proposal penelitian Skripsi ini dapat berguna bagi para pembaca dan bagi penulis khususnya. Penulis Mengerti jika laporan ini masihjauh dari kata sempurna. Untuk Itu, penulis dengan tangan terbuka mengharapkan kritik dan saran membangun dari para pembaca sekalian.

|  |
| --- |
| Surabaya, 25 Maret 2024 |
| Penulis |

# DAFTAR ISI

[PROPOSAL PENELITIAN 1](#_Toc182414862)

[HALAMAN PERSETUJUAN ii](#_Toc182414863)

[KATA PENGANTAR iii](#_Toc182414864)

[DAFTAR ISI v](#_Toc182414865)

[DAFTAR TABEL vii](#_Toc182414866)

[DAFTAR GAMBAR viii](#_Toc182414867)

[BAB I PENDAHULUAN 1](#_Toc182414868)

[A. Latar Belakang 1](#_Toc182414869)

[B. Rumusan Masalah 5](#_Toc182414870)

[C. Tujuan Penelitian 5](#_Toc182414871)

[D. Manfaat Penelitian 5](#_Toc182414872)

[E. Batasan Penelitian 5](#_Toc182414873)

[F. Asumsi Penelitian 6](#_Toc182414874)

[BAB II KAJIAN PUSTAKA 8](#_Toc182414875)

[A. Landasan Teori 8](#_Toc182414876)

[B. Penelitian Relevan 28](#_Toc182414877)

[C. Kerangka Konseptual 31](#_Toc182414878)

[D. Hipotesis 36](#_Toc182414879)

[BAB III METODE PENELITIAN 37](#_Toc182414880)

[A. Jenis dan Desain Penelitian 37](#_Toc182414881)

[B. Tempat dan Waktu Penelitian 39](#_Toc182414882)

[C. Populasi dan Sampel 41](#_Toc182414883)

[E. Instrumen Penelitian 41](#_Toc182414884)

[F. Teknik Pengumpulan Data 43](#_Toc182414885)

[G. Teknik Analisis Data 46](#_Toc182414886)

[H. Blok Diagram Penelitian 47](#_Toc182414887)

[DAFTAR PUSTAKA 50](#_Toc182414888)

# 

# DAFTAR TABEL

No Uraian Hal

[Table 1.1 Survei awal persepsi mahasiswa terhadap tugas kuliah 2](#_Toc182823905)

[Table 2.1 Indeks Sentimen Mahasiswa 26](#_Toc182823906)

[Table 3.1 Waktu Penelitian 40](#_Toc182823907)

# 

# DAFTAR GAMBAR

No Uraian Hal

[Gambar 2.1 Alur Metode *K-Nearest Neighbor* 27](#_Toc182824254)

[Gambar 2.2 Kerangka Konseptual 35](#_Toc182824255)

[Gambar 3.1 Teknik Pengumpulan Data 43](file:///C:\Users\Vina\Downloads\20050874002_Ervina%20Tryastuti%20Rahayu_revisi0112.docx#_Toc184015728)

[Gambar 3.2 Scraping Data 44](file:///C:\Users\Vina\Downloads\20050874002_Ervina%20Tryastuti%20Rahayu_revisi0112.docx#_Toc184015729)

[Gambar 3.3 Preprocessing Data 45](file:///C:\Users\Vina\Downloads\20050874002_Ervina%20Tryastuti%20Rahayu_revisi0112.docx#_Toc184015730)

[Gambar 3.4 Blok Diagram Penelitian 48](#_Toc184015731)

# BAB I PENDAHULUAN

## A. Latar Belakang

Tugas kuliah merupakan salah satu komponen penting dalam proses pembelajaran di perguruan tinggi Tugas-tugas ini dirancang untuk membantu mahasiswa mengembangkan kemampuan berpikir kritis, keterampilan pemecahan masalah, dan penerapan teori ke dalam praktik nyata. Idealnya, tugas kuliah dapat mendukung pencapaian kompetensi yang diharapkan, baik secara individu maupun kelompok [1]. Namun, realitas di lapangan menunjukkan bahwa tidak semua tugas kuliah diterima dengan baik oleh mahasiswa, sehingga menciptakan perbedaan antara harapan institusi dan pengalaman mahasiswa.

Dalam era digital saat ini, media sosial seperti Twitter atau X telah menjadi wadah yang sering digunakan oleh mahasiswa untuk menyuarakan pendapat mereka. Platform ini memungkinkan mahasiswa untuk mengungkapkan opini terkait berbagai aspek pembelajaran, termasuk tugas kuliah. Mahasiswa sering kali memberikan pujian terhadap tugas yang dianggap relevan dan bermanfaat, tetapi tidak jarang juga menyampaikan keluhan tentang beban kerja yang terlalu berat. Fenomena ini memberikan peluang untuk memahami persepsi mahasiswa melalui analisis sentimen [2].

Analisis sentimen menjadi pendekatan yang relevan dalam memahami opini mahasiswa secara lebih objektif. Dengan memanfaatkan data unggahan dari media sosial, institusi pendidikan dapat memperoleh gambaran yang lebih jelas tentang bagaimana mahasiswa memandang tugas kuliah. Pendekatan ini juga dapat digunakan untuk mengevaluasi efektivitas desain tugas serta menentukan langkah-langkah perbaikan yang diperlukan.

Berdasarkan survei awal terhadap 500 unggahan mahasiswa di Twitter/X, persepsi mahasiswa terhadap tugas kuliah terbagi menjadi tiga kategori sentimen: positif, netral, dan negatif. Data menunjukkan bahwa 30% mahasiswa memberikan respons positif terhadap tugas kuliah, 40% bersikap netral, dan 30% lainnya menunjukkan sentimen negatif. Hasil ini menggambarkan keberagaman pengalaman mahasiswa dalam menghadapi tugas kuliah yang diberikan oleh dosen.

Distribusi data dari survei ini dapat dilihat pada tabel berikut:

Table 1.1 Survei awal persepsi mahasiswa terhadap tugas kuliah

|  |  |  |  |
| --- | --- | --- | --- |
| Kategori Sentimen | Jumlah Data | Persentase (%) | Contoh Unggahan |
| Positif | 150 | 30% | "Tugasnya bikin mikir kreatif, dosennya keren banget!" |
| Netral | 200 | 40% | "Tugas minggu ini lumayan, semoga bisa selesai tepat waktu." |
| Negatif | 150 | 30% | "Ini tugas apa hukuman? Sungguh memberatkan." |

Sumber: Data primer diolah (2024)

Hasil survei ini menunjukkan adanya kesenjangan antara ekspektasi mahasiswa dan desain tugas yang diberikan. Sebagian besar mahasiswa mengeluhkan bahwa tugas kuliah yang mereka terima tidak relevan dengan kebutuhan praktis atau konteks kehidupan nyata. Sebagai contoh, salah satu unggahan menyebutkan, *"Tugas ini tidak membantu kami memahami materi lebih baik, malah bikin stres"*. Keluhan seperti ini menunjukkan adanya ruang untuk perbaikan dalam desain tugas kuliah [3].

Di sisi lain, tugas kuliah sebenarnya merupakan alat penting bagi institusi pendidikan untuk mencapai hasil pembelajaran yang sesuai dengan standar kurikulum. Namun, apabila tugas tersebut menimbulkan keluhan yang signifikan di kalangan mahasiswa, dampaknya dapat menjadi kontraproduktif. Oleh karena itu, penting bagi institusi untuk mendengarkan dan memahami opini mahasiswa dalam rangka menciptakan tugas kuliah yang lebih efektif dan relevan.

Studi sebelumnya menunjukkan bahwa analisis sentimen berbasis algoritma dapat memberikan wawasan yang berharga dalam memahami opini publik. Salah satu metode yang sering digunakan adalah algoritma K-Nearest Neighbors (KNN). Algoritma ini mampu mengklasifikasikan data berdasarkan jarak terdekat, sehingga sangat cocok untuk menganalisis sentimen [1].

Penelitian oleh Wenando et al. (2022, p. 153) menunjukkan keberhasilan KNN dalam menganalisis respons publik terhadap kebijakan pendidikan. Dalam konteks ini, KNN digunakan untuk memetakan sentimen terhadap kebijakan Merdeka Belajar Kampus Merdeka. Hasil penelitian tersebut mengindikasikan bahwa KNN mampu menangkap dinamika opini dengan cepat dan efisien, bahkan dalam dataset yang besar [4].

Hal serupa juga ditemukan dalam penelitian Husdi & Kamaruddin, yang menggunakan KNN untuk menganalisis sentimen mahasiswa baru terhadap mata kuliah tertentu. Dengan memanfaatkan teknik auto-labeling, algoritma ini berhasil mengelompokkan data secara otomatis berdasarkan pola sentimen yang teridentifikasi. Temuan ini menunjukkan bahwa KNN dapat digunakan tidak hanya untuk klasifikasi data, tetapi juga untuk proses otomatisasi dalam analisis data besar [2].

Dalam konteks tugas kuliah, algoritma KNN memiliki potensi besar untuk memberikan pemetaan sentimen mahasiswa secara lebih rinci. Dengan menganalisis data unggahan di media sosial, institusi dapat memperoleh gambaran yang lebih lengkap tentang persepsi mahasiswa terhadap tugas yang mereka terima. Hal ini dapat menjadi dasar untuk merancang tugas yang lebih relevan dan mendukung tujuan pembelajaran secara lebih optimal.

Melalui penelitian ini, algoritma KNN digunakan untuk menganalisis 500 unggahan mahasiswa di Twitter/X. Data yang dianalisis mencakup opini positif, netral, dan negatif. Hasil analisis ini diharapkan dapat memberikan wawasan yang lebih mendalam tentang persepsi mahasiswa terhadap tugas kuliah, sekaligus menjadi alat bantu bagi dosen dalam mengevaluasi desain tugas.

Penelitian ini juga diharapkan dapat memberikan kontribusi nyata dalam pengembangan sistem pembelajaran di perguruan tinggi. Dengan memahami sentimen mahasiswa, institusi dapat merancang tugas kuliah yang lebih relevan dengan kebutuhan mahasiswa, sekaligus meningkatkan kualitas pengalaman belajar secara keseluruhan.

Dengan demikian, Dengan demikian, berdasarkan uji validitas dari penelitian-penelitian sebelumnya, penulis mengambil judul **"Analisis Sentimen Mahasiswa terhadap Tugas Kuliah Menggunakan Algoritma K-Nearest Neighbors"** untuk memberikan kontribusi pada pengembangan tugas kuliah yang lebih relevan dengan kebutuhan pembelajaran. Penelitian ini tidak hanya menjadi respons terhadap keluhan mahasiswa, tetapi juga langkah proaktif untuk menciptakan proses pembelajaran yang lebih efektif. Upaya ini diharapkan mampu menjembatani kesenjangan antara harapan mahasiswa dan realitas yang mereka hadapi dalam pendidikan tinggi.

## B. Rumusan Masalah

Berdasarkan uraian dari latar belakang tersebut, maka penulis merumuskan permasalahan penelitian yaitu:

1. Bagaimana sentimen mahasiswa terhadap tugas kuliah dapat dianalisis menggunakan data dari Twitter/X?
2. Bagaimana metode K-Nearest Neighbors (KNN) dapat digunakan untuk mengklasifikasikan sentimen mahasiswa menjadi positif, negatif, dan netral?

## C. Tujuan Penelitian

Berdasarkan rumusan masalah tersebut, maka tujuan dari rencana penelitian ini adalah:

1. Menganalisis sentimen mahasiswa terhadap tugas-tugas kuliah dengan memanfaatkan data dari platform media sosial Twitter/X.
2. Menggunakan metode K-Nearest Neighbors (KNN) untuk mengklasifikasikan sentimen mahasiswa menjadi positif, negatif, dan netral.

## D. Manfaat Penelitian

1. Manfaat Teoretis

Penelitian ini diharapkan dapat menambah referensi ilmiah dalam bidang analisis sentimen dan aplikasi algoritma K-Nearest Neighbors (KNN).

1. Bagi Penulis

Memberikan rekomendasi kepada pengajar dan institusi pendidikan terkait pola dan metode pemberian tugas kuliah berdasarkan hasil analisis sentimen mahasiswa.

## E. Batasan Penelitian

* 1. Data penelitian diambil dari platform media sosial Twitter/X, dengan jumlah tweet yang diambil sebanyak 1000 tweet. Data ini akan dikumpulkan dengan rentang waktu tertentu, yaitu mulai tanggal 1 Desember 2023 hingga 30 November 2024. Rentang waktu ini dipilih untuk memastikan bahwa data yang dianalisis mencerminkan tren sentimen mahasiswa terhadap tugas-tugas kuliah selama satu tahun penuh, sehingga dapat memberikan gambaran yang lebih komprehensif dan relevan dengan topik penelitian.
  2. Analisis sentimen terbatas pada tiga kategori utama, yaitu positif, negatif, dan netral. Pembatasan ini bertujuan untuk menyederhanakan analisis dan fokus pada persepsi dasar mahasiswa terhadap tugas kuliah.
  3. Subjek penelitian adalah mahasiswa aktif di perguruan tinggi. Penelitian ini hanya melibatkan mahasiswa yang terlibat langsung dalam kegiatan perkuliahan untuk memastikan data yang diperoleh mewakili persepsi mereka terhadap tugas kuliah.

## F. Asumsi Penelitian

Asumsi atau tanggapan dasar ini merupakan suatu gambaran sangkaan, perkiraan, satu pendapat atau kesimpulan sementara, atau suatu toeri sementara yang belum dibuktikan. Menurut Arikuntoasumsi atau anggapan dasar adalah sebuah titik tolak pemikiran yang kebenarannya diterima oleh penyelidik [6]. Penelitian ini didasarkan pada beberapa asumsi yang perlu diperhatikan dalam interpretasi hasil analisis. Pertama, diasumsikan bahwa data yang diambil dari Twitter/X dapat merepresentasikan opini mahasiswa secara umum. Hal ini didasarkan pada fakta bahwa Twitter/X merupakan platform sosial yang sering digunakan oleh mahasiswa untuk menyuarakan pendapat dan berbagi pengalaman terkait pendidikan, termasuk tugas kuliah. Menurut penelitian yang dilakukan oleh Furqan et al. (2022), platform media sosial seperti Twitter/X telah terbukti menjadi sumber yang efektif untuk menganalisis opini publik terkait berbagai topik, termasuk pendidikan [7].

Kedua, penelitian ini mengasumsikan bahwa mahasiswa memberikan opini yang jujur terkait tugas kuliah. Meskipun di platform sosial media seperti Twitter/X terdapat potensi bias atau opini yang tidak sepenuhnya objektif, diasumsikan bahwa mahasiswa cenderung memberikan pendapat yang jujur terkait pengalaman mereka dengan tugas kuliah, baik itu positif, negatif, maupun netral. Hal ini sejalan dengan penelitian yang dilakukan oleh Husdi & Kamaruddin (2024), yang mengidentifikasi bahwa di media sosial, terutama di Twitter, mahasiswa seringkali berbagi pengalaman pribadi secara terbuka, yang dapat dianggap sebagai bentuk ekspresi opini yang autentik [2].

Dengan batasan dan asumsi tersebut, penelitian ini bertujuan untuk memberikan gambaran yang jelas dan terfokus mengenai sentimen mahasiswa terhadap tugas kuliah, serta implementasi metode KNN dalam menganalisis data dari Twitter/X.

# BAB II KAJIAN PUSTAKA

## Landasan Teori

* 1. **Pengertian Sentimen**
     1. **Definisi Sentimen**

Sentimen adalah ekspresi perasaan, opini, atau sikap seseorang terhadap suatu objek, isu, atau kejadian yang dapat dilihat dalam bentuk positif, negatif, atau netral [8]. Menurut Cai (2023), sentimen merujuk pada sikap atau pandangan yang diekspresikan melalui teks, yang dapat dianalisis untuk mengetahui apakah sebuah pendapat bersifat baik, buruk, atau netral [8]. Dalam konteks ini, analisis sentimen digunakan untuk menggali reaksi atau tanggapan individu terhadap suatu hal, seperti kebijakan pendidikan atau pengalaman belajar [9].

Sementara itu, Yadav (2023) mendefinisikan sentimen sebagai reaksi emosional atau opini yang tercermin dalam komunikasi digital, yang kemudian dianalisis untuk memahami perasaan orang terhadap suatu kejadian atau objek [10]. Dengan demikian, sentimen bukan hanya mencerminkan opini semata, tetapi juga emosi yang terlibat dalam respons terhadap informasi yang diterima.

* + 1. **Jenis-jenis Sentimen (Positif, Negatif, Netral)**

Sentimen dapat dibagi menjadi tiga kategori utama [10], yakni:

1. Sentimen Positif: Merupakan ekspresi yang menunjukkan perasaan atau penilaian yang baik terhadap sesuatu. Dalam analisis sentimen, kata-kata seperti "terbaik", "mengesankan", atau "memuaskan" sering digunakan untuk menggambarkan sentimen positif.
2. Sentimen Negatif: Ini menggambarkan perasaan atau opini yang tidak menyenangkan atau tidak puas terhadap suatu hal. Kata-kata seperti "buruk", "kecewa", atau "terbebani" sering dijumpai dalam sentimen negatif, misalnya dalam komentar mahasiswa mengenai tugas kuliah yang dirasa terlalu sulit atau memberatkan.
3. Sentimen Netral: Merupakan kategori yang tidak menunjukkan emosi atau penilaian yang jelas, baik positif maupun negatif. Pada umumnya, kalimat yang bersifat informatif tanpa adanya penekanan emosi digolongkan ke dalam sentimen ini.

Jenis-jenis sentimen ini penting dalam memahami bagaimana suatu topik atau isu diterima oleh masyarakat atau kelompok tertentu, termasuk dalam konteks mahasiswa terhadap tugas-tugas kuliah yang diberikan.

* + 1. **Pengukuran Sentimen dalam Analisis Data**

Pengukuran sentimen dilakukan untuk memahami bagaimana orang bereaksi terhadap suatu topik atau kejadian berdasarkan teks atau komentar yang mereka buat. Ada dua pendekatan utama dalam pengukuran sentimen: berbasis lexicon dan berbasis algoritma.

1. Pendekatan Berbasis Lexicon: Dalam pendekatan ini, sentimen diukur berdasarkan kata-kata yang terdapat dalam kamus sentimen, yang masing-masing kata sudah diberi label positif, negatif, atau netral. Misalnya, kata-kata seperti "bagus" atau "memuaskan" diberi label positif, sementara kata-kata seperti "buruk" atau "kecewa" diberi label negatif. Hal ini memungkinkan pengklasifikasian otomatis dari teks yang dianalisis [11].
2. Pendekatan Berbasis Algoritma: Dalam analisis berbasis algoritma, metode seperti K-Nearest Neighbor (KNN) digunakan untuk mengklasifikasikan teks berdasarkan kedekatannya dengan data pelatihan yang sudah diberi label sentimen. Supriyanto et al. (2023) menggunakan KNN untuk menganalisis sentimen publik terhadap pembelajaran daring, di mana data teks yang diolah mengarah pada klasifikasi sentimen positif, negatif, atau netral berdasarkan kedekatannya dengan data referensi [12].

Penerapan teknik-teknik ini memberikan wawasan yang lebih mendalam tentang bagaimana perasaan individu atau kelompok terhadap isu tertentu, seperti kebijakan atau pengalaman belajar yang sedang dianalisis [4]. Dengan demikian, pengukuran sentimen menjadi alat yang efektif untuk mengevaluasi persepsi publik secara otomatis dalam berbagai konteks, termasuk pendidikan dan pekerjaan mahasiswa.

* 1. **Analisis Sentimen**
     1. **Konsep dan Pentingnya Analisis Sentimen**

Analisis sentimen adalah proses untuk mengidentifikasi, mengklasifikasikan, dan memahami perasaan atau opini yang terkandung dalam sebuah teks, apakah itu positif, negatif, atau netral. Analisis ini sering digunakan untuk mengevaluasi respon publik terhadap suatu kejadian, produk, atau layanan berdasarkan teks yang dihasilkan di media sosial, forum online, atau ulasan pengguna. Analisis sentimen membantu dalam memahami opini masyarakat secara otomatis dan mengungkap pola perilaku serta sikap konsumen atau pengguna terhadap produk atau layanan tertentu [9].

Pentingnya analisis sentimen terletak pada kemampuannya untuk memberikan wawasan yang lebih mendalam tentang persepsi publik tanpa memerlukan interaksi langsung. Analisis ini digunakan untuk memantau citra merek, mengevaluasi efektivitas kampanye pemasaran, atau merespons keluhan pelanggan secara lebih cepat dan efektif [8]. Misalnya, dalam konteks politik atau media sosial, analisis sentimen dapat membantu dalam memahami bagaimana masyarakat merespons isu atau kebijakan tertentu, yang sangat penting bagi pengambil keputusan.

* + 1. **Metode dalam Analisis Sentimen**

Terdapat dua metode utama yang digunakan dalam analisis sentimen, yaitu metode berbasis lexicon dan metode berbasis *machine learning*.

1. Metode Lexicon-Based

Metode lexicon-based adalah pendekatan yang mengandalkan daftar kata atau kamus yang sudah diberi label sentimen tertentu, baik itu positif, negatif, atau netral. Setiap kata dalam teks dianalisis dan dibandingkan dengan kata-kata dalam kamus tersebut untuk menentukan nilai sentimen keseluruhan dari teks tersebut. Menurut Silge dan Robinson (2017), metode ini sangat bergantung pada kualitas dan kelengkapan kamus yang digunakan, dan umumnya lebih sederhana dibandingkan dengan pendekatan lainnya [11]. Kelemahan dari metode ini adalah ketidakmampuannya dalam menangani konteks kalimat secara menyeluruh, seperti ketika kata-kata yang memiliki sentimen positif dalam konteks tertentu bisa berubah makna jika digunakan dalam konteks lain.

1. Metode *Machine learning*

Metode *machine learning* lebih kompleks dan melibatkan algoritma yang dapat "belajar" dari data dan membuat prediksi berdasarkan pola yang ditemukan dalam data. Dalam analisis sentimen, algoritma seperti Naïve Bayes, Support Vector Machines (SVM), dan K-Nearest Neighbor (KNN) sering digunakan untuk mengklasifikasikan teks berdasarkan sentimen yang terkandung. Aggarwal (2018) mengungkapkan bahwa metode ini lebih unggul dibandingkan metode lexicon-based karena kemampuannya untuk menangkap konteks dan hubungan antar kata yang lebih kompleks [9]. Penggunaan *machine learning* juga memungkinkan sistem untuk terus belajar dan meningkatkan akurasi prediksinya seiring dengan bertambahnya data yang dianalisis.

* + 1. **Aplikasi Analisis Sentimen dalam Berbagai Bidang**

Analisis sentimen memiliki beragam aplikasi di berbagai bidang, yang memungkinkan penggunaannya untuk memahami reaksi publik dan mengambil keputusan yang lebih baik.

1. **Aplikasi dalam Media Sosial**

Di era digital saat ini, media sosial menjadi salah satu platform utama untuk menganalisis sentimen. Dengan miliaran postingan dan komentar yang dihasilkan setiap hari, platform seperti Twitter, Facebook, dan Instagram menjadi sumber data yang sangat kaya untuk analisis sentimen. Misalnya, analisis sentimen dapat digunakan untuk mengukur respon publik terhadap suatu produk atau kampanye pemasaran. Wenando et al. (2022) menunjukkan bagaimana analisis sentimen dapat digunakan untuk menganalisis opini masyarakat terhadap kebijakan tertentu melalui media sosial, seperti kebijakan "Merdeka Belajar" yang diimplementasikan di Indonesia [4]. Dengan memanfaatkan analisis sentimen, perusahaan atau pemerintah dapat mendapatkan gambaran tentang bagaimana masyarakat merespons isu atau produk yang mereka tawarkan.

1. **Aplikasi dalam Pendidikan**

Dalam bidang pendidikan, analisis sentimen dapat diterapkan untuk mengevaluasi pengalaman belajar siswa, terutama dalam konteks pembelajaran daring. Sebagai contoh, Supriyanto et al. (2023) menggunakan analisis sentimen untuk mengukur opini siswa terhadap pembelajaran daring menggunakan algoritma K-Nearest Neighbor (KNN), yang membantu dalam menilai kepuasan dan tantangan yang dihadapi oleh siswa selama pembelajaran online [12]. Selain itu, analisis sentimen juga dapat digunakan untuk menilai efektivitas metode pengajaran atau materi pembelajaran dengan mengolah feedback atau umpan balik yang diberikan oleh siswa dalam bentuk komentar atau ulasan.

Dengan demikian, analisis sentimen memiliki potensi yang besar untuk digunakan dalam berbagai bidang, mulai dari pemasaran, kebijakan publik, hingga pendidikan, dengan memberikan informasi yang sangat berguna dalam memahami opini dan perasaan masyarakat atau pengguna terhadap suatu topik atau isu.

* 1. **Mahasiswa**

1. Definisi Mahasiswa

Mahasiswa adalah peserta didik yang terdaftar dan belajar pada perguruan tinggi. Seorang mahasiswa dikategorikan pada tahap perkembangan yang usianya 18 sampai 25 tahun [13]. Tahap ini dapat digolongkan pada masa remaja akhir sampai masa dewasa awal dan dilihat dari segi perkembangan, tugas perkembangan pada usia mahasiswa ini ialah pemantapan pendirian hidup [14]. Mahasiswa dapat di definisikan sebagai individu yang sedang menuntut ilmu ditingkat perguruan tinggi, baik negri maupun swasta atau lembaga lain yang setingkat dengan perguruan tinggi [15].

Berdasarkan uraian diatas dapat disimpulkan bahwa mahasiswa ialah seorang peserta didik berusia 18 sampai 25 tahun yang terdaftar dan menjalani pendidikannnya di perguruan tinggi baik dari akademik, politeknik, sekolah tinggi, institut dan universitas. Sedangkan dalam penelitian ini, subyek yang digunakan ialah dua mahasiswa yang berusia 23 tahun dan masih tercatat sebagai mahasiswa aktif.

Belajar di perguruan tinggi sangat berbeda dari belajar di sekolah [16], siswa lebih banyak berperan sebagai penerima ilmu pengetahuan sementara pengajar berfungsi sebagai fasilitator yang membantu mahasiswa mencapai tujuan pembelajaran yang telah disepakati. Berdasarkan beberapa pendapat ahli diatas, peneliti menyimpulkan bahwa mahasiswa adalah orang yang sedang menjalani pendidikan tinggi di sebuah universitas atau perguruan tinggi

1. Ciri-Ciri Mahasiswa

Menurut Mustofa et al. (2020) mahasiswa merupakan anggota masyarakat yang mempunyai ciri-ciri tertentu [18], antara lain:

1. Mempunyai kemampuan dan kesempatan untuk belajar diperguruan tinggi, sehingga dapat digolongkan sebagai kaum intelegansi.
2. Karena kesempatan yang ada, mahasiswa diharapkan nantinya dapat bertindak sebagai pemimpin yang mampu dan terampil, baik sebagai pemimpin masyarakat ataupun dalam dunia kerja.
3. Diharapkan dapat menjadi daya penggerakan yang dinamis bagi proses modernisasi.
4. Diharapkan dapat memasuki dunia kerja sebagai tenaga yang berkualitas.

Khairun & Nurmalamenguraikan beberapa ciri dari mahasiswa [15], yaitu sebagai berikut:

1. Menerima keadaan fisiknya; perubahan fisiologis dan organis yang sedemikian hebat pada tahun-tahun sebelumnya, pada masa remaja akhir sudah tenang. Struktur dan penampilan fisik sudah menatap dan harus diterima sebagaimana adanya. Kekecewaan karena kondisi fisik tertentu tidak lagi mengunggu dan sedikit demi sedekit mulai menerima keadaan.
2. Memperoleh kebebasan emosional; masa remaja akhir sedang pada masa proses melepaskan diri dari ketergantungan secara emosional dari orang yang dekat dalam hidupnya (orangtua). Kehidupan emosi yang sebelumnya banyak mendominasi sikap dan tindakannya mulai terintegrasi dengan fungsi-fungsi lain sehingga lebih stabil dan lebih terkendali. Dia mampu mengungkapkan pendapat dan perasaannya dengan sikap yang sesuai dengan lingkungan dan kebebasan emosionalnya
3. Mampu bergaul; dia mulai mengembangkan kemampuan mengadakan hubungan sosial baik dengan teman sebaya maupun orang lain yang berbeda tingkat kematangan sosialnya. Dia mampu menyesuaikan dan memperlihatkan kemampuan bersosialisasi dalam tingkat kematangan sesuai dengan norma sosial yamg ada.
4. Menemukan model untuk identifikasi; dalam proses ke arah kematangan pribadi, tokoh indentifikasi sering kali menjadi faktor penting, tanpa tokoh identifikasi timbul kekaburan akan model yang akan ditiru dan memberikan pengarahan bagamana bertingkah laku dan bersikap sebaikbaiknya.
5. Mengetahui dan menerima kemampuan sendiri; pengertian dan penelitian yang objektif mengenai keadaan diri sendiri mulai terpuruk. Kekurngan dan kegagalan yang bersumber pada keadaan kemampuan tidak lagi mengganggu berfungsinya kepribadian dan menghambat prestasi yang ingin dicapai
6. Memperkuat penguasaan diri atas dasar skala nilai dan norma; nilai pribadi yang tadinya menjadi norma dalam melakukan sesuatu tindakan bergeser ke arah penyesuaian terhadap norma di luar dirinya. Baik yang berhubungan dengan nilai sosial ataupun moral. Nilai pribadi adakalanya harud disesuaikan dengan nilai-nilai umum (positif) yang berlaku dilingkungannya.
7. Meninggalkan reaksi dan cara penyesuaian kanak-kanakan; dunia remaja mulai ditinggalkan dan dihadapannya terbentang dunia dewasa yang akan dimasuki. Ketergantungan secara psikis mulai ditinggalkan dan ia mampu mengurus dan menentukan sendiri. Dapat dikatakan masa ini ialah masa persiapan ke arah tahapan perkembangan berikutnya yakni masa dewasa muda
8. Peran dan Fungsi Mahasiswa Sebagai mahasiswa

berbagai macam label pun disandang, ada beberapa macam label yang melekat pada diri mahasiswa, misalnya:

1. *Direct Of Change*, mahasiswa bisa melakukan perubahan langsung karena sumber daya manusianya yang banyak.
2. *Agent Of Change*, mahasiswa adalah perubahan, maksudnya sumber daya manusia untuk melakukan perubahan.
3. *Iron Stock*, sumber daya manusia dari mahasiswa itu tidak akan pernah habis.
4. *Moral Force*, mahasiswa itu kumpulan orang yang memiliki moral yang baik.
5. *Social Control*, mahasiswa itu pengontrol kehidupan sosial, seperti mengontrol kehidupan sosial yang dilakukan masyarakat [20].

Namun secara garis besar, setidaknya ada 3 peran dan fungsi yang sangat penting bagi mahasiwa, yaitu :

1. Peranan Moral Dunia kampus merupakan dunia di mana setiap mahasiswa dengan bebas memilih kehidupan yang mereka mau. Disinilah dituntut suatu tanggung jawab moral terhadap diri masing-masing sebagai indidu untuk dapat menjalankan kehidupan yang bertanggung jawab dan sesuai dengan moral yang hidup dalam masyarakat.
2. Peranan Sosial Selain tanggung jawab individu, mahasiswa juga memiliki peranan sosial, yaitu bahwa keberadaan dan segala perbuatannya tidak hanya bermanfaat untuk dirinya sendiri tetapi juga harus membawa manfaat bagi lingkungan sekitarnya.
3. Peranan Intelektual Mahasiswa sebagai orang yang disebut-sebut sebagai insan intelek haruslah dapat mewujudkan status tersebut dalam ranah kehidupan nyata. Dalam arti menyadari betul bahwa fungsi dasar mahasiswa adalah bergelut dengan ilmu pengetahuan dan memberikan perubahan yang lebih baik dengan intelektualitas yang ia miliki selama menjalani pendidikan [21].
4. Karakteristik Perkembangan Mahasiswa

Seperti halnya transisi dari sekolah dasar menuju sekolah menengah pertama yang melibatkan perubahan dan kemungkinan stres, begitu pula masa transisi dari sekolah menengah atas menuju universitas. Dalam banyak hal, terdapat perubahan yang sama dalam dua transisi itu. Transisi ini melibatkan gerakan menuju satu struktur sekolah yang lebih besar dan tidak bersifat pribadi, seperti interaksi dengan kelompok sebaya dari daerah yang lebih beragam dan peningkatan perhatian pada prestasi dan penilaiannya [22].

Perguruan tinggi dapat menjadi masa penemuan intelektual dan pertumbuhan kepribadian. Mahasiswa berubah saat merespon terhadap kurikulum yang menawarkan wawasan dan cara berpikir baru seperti; terhadap mahasiswa lain yang berbeda dalam soal pandangan dan nilai, terhadap kultur mahasiswa yang berbeda dengan kultur pada umumnya dan terhadap anggota fakultas yang memberikan model baru. Pilihan perguruan tinggi dapat mewakili pengejaran terhadap hasrat yang menggebu atau awal dari karir masa depan [23]. Ciri-ciri perkembangan remaja lanjut atau remaja akhir (usia 18 sampai 21 tahun) dapat dilihat dalam tugas-tugas perkembangan.

1. Faktor-faktor yang Mempengaruhi Sentimen Mahasiswa terhadap Tugas Kuliah

Sentimen mahasiswa terhadap tugas kuliah dipengaruhi oleh berbagai faktor, termasuk tingkat kesulitan tugas, beban kerja, dukungan dari dosen, serta pemahaman terhadap materi yang diajarkan. Cai (2023) menjelaskan bahwa tugas yang dianggap terlalu sulit atau tidak relevan dengan pembelajaran dapat menimbulkan perasaan frustrasi atau kebosanan di kalangan mahasiswa, yang pada gilirannya mempengaruhi sentimen mereka secara negatif [8]. Selain itu, kurangnya komunikasi yang efektif antara dosen dan mahasiswa dapat memperburuk persepsi mahasiswa terhadap tugas kuliah, membuat mereka merasa tidak dihargai atau tidak didukung [10].

Faktor-faktor lain yang mempengaruhi sentimen ini adalah fleksibilitas waktu pengerjaan tugas, serta adanya kesempatan untuk memberikan umpan balik atau bertanya kepada [12]. Mahasiswa cenderung memiliki sentimen positif terhadap tugas yang dianggap menantang namun bermanfaat dan relevan dengan tujuan pembelajaran mereka. Sebaliknya, tugas yang diberikan tanpa pertimbangan beban yang ada atau yang dianggap terlalu monoton dan tidak menambah wawasan akademik dapat menurunkan semangat dan kepuasan mahasiswa [9].

* 1. **K-Nearest Neighbors**

1. Pengertian K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) adalah sebuah metode untuk mengklasifikasikan objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut [24], Teknik ini sangat sederhana dan mudah diimplementasikan.

Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data pembelajaran. Sebuah titik pada ruang ini ditandai kelas c jika kelas c merupakan klasifikasi yang paling banyak ditemui pada k buah tetangga terdekat titik tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak Euclidean. Pendapat lain mengungkapkan jika Algoritma Nearest Neighbor adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada [25].

Algoritma K-NN adalah suatu metode yang menggunakan algoritma *supervised*. Perbedaan antara *supervised learning* dengan *unsupervised* *learning* adalah pada *supervised learning* bertujuan untuk menemukan pola baru dalam data dengan menghubungkan pola data yang sudah ada dengan data yang baru. Sedangkan pada *unsupervised learning*, data belum memiliki pola apapun, dan tujuan *unsupervised learning* untuk menemukan pola dalam sebuah data. Tujuan dari algoritma K-NN adalah untuk mengklasifikasi objek baru berdasarkan atribut dan *training samples* [26]*.*

*K-Nearest Neighbor* berupa sebuah metode non parametrik yang dapat diterapkan untuk mengklasifikasi berdasarkan tetangga k terdekat dan regresi. Algoritma K-Nearest Neighbor merupakan metode untuk mengklasifikasikan objek yang paling dekat dengan objek berdasarkan data pembelajaran [27]. Algortima *K-Nearest Neighbor* bekerja dengan mencari jarak antara data yang akan dievaluasi (*data training*) dan himpunan terdekat dari k tetangga (*neighbor*) terdekat dalam data baru (*data testing*). Ketika masuk dalam data kategori baru yang tidak diketahui untuk diklasifikasikan, maka kategori data baru tersebut harus ditentukan kategorinya berdasarkan sampel yang lain.

Karakteristik data yang akan diklasifikasikan wajib diekstraksi dan dibandingkan dengan karakteristik dari setiap data kategori yang telah diketahui dalam data testing, kemudian data tetangga terdekat k harus diambil dari data testing untuk menghitung kategori yang mana sebagian besar data berada [25]. Paduan nilai terdekat berdasarkan dengan prosedur perkiraan waktu yang diharapkan, yang kemudian digunakan sebagai nilai masa depan seperti yang diharapankan. Prediksi berdasarkan K-Nearest Neighbor mengacu pada pola urutan yang diamati berulang dari waktu ke waktu [29]. Jika pola perilaku sebelumnya dapat diidentifikasi yang mirip dengan pola perilaku deret waktu saat ini, nilai-nilai selanjutnya berdasarkan nilai sebelumnya dapat digunakan untuk memprediksi pola atau nilai perilaku pada periode mendatang [30].

Algoritma *Nearest Neighbor* ialah algoritma yang dilakukan untuk memperoleh kedekatan dalam memperoleh kasus saat memperhitungkan kedekatan kasus yang baru dengan kasus yang lama, yakni berlandaskan dengan mencocokkan kesesuaian bobot atas jumlah fitur yang telah ada [31]. *K-Nearest Neighbor* merupakan algoritma dengan tujuan mengklasifikasikan obyek data baru, bentuk dari proses tersebut dilatih dengan berdasarkan berbagai atribut serta data sampel latih. Permodelan atas pengelompokkan dari obyek tersebut hanya bisa didasarkan terhadap memori. Pengapliasian metode tersebut hanya bisa berjalan dengan mencari sejumlah k obyek data (data latih) yang mendekati atau terdekat atas data uji yang disajikan, selanjutnya dipilih klasifikasi menggunakan total *voting* paling banyak.

Cara tersebut dianggap sederhana serta sangat mudah diterapkan dalam penggunaannya. Berbagai data pembelajaran yang diproyeksikan dalam ruang dimensi memiliki banyak ruang, sehingga tereprentasi dimensi dengan berbagai fitur atas data tersebut. Dimensi ruang tersebut kemudian dibagi atas berbagai bagian klasifikasi dari berbagai data pembelajaran tersebut. Titik dari ruang dimensi tersebut diberi tanda dengan kelas c, apabila kelas c berupa sebuah klasifikasi yang terbanyak ditemui terhadap k buah tetangga yang dekat dari titik tersebut, maka titik dekat atau titik jauh dari tetangga kebanyakan dihitung menggunakan jarak metriks Euclidean.

Adapun langkah-langkahnya yaitu:

1. Penentu parameter K (jumlah tetangga terdekat)
2. Penghitung jumlah kuadrat atas jarak dimensi *eucliden* obyek atas berbagai data uji coba yang telah diberikan
3. Mengurutkan perhitungan kuadrat dengan *ascending* (tinggi ke rendah)
4. Mengelompokkan kategori Y (klasifikasi *Nearest Neighbor* berdasarkan nilai K)
5. Penggunaan algoritma *Nearest Neighbor* dengan manyoritas sehingga bisa terprediksi oleh obyek yang baru

K dalam *Nearest Neighbor* merupakan banyaknya tetangga terdekat yang akan dipakai sebagai titik dalam melakukan klasifikasi terhadap data atau obyek baru. Untuk menentukan nilai dari k maka lebih baik memakai angka yang ganjil, Penentuan nilai k dalam algoritma klasifikasi *K-Nearest Neighbor* dapat dicari berdasarkan nilai sampel k terdekat 𝑘1, 𝑘2, … , 𝑘𝑠 . Semakin banyak data yang ada semakin kecil jumlah k yang dipilih, sedangkan jika ukuran dimensi data yang ada lebih besar, jumlah k yang dipilih harus lebih tinggi. Dalam menentukan nilai k, lebih baik menggunakan angka ganjil seperti k = 1, 3, 5, …, dst.

Nilai k harus memenuhi syarat yaitu k < N dimana N merupakan jumlah dari dataset latih, karena nilai k digunakan untuk mencari jumlah mayoritas dari kelas/label pada data latih maka nilai k tidak boleh lebih dari jumlah dataset latih [32]. Mencari tetangga terdekat atau jarak pada algoritma K-NN terdapat 5 cara yaitu Jarak Euclidean, Jarak Manhattan, Jarak Cosine, Jarak Correlation, Jarak Hamming. Jarak antara dua tetangga k terdekat berdasarkan nilai kemiripan dapat dihitung menggunakan jarak Euclidean yang didefinisikan sebagai berikut [33]:

*Dist* (*X*, *Y*) = (2.1)

dengan:

*dist* (X, Y) = Jarak antar obyek *x*1i dan*x*2i

*x*i = Data uji atau data *testing*

*y*i = Data *training*

*D* = dimensi data

*i*  = variabel data

Kinerja dari setiap model tetangga k terdekat dapat dicari dengan menggunakan alat ukur sebagai parameter, misalnya dengan menggunakan *Crossvalidation*. yang merupakan teknik untuk memvalidasi suatu kinerja dan akurasi. *K-Nearest Neighbor* memiliki kelebihan dan kekurangannya sebagai berikut:

1. Kelebihan *K-Nearest Neighbor*
2. Mudah diimplementasikan
3. Tidak perlu membangun model, menentukan banyak parameter atau membuat asumsi tambahan
4. Tangguh terhadap training data yang noise
5. Efektif apabila data latihnya besar
6. Kekurangan *K-Nearest Neighbor*
7. *K-Nearest Neighbor* perlu menentukan nilai yang tepat dari parameter k (jumlah dari tetangga terdekat).
8. Pembelajaran berdasarkan jarak. Namun jarak dan atribut yang digunakan tidak jelas dalam memperoleh hasil yang terbaik.
9. Biaya cukup tinggi karena diunakan untuk menghitung jarak setiap sampel uji keseluruhan sampel training

Algoritma *K-Nearest Neighbor* dapat dituliskan sebagai berikut [27]:

1. Menentukan nilai K tetangga terdekat
2. Menghitung jarak dari data baru dengan data training, ukuran dari jarak yang digunakan adalah jarak metrik Euclidean persamaan (2.1)
3. Mengurutkan antar jarak terdekat terhadap jarak yang jauh
4. Memeriksa kelas K dalam tetangga yang dekat
5. Kelas data baru sama dengan kelas mayoritas dengan tetangga terdekat

Transformasi data merupakan suatu proses pengubahan atribut data kedalam atribut data baru bentuk lain yang sesuai [34]. Dalam penentuan kelas/label dari data indeks kedalaman keterlambatan pemabayaran dapat menggunakan proses transformasi data. Kriteria klasifikasi yang digunakan dalam penentuan kelas/label pada indeks Keterlambatan Pembayaran sebagai berikut [27]:

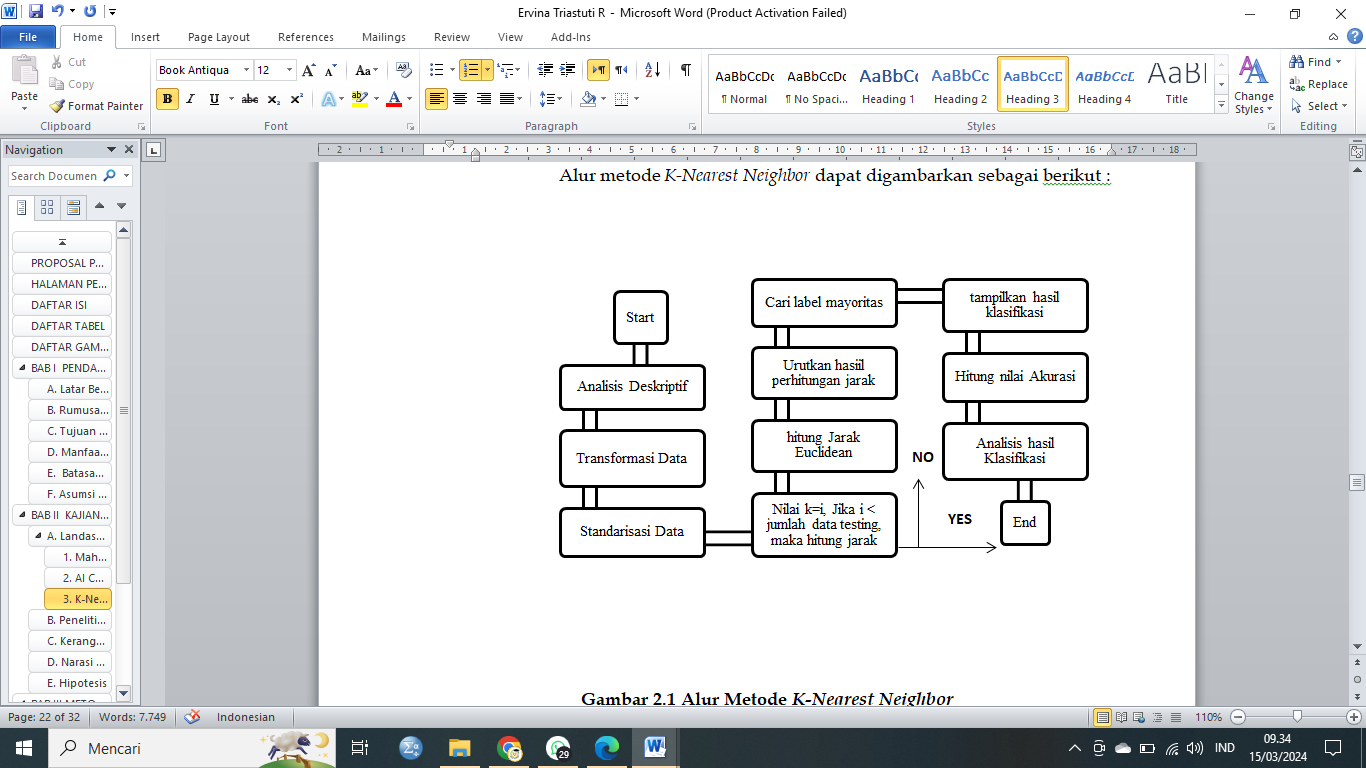
|  |  |
| --- | --- |
| **Label/Kelas** | **Indeks Sentimen Mahasiswa** |
| Kelas 0 | ISM < Rata-rata |
| Kelas 1 | ISM > Rata-rata |

Table 2.1 Indeks Sentimen Mahasiswa

Berikut merupakan rumus yang dapat digunakan dalam menentukan atau mencari nilai rata-rata:

(2.2)

Alur metode *K-Nearest Neighbor* dapat digambarkan sebagai berikut :



Gambar 2.1 Alur Metode *K-Nearest Neighbor*

Tahap dari penelitian yang dikerjakan dalam penelitian ini mencangkup berbagai prosedur pengerjaan serta dapat diperoleh garis bersar dengan menggunakan tahap *(fase),* diantaranya : (1) Start (2) Analisa Deskriptif (3) Transformasi Data (4) Standarisasi Data (5) Nilai k=i, Jika i < jumlah data testing, maka hitung jarak (apabila tidak sesuai maka selesai, namun jika sesuai terus berlanjut) (6) Hitung Jarak Euclidean (7) Hitung Jarang Euclidean (8) Urutkan Hasil Perhitungan Jarak (9) Cari Label Mayoritas (10) Tampilkan Hasil Klasifikasi (11) Hitung Nilai Akurasi (12) Analisis Hasil Klasifikasi (*End*).

## Penelitian Relevan

* 1. Penelitian oleh Tanggu Mara, A. A. P., Sediyono, E., & Purnomo, H. (2021)

Penelitian ini berjudul *"Penerapan algoritma K-Nearest Neighbors pada analisis sentimen metode pembelajaran dalam jaringan (DARING) di Universitas Kristen Wira Wacana Sumba"* yang diterbitkan dalam *JOINTER: Journal of Informatics Engineering* (Vol. 01, No. 02). Penelitian ini menggunakan KNN untuk menganalisis sentimen dari komentar mahasiswa mengenai pembelajaran daring. Hasil penelitian menunjukkan bahwa akurasi klasifikasi sentimen dengan KNN mencapai 85%, dengan sentimen positif mendominasi sebanyak 58%, sedangkan sentimen negatif hanya mencapai 22%, dan sisanya (20%) adalah sentimen netral [1].

* 1. Penelitian oleh Wenando, F. A., et al. (2022)

Dalam penelitian yang berjudul *“K-Nearest Neighbor (KNN) untuk menganalisis sentimen terhadap kebijakan Merdeka Belajar Kampus Merdeka pada komentar Twitter”*, peneliti menggunakan 3.000 tweet yang mengandung hashtag terkait kebijakan Merdeka Belajar. Model KNN yang digunakan memiliki akurasi 83%, dengan 54% tweet bernada positif, 32% bernada negatif, dan 14% netral. Model ini berhasil mengidentifikasi sentimen positif dengan recall sebesar 80% dan precision sebesar 85% pada sentimen negatif [4].

* 1. Penelitian oleh Riany, J., Fajar, M., & Lukman, M. P. (2016)

Penelitian ini berjudul *“Penerapan deep sentiment analysis pada angket penilaian terbuka menggunakan K-nearest neighbor”* yang diterbitkan dalam *Jurnal SISFO*. Dengan menggunakan 1.000 data angket terbuka, penelitian ini menemukan bahwa KNN dapat mengklasifikasikan sentimen positif dan negatif dengan akurasi 87%. Hasil uji f1-score menunjukkan angka tertinggi pada kelas sentimen positif, yaitu 85%, dengan precision dan recall masing-masing sebesar 83% dan 87% [35].

* 1. Penelitian oleh Putra, T. W., et al. (2022)

Dalam studi *“Analisis sentimen pembelajaran daring menggunakan metode Naïve Bayes, KNN, dan Decision Tree”*, yang diterbitkan dalam *Jurnal JTIK*, hasil penelitian menunjukkan bahwa KNN memiliki akurasi sebesar 90%, lebih tinggi dibandingkan Naïve Bayes (85%) dan Decision Tree (80%). Hasil analisis sentimen terhadap 2.500 data komentar mahasiswa menunjukkan 62% sentimen positif, 25% negatif, dan 13% netral, dengan KNN mampu memprediksi dengan baik sentimen positif dengan precision 91% [36].

* 1. Penelitian oleh Akhmad Deviyanto (2018)

Dalam penelitian *“Penerapan analisis sentimen pada pengguna Twitter menggunakan metode K-Nearest Neighbor”* yang diterbitkan dalam *Jurnal Ilmiah Sistem Komputer dan Aplikasi (JISKA)*, peneliti mengaplikasikan KNN untuk menganalisis 2.000 tweet yang terkait dengan topik tertentu. Penelitian ini menunjukkan akurasi model KNN mencapai 88%, dengan recall sebesar 85% dan precision sebesar 87%, yang mengindikasikan kemampuan KNN dalam mengklasifikasikan sentimen terhadap data teks [37].

* 1. Penelitian oleh Halimi, A., Kusrini, & Arief, M. R. (2021)

Penelitian berjudul *“Analisis Sentimen Masyarakat Indonesia Terhadap Pembelajaran Online dari Media Sosial Twitter Menggunakan Lexicon dan K-Nearest Neighbor”* ini menggunakan 5.000 tweet yang diambil dari Twitter. Hasil analisis menunjukkan bahwa sentimen terhadap pembelajaran online sebagian besar negatif, dengan 55% tweet bernada negatif, 30% positif, dan 15% netral. KNN menghasilkan akurasi 82% dengan precision dan recall masing-masing sebesar 80% dan 83% [3].

* 1. Penelitian oleh Furqan, M., Sriani, & Sari, S. M. (2022)  
     Penelitian ini berjudul *“Analisis sentimen menggunakan K-Nearest Neighbor terhadap new normal masa Covid-19 di Indonesia”*, yang diterbitkan dalam *Techno.Com*. Peneliti menggunakan lebih dari 3.500 tweet terkait kebijakan new normal di Indonesia. Hasil penelitian menunjukkan bahwa KNN mampu mengklasifikasikan sentimen dengan akurasi 86%, dengan 55% sentimen positif dan 28% negatif. Model ini memiliki f1-score sebesar 84%, yang menunjukkan kemampuannya dalam mengklasifikasikan tweet secara akurat [7].
  2. Penelitian oleh Husdi, & Kamaruddin, A. (2024)  
     Dalam penelitian *“Auto labeling untuk analisis sentimen opini mahasiswa baru terhadap pembelajaran mata kuliah dengan algoritma K-Nearest Neighbor”* yang diterbitkan dalam *Jurnal Restikom*, hasilnya menunjukkan bahwa KNN memiliki akurasi 89% dalam mengklasifikasikan opini mahasiswa baru terhadap mata kuliah. Penelitian ini menggunakan 1.200 data opini mahasiswa dengan distribusi sentimen 60% positif, 25% negatif, dan 15% netral [2].
  3. Penelitian oleh Narulita, L. F. (2017)

Penelitian ini berjudul *“Analisa sentimen pada tinjauan buku dengan algoritma K-Nearest Neighbour”* yang diterbitkan dalam *KONVERGENSI*.

Dengan menggunakan 500 data ulasan buku, hasil penelitian menunjukkan bahwa KNN dapat mengklasifikasikan sentimen dengan akurasi 84%, dengan 70% sentimen positif, 18% negatif, dan 12% netral. Precision pada sentimen positif mencapai 79%, dengan recall sebesar 80% [38].

* 1. Penelitian oleh Supriyanto, J., Alita, D., & Isnain, A. R. (2023)  
     Penelitian berjudul *“Penerapan algoritma K-Nearest Neighbor (K-NN) untuk analisis sentimen publik terhadap pembelajaran daring”* ini menggunakan lebih dari 3.000 data komentar terkait pembelajaran daring. Hasilnya menunjukkan bahwa KNN dapat mengklasifikasikan sentimen dengan akurasi 91%, dengan 64% komentar bernada positif, 26% negatif, dan 10% netral. Precision dan recall untuk sentimen positif masing-masing mencapai 93% dan 90% [12].

## Kerangka Konseptual

Kerangka konseptual penelitian ini menggambarkan alur hubungan antara berbagai tahapan yang dilalui dalam analisis sentimen terhadap tugas kuliah mahasiswa, mulai dari pengumpulan data hingga hasil klasifikasi sentimen. Penelitian ini memanfaatkan data dari platform Twitter (X), yang kemudian diolah melalui beberapa tahapan. Tahapan tersebut meliputi preprocessing data, ekstraksi fitur, dan penggunaan algoritma K-Nearest Neighbor (KNN) untuk mengklasifikasikan sentimen.

Tahapan pertama adalah pengumpulan data, di mana tweet dari mahasiswa terkait tugas kuliah diambil menggunakan API resmi Twitter. Data ini merupakan sumber utama dalam penelitian untuk menggambarkan opini mahasiswa terhadap tugas mereka. Selanjutnya, data tersebut melalui proses preprocessing, yang mencakup pembersihan teks, tokenisasi, normalisasi, dan ekstraksi fitur. Ekstraksi fitur dilakukan menggunakan teknik seperti TF-IDF atau word embeddings untuk merepresentasikan teks dalam bentuk numerik yang dapat dianalisis oleh model.

Kemudian, data yang telah diproses dimasukkan ke dalam model klasifikasi berbasis algoritma K-Nearest Neighbor (KNN). Algoritma ini digunakan untuk mengklasifikasikan sentimen menjadi tiga kategori utama: positif, negatif, dan netral. Proses ini mencakup pelatihan model menggunakan data yang telah dilabeli dan klasifikasi data baru berdasarkan kedekatan jarak fitur dalam ruang vektor.

Hasil akhir dari klasifikasi tersebut memberikan wawasan tentang persepsi mahasiswa terhadap tugas kuliah, yang terbagi menjadi tiga kategori:

1. Sentimen Positif: Menunjukkan pandangan atau perasaan positif terhadap tugas.
2. Sentimen Negatif: Menunjukkan ketidakpuasan atau pandangan negatif.
3. Sentimen Netral: Menunjukkan sikap netral tanpa opini yang kuat.

Berikut adalah diagram kerangka konseptual penelitian ini yang menunjukkan hubungan antar tahapan secara visual:

Data Input

Preprocessing

Feature

Extravtion

Tokenisasi

Normalisasi Teks

Pembersihan Data

Hasil Klasifikasi

Model KKN

KKN

Netral

Positif

Negatif

Gambar 2.2 Kerangka Konseptual

## Hipotesis

Hipotesis dalam penelitian dirumuskan untuk menjawab permasalahan proposal penelitian ini, Hipotesis perlu melalui pengujian untuk diuji dan dibuktikan kebenarannya berdasarkan data yang telah diperoleh dari sampel penelitian [39]. Hipotesis dalam penelitian ini adalah:

*Sentimen mahasiswa terhadap tugas kuliah cenderung lebih negatif daripada positif atau netral*.

Hipotesis ini berangkat dari asumsi bahwa banyak mahasiswa menganggap tugas kuliah sebagai beban yang berat, yang dapat mempengaruhi sentimen mereka. Berdasarkan penelitian sebelumnya, sentimen negatif sering kali lebih dominan terkait dengan beban akademik [2].

# BAB III METODE PENELITIAN

## A. Jenis dan Desain Penelitian

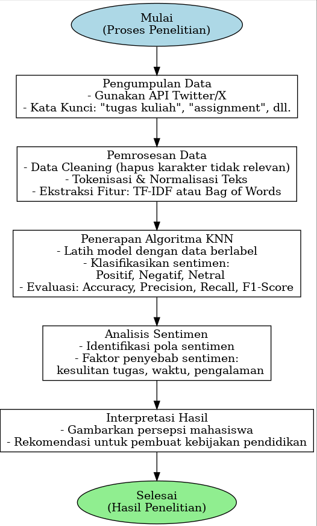
1. Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *Machine learning*, khususnya algoritma K-Nearest Neighbor (KNN), untuk menganalisis sentimen mahasiswa terhadap tugas kuliah yang dibagikan melalui platform media sosial Twitter/X. Pendekatan kuantitatif dipilih karena memungkinkan analisis data yang bersifat numerik dan dapat diukur secara objektif, serta memberikan gambaran yang lebih komprehensif mengenai pandangan atau perasaan mahasiswa terhadap suatu fenomena [8].

1. Rancangan Penelitian

Rancangan penelitian ini bersifat eksploratif dengan tujuan untuk mengeksplorasi sentimen mahasiswa yang diungkapkan melalui komentar mereka terkait tugas kuliah yang diposting di platform media sosial Twitter/X. Pendekatan eksploratif dipilih karena penelitian ini tidak bertujuan untuk menguji suatu hipotesis tertentu, melainkan untuk memperoleh gambaran umum mengenai perasaan dan pandangan mahasiswa terhadap tugas kuliah dalam konteks media sosial. Penelitian ini akan mengumpulkan data berupa tweet-tweet yang relevan, kemudian mengklasifikasikan dan menganalisis sentimen di balik komentar-komentar tersebut menggunakan algoritma *Machine Learning*, khususnya algoritma K-Nearest Neighbor (KNN).

Tahapan penelitian ini dapat dibagi menjadi beberapa bagian yang saling terkait, yaitu:



Gambar 3. 1 Rancangan Penelitian

Tahapan-tahapan tersebut meliputi:

1. Pengumpulan Data: Data tweet dikumpulkan dari Twitter/X menggunakan API resmi berdasarkan kata kunci terkait tugas kuliah, seperti "tugas kuliah", "assignment", dan lainnya. Data yang diambil dibatasi dalam rentang waktu tertentu untuk menjaga relevansi.
2. Pemrosesan Data: Data mentah yang telah dikumpulkan melalui tahap preprocessing, mencakup pembersihan data, tokenisasi, normalisasi teks, dan transformasi teks menjadi format numerik menggunakan metode seperti TF-IDF.
3. Penerapan Algoritma KNN: Algoritma KNN diterapkan untuk mengklasifikasikan sentimen tweet ke dalam tiga kategori: positif, negatif, dan netral. Model dievaluasi menggunakan metrik kinerja seperti akurasi dan F1-score.
4. Analisis Sentimen: Hasil klasifikasi dianalisis untuk mengidentifikasi pola sentimen dan faktor-faktor yang memengaruhinya, memberikan wawasan yang lebih dalam tentang persepsi mahasiswa terhadap tugas kuliah.
5. Interpretasi Hasil: Tahap akhir penelitian adalah menyimpulkan temuan utama dan memberikan rekomendasi berdasarkan hasil analisis.

## B. Tempat dan Waktu Penelitian

1. Tempat Penelitian

Penelitian ini dilakukan secara daring dengan sumber data utama yang berasal dari platform Twitter/X. Penelitian ini akan mencakup tweet yang relevan yang diposting dalam rentang waktu tertentu untuk mendapatkan gambaran yang representatif tentang pandangan mahasiswa terhadap tugas kuliah.

Twitter/X dipilih juga karena memiliki berbagai fitur yang mendukung analisis data, seperti hashtag, mention, dan berbagai ekspresi dalam bentuk teks yang memudahkan pengumpulan data yang terstruktur dan relevan [9].

1. Waktu Penelitian

Table 3.1 Waktu Penelitian

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Kegiatan | November 2024 | Desember 2024 | Desember 2024 | Januari 2025 | Februari 2025 |
| Pengumpulan Data |  |  |  |  |  |
| Pengembangan Model |  |  |  |  |  |
| Implementasi |  |  |  |  |  |
| Analisa Hasil |  |  |  |  |  |
| Pembuatan Laporan |  |  |  |  |  |

## C. Populasi dan Sampel

Populasi dalam penelitian ini adalah semua mahasiswa yang aktif memberikan opini atau komentar terkait tugas kuliah di platform Twitter/X. Populasi ini terdiri dari berbagai kalangan mahasiswa dari berbagai universitas dan disiplin ilmu yang menggunakan Twitter/X untuk berbagi pengalaman akademik mereka.

Sampel yang digunakan dalam penelitian ini adalah data 1000 tweet yang relevan dengan topik tugas kuliah. Pemilihan sampel ini bertujuan untuk mendapatkan representasi yang cukup luas mengenai pandangan mahasiswa terhadap tugas kuliah. Sampel akan dipilih secara acak dari tweet yang memenuhi kriteria tertentu, seperti penggunaan kata kunci yang relevan dan tingkat relevansi tweet dengan topik penelitian.

## E. Instrumen Penelitian

Instrumen penelitian utama dalam studi ini adalah dataset tweet yang dikumpulkan dari Twitter/X menggunakan teknik web scraping melalui API Twitter. Instrumen-instrumen yang digunakan dalam penelitian ini adalah sebagai berikut:

1. **Dataset**: Dataset penelitian terdiri dari tweet yang relevan dengan topik tugas kuliah, yang dikumpulkan menggunakan API resmi dari Twitter/X. Data akan diperoleh melalui proses *scraping* yang mencakup pengumpulan tweet berdasarkan kata kunci yang telah ditentukan. Proses ini akan dilakukan dengan memperhatikan etika penggunaan data dan batasan yang ditetapkan oleh platform Twitter/X.
2. **Validasi Dataset**: Setelah pengumpulan data, dataset yang terkumpul akan divalidasi oleh ahli bahasa dan pakar dalam bidang data science untuk memastikan akurasi dan relevansi dalam klasifikasi sentimen. Validasi ini dilakukan agar hasil analisis dapat dipertanggungjawabkan dan memiliki keandalan yang tinggi. Validasi bertujuan untuk memastikan bahwa tweet yang dikumpulkan relevan dengan topik tugas kuliah dan bahwa teks yang diproses telah disaring dengan benar dari tweet yang tidak relevan [10].
3. **Alat Analisis**: Untuk menganalisis sentimen, penelitian ini akan menggunakan perangkat lunak dan pustaka pemrograman seperti Python dengan pustaka *scikit-learn* untuk implementasi algoritma KNN. Selain itu, pustaka lainnya seperti *NLTK* (Natural Language Toolkit) dan *spaCy* akan digunakan untuk pemrosesan bahasa alami (NLP) dan klasifikasi sentimen.
4. **Instrumen Evaluasi**: Evaluasi model sentimen akan dilakukan menggunakan metrik *precision*, *recall*, *accuracy*, dan *F1-score* untuk mengukur kinerja model KNN dalam mengklasifikasikan tweet berdasarkan sentimen yang terkandung di dalamnya. Metrik evaluasi ini digunakan untuk memastikan bahwa hasil analisis sentimen dapat dipercaya dan menggambarkan pandangan mahasiswa secara akurat.

## F. Teknik Pengumpulan Data

Start

Scraping Data

Preprocessing Data

End

Gambar 3. 2 Teknik Pengumpulan Data

Terdapat beberapa teknik dalam pengumpulan data yang dilakukan oleh peneliti. Teknik pengumpulan data yang ada di dalam pengumpulan data ini didapatkan dari berbagai sumber yang berkaitan dengan penelitian yang sedang dilakukan oleh peneliti. Teknik pengumpulan data merupakan langkah yang paling strategis dalam penelitian, karena tujuan utama dari penelitian adalah mendapatkan data untuk diteliti lebih lanjut [40]. Teknik pengumpulan data yang dilakukan peneliti dalam penelitian ini adalah sebagai berikut:

* 1. *Scraping Data*:

Start

Access Twitter API

Set Parameters: Menentukan kata kunci, rentang tanggal (1 Desember 2023 - 30 November 2024), dan jumlah tweet (1000 tweet).

Fetch Data: Mengambil data sesuai parameter.

Preprocess Data:

Menghapus duplikasi dan teks yang tidak relevan.

Save Data:

Menyimpan data

End

Gambar 3. 3 Scraping Data

Data tweet akan diambil menggunakan API resmi Twitter/X. Menurut Cai (2023, p. 3), API Twitter merupakan metode yang dapat diandalkan untuk mengumpulkan data teks dari platform sosial media seperti Twitter, yang sangat berguna dalam analisis sentimen dan pemrosesan bahasa alami [8].

* 1. *Preprocessin*g Data:

Start

**Tokenisasi**: Memecah teks menjadi kata-kata.

**Penghapusan Stopwords**: Menghapus kata tidak penting.

**Stemming:** Mengubah kata ke bentuk dasar.

****

**Pelabelan:** Memberi label sentimen (positif, negatif, netral)

End

Gambar 3. 4 Preprocessing Data

1. *Tokenization*:

Memecah data teks menjadi token-token (kata atau frasa) yang lebih kecil untuk memudahkan analisis lebih lanjut [11].

1. Penghapusan *Stopwords*:

Stopword seperti kata penghubung atau kata yang sering muncul namun tidak memberikan banyak informasi (seperti "dan", "di", "itu") akan dihapus agar analisis lebih fokus pada kata yang lebih bermakna.

1. *Stemming*:

Mengubah kata menjadi bentuk dasarnya (misalnya "berlari" menjadi "lari") [9].

1. Pelabelan Data (*Labeling*):

Pelabelan sentimen pada tweet yang telah diproses menjadi kategori seperti positif, negatif, atau netral untuk memudahkan analisis lebih lanjut.

## G. Teknik Analisis Data

* 1. Pembobotan Fitur:

Metode **Term Frequency-Inverse Document Frequency (TF-IDF)** digunakan untuk memberikan bobot pada setiap kata dalam tweet, yang akan membantu dalam menentukan pentingnya kata tersebut dalam konteks analisis sentimen. TF-IDF memperhitungkan seberapa sering kata muncul dalam tweet tertentu serta seberapa jarang kata tersebut muncul secara keseluruhan dalam dataset, yang membantu untuk menyoroti kata-kata yang lebih informatif [38].

* 1. Klasifikasi dengan K-Nearest Neighbor (KNN):

Setelah pembobotan fitur dilakukan, algoritma K-Nearest Neighbor (KNN) akan digunakan untuk mengklasifikasikan tweet ke dalam tiga kategori sentimen utama: positif, negatif, dan netral. Dalam KNN, setiap tweet yang memiliki kata atau konteks serupa dengan tweet lainnya akan dikelompokkan dalam kategori yang sama. Metode ini terbukti efektif dalam analisis sentimen karena kemampuannya untuk memahami hubungan antara kata-kata dan konteks dalam data teks [12]. Selain itu, metode KNN juga memungkinkan evaluasi menggunakan teknik k-fold*cross-validation* untuk mengukur akurasi dan menghindari overfitting [12].

* 1. Evaluasi Model dengan K-Fold *Cross Validation*:

Model yang dibangun menggunakan algoritma KNN akan dievaluasi dengan teknik k-fold cross-validation. Teknik ini membagi data menjadi beberapa subset, dan model akan dilatih serta diuji pada masing-masing subset untuk memastikan generalisasi model yang lebih baik dan meningkatkan akurasi [10]. Ini adalah langkah penting untuk memvalidasi ketepatan klasifikasi sentimen yang dilakukan oleh model.

* 1. Visualisasi Hasil:

Setelah klasifikasi sentimen dilakukan, hasil analisis akan disajikan dalam bentuk visual seperti grafik batang, diagram lingkaran, atau tabel distribusi sentimen untuk memudahkan interpretasi. Visualisasi ini akan memperlihatkan perbandingan antara jumlah tweet yang termasuk dalam kategori positif, negatif, dan netral, sehingga hasil penelitian dapat dipahami dengan lebih jelas [1].

## H. Blok Diagram Penelitian

Proses penelitian ini akan diilustrasikan dalam bentuk blok diagram sebagai berikut:

1. **Input:** Data tweet yang diperoleh melalui API resmi Twitter/X.
2. **Preprocessing:** Data yang terkumpul akan diproses melalui tahap pembersihan (cleaning), tokenisasi, penghapusan stopword, stemming, dan pelabelan data untuk mempersiapkan data bagi analisis lebih lanjut.
3. **Proses:** Penggunaan algoritma K-Nearest Neighbor (KNN) untuk mengklasifikasikan tweet ke dalam kategori sentimen (positif, negatif, atau netral), dengan evaluasi menggunakan teknik k-fold cross-validation.
4. **Output:** Klasifikasi sentimen tweet menjadi tiga kategori utama: positif, negatif, atau netral.



Gambar 3.5 Blok Diagram Penelitian

* + 1. Validasi dan Keterkaitan Data

1. **Validasi Instrumen**:

Untuk memastikan akurasi hasil klasifikasi, proses validasi instrumen dilakukan dengan melibatkan ahli bahasa dan ahli data science. Validasi ini bertujuan untuk menilai kualitas instrumen yang digunakan dalam penelitian, seperti proses pembersihan data (preprocessing), pelabelan sentimen, serta pengaplikasian algoritma K-Nearest Neighbor (KNN). Ahli bahasa memberikan masukan terkait ketepatan pemahaman konteks linguistik dalam tweet, sementara ahli data science memastikan bahwa pendekatan teknis, seperti pemilihan fitur dan algoritma yang digunakan, sudah sesuai dan efektif untuk analisis sentimen [12]. Validasi yang komprehensif ini penting untuk memastikan bahwa data yang digunakan relevan, akurat, dan sesuai dengan tujuan penelitian [9].

1. **Penyajian Data:**

Data yang diperoleh akan disajikan dengan jelas menggunakan tabel, grafik, dan visualisasi lainnya untuk menggambarkan proses analisis yang dilakukan. Tabel akan digunakan untuk menyajikan distribusi sentimen dalam setiap kategori (positif, negatif, dan netral), sedangkan grafik batang dan diagram lingkaran akan memvisualisasikan perbandingan frekuensi setiap sentimen. Penyajian ini bertujuan untuk memudahkan pembaca dalam memahami proses dan hasil penelitian secara lebih sistematis dan [11]. Selain itu, persamaan matematika yang menggambarkan perhitungan bobot fitur dalam metode TF-IDF dan evaluasi model menggunakan k-fold cross-validation akan disertakan untuk menunjukkan langkah-langkah teknis yang digunakan dalam analisis data [38].

Pentingnya penyajian data yang jelas dan terstruktur dapat membantu meningkatkan pemahaman tentang hasil analisis, terutama dalam penelitian yang melibatkan data besar dan kompleks [10].

# BAB IV PEMBAHASAN

## A. Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dari kumpulan tweet berbahasa Indonesia yang membahas opini mahasiswa mengenai tugas kuliah. *Tweet* tersebut dikumpulkan menggunakan API resmi Twitter/X dengan kata kunci yang relevan seperti “tugas kuliah”, “assignment”, “dosen”, dan “kampus” dalam rentang waktu 1 Desember 2023 hingga 30 November 2024. Data yang digunakan dalam penelitian ini terdiri dari dua jenis, yaitu data latih dan data uji. Data latih yang digunakan merupakan kumpulan *tweet* yang telah dilabeli secara manual dengan kategori sentimen positif, netral, dan negatif yang berjumlah 70% dari total data yang terkumpul. Data ini berfungsi sebagai acuan dalam membentuk model analisis sentimen dengan menggunakan algoritma K-*Nearest Neighbors* (KNN). Sementara itu, 30% dari data yang terkumpul akan digunakan sebagai data uji. Data uji ini terdiri dari *tweet* yang belum dilabeli sentimen dan akan digunakan untuk mengukur performa model yang telah dibangun. Rincian *dataset* yang didapatkan melalui proses data *scraping* dapat dilihat pada Tabel 4.1.

Table 4.1 Detail *Dataset* dari Hasil *Scraping* *Data*

|  |  |  |
| --- | --- | --- |
| Kategori Sentimen | Jumlah Data Latih | Jumlah Data Uji |
| Positif | 288 | 195 |
| Netral | 238 | 65 |
| Negatif | 174 | 40 |

## B. Pemrosesan Data

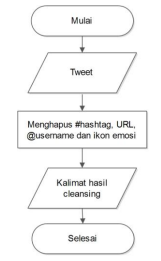
Pemrosesan data dilakukan untuk menghindari data yang tidak sempurna, *noise* pada data, dan data yang tidak konsisten. Adapun langkah-langkah pemrosesan data adalah *cleaning* *data*, tokenisasi, normalisasi teks, dan transformasi teks menjadi format numerik menggunakan metode seperti TF-IDF.

1. *Cleaning Data*

Proses *cleaning* *data* dilakukan untuk menghilangkan elemen-elemen yang tidak diperlukan dari data mentah yang telah dikumpulkan. Langkah ini dilakukan untuk mengurangi noise pada data yang dapat mempengaruhi hasil analisis sentimen.

Tahap *cleaning* *data* meliputi penghapusan karakter-karakter yang tidak relevan seperti *hashtag* (#), *username* (@username), URL (http://situs.com), ikon emosi, dan alamat email (nama@situs.com). Langkah ini bertujuan untuk memastikan hanya data teks yang relevan yang akan diproses lebih lanjut.

Diagram alir untuk proses *cleaning* *data* dapat dilihat pada Gambar 4.1.



Gambar 4.1 Diagram Alir *Cleaning*

Penjelasan dari diagram alir di atas adalah sebagai berikut:

1. Data masukan yang digunakan adalah *tweet* mentah.
2. *Tweet* akan diperiksa apakah mengandung karakter-karakter yang tidak relevan seperti *hashtag*, *username*, URL, ikon emosi, atau email.
3. Jika karakter-karakter tersebut ditemukan, maka karakter-karakter tersebut akan dihapus.

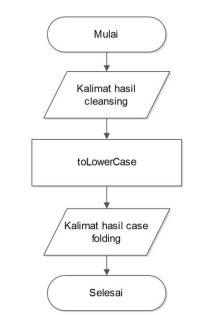
Hasil dari proses pembersihan data dapat dilihat pada Tabel 4.2.

Table 4.2 Hasil *Cleaning*

|  |  |  |
| --- | --- | --- |
| Kode | *Input Tweet* Asli | *Output Cleaning* |
| 1 | @aauniawatif Habis paper dah eh tepp? Mantappp!! Assignment aku pon baru settle | Habis paper dah eh tepp? Mantappp!! Assignment aku pon baru settle |
| 1 | @Celcom tolong periksa sikit line di Batu Satu Temerloh. Saya nak buat assignment pun terbantut ni. Terima kasi https://t.co/Y7p80BnROD | tolong periksa sikit line di Batu Satu Temerloh. Saya nak buat assignment pun terbantut ni . Terima kasi |
| 0 | @Farizizadi Ya faham faham aja pak dosen | Ya faham faham aja pak dosen |
| 1 | @MhdAfqWhb I belajar dulu Makan tidur berak muntah semua i fikir pasal ASSIGNMENT Takde nak pakai socmed lang https://t.co/r8eG6MyLQS | I belajar dulu Makan tidur berak muntah semua i fikir pasal ASSIGNMENT Takde nak pakai socmed lang |
| 0 | @VRRCAA Langsung wae BU DOSEN | Langsung wae BU DOSEN |
| 1 | Nak madah Im not used to being alone dari zaman study aku biasa sorg2 waima g makan sekalipun. Molah assignment k https://t.co/1egLtMOXQ1 | Nak madah Im not used to being alone dari zaman study aku biasa sorg2 waima g makan sekalipun. Molah assignment k |

1. Tokenisasi dan Normalisasi Teks

Sebelum melakukan tokenisasi, semua huruf akan diubah menjadi huruf kecil. Diagram alir untuk proses *case folding* dapat dilihat pada Gambar 3.3.



Gambar 4.3 Diagram Alir *Tokenisasi*

Penjelasan dari diagram alir di atas adalah sebagai berikut:

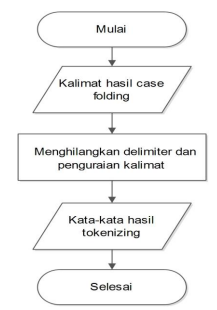
1. Kalimat yang digunakan adalah kalimat hasil *cleaning*.
2. Hasil *cleaning* akan diperiksa apakah karakter yang menggunakan huruf besar, maka huruf tersebut akan diubah menjadi huruf kecil.

Hasil dari *case folding* dapat dilihat pada Tabel 4.3.

Table 4.3 Hasil *Case Folding*

|  |  |  |
| --- | --- | --- |
| Kode | *Input Cleaning Data* | *Output Case Folding* |
| 1 | Habis paper dah eh tepp? Mantappp!! Assignment aku pon baru settle | habis paper dah eh tepp? mantappp!! assignment aku pon baru settle |
| 1 | tolong periksa sikit line di Batu Satu Temerloh. Saya nak buat assignment pun terbantut ni . Terima kasi | tolong periksa sikit line di batu satu temerloh. saya nak buat assignment pun terbantut ni . terima kasi |
| 0 | Ya faham faham aja pak dosen | ya faham faham aja pak dosen |
| 1 | I belajar dulu Makan tidur berak muntah semua i fikir pasal ASSIGNMENT Takde nak pakai socmed lang | i belajar dulu makan tidur berak muntah semua i fikir pasal assignment takde nak pakai socmed lang |
| 0 | Langsung wae BU DOSEN | langsung wae bu dosen |
| 1 | Nak madah Im not used to being alone dari zaman study aku biasa sorg2 waima g makan sekalipun. Molah assignment k | nak madah im not used to being alone dari zaman study aku biasa sorg2 waima g makan sekalipun. molah assignment k |

Pada tahap tokenisasi, kita akan memisahkan kalimat menjadi beberapa kata dan memeriksa kata dari karakter pertama hingga terakhir. Tokenisasi berarti memecah data teks menjadi token-token (kata atau frasa) yang lebih kecil untuk memudahkan analisis lebih lanjut. Jika karakter ke-i bukan merupakan tanda pemisah kata seperti titik(.), koma(,), spasi, dan tanda pemisah lainnya, maka karakter tersebut akan digabungkan dengan karakter berikutnya. Diagram alir untuk proses tokenisasi dapat dilihat pada Gambar 4.4.



Gambar 4.4 Diagram Alir *Tokenisasi*

Penjelasan dari diagram alir di atas adalah sebagai berikut:

1. Kalimat yang digunakan adalah kalimat hasil *case folding*.
2. Potong setiap kata pada kalimat tersebut berdasarkan pemisah kata seperti titik(.), koma(,), dan spasi. Bagian yang hanya memiliki satu karakter non-alfabet dan angka akan dibuang. Sehingga menghasilkan kata-kata yang membentuk kalimat.

Hasil dari tokenisasi dapat dilihat pada Tabel 4.4.

Table 4.4 Hasil Tokenisasi

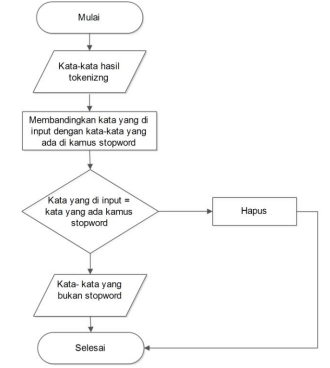
|  |  |  |
| --- | --- | --- |
| Kode | *Input Case Folding* | *Output* Tokenisasi |
| 1 | habis paper dah eh tepp? mantappp!! assignment aku pon baru settle | ['habis', 'paper', 'dah', 'eh', 'tepp', 'mantappp', 'assignment', 'aku', 'pon', 'baru', 'settle'] |
| 1 | tolong periksa sikit line di batu satu temerloh. saya nak buat assignment pun terbantut ni . terima kasi | ['tolong', 'periksa', 'sikit', 'line', 'di', 'batu', 'satu', 'temerloh', 'saya', 'nak', 'buat', 'assignment', 'pun', 'terbantut', 'ni', 'terima', 'kasi'] |
| 0 | ya faham faham aja pak dosen | ['ya', 'faham', 'faham', 'aja', 'pak', 'dosen'] |
| 1 | i belajar dulu makan tidur berak muntah semua i fikir pasal assignment takde nak pakai socmed lang | ['i', 'belajar', 'dulu', 'makan', 'tidur', 'berak', 'muntah', 'semua', 'i', 'fikir', 'pasal', 'assignment', 'takde', 'nak', 'pakai', 'socmed', 'lang'] |
| 0 | langsung wae bu dosen | ['langsung', 'wae', 'bu', 'dosen'] |
| 1 | nak madah im not used to being alone dari zaman study aku biasa sorg2 waima g makan sekalipun. molah assignment k | ['nak', 'madah', 'im', 'not', 'used', 'to', 'being', 'alone', 'dari', 'zaman', 'study', 'aku', 'biasa', 'sorg2', 'waima', 'g', 'makan', 'sekalipun', 'molah', 'assignment', 'k'] |

Setelah itu masuk ke normalisasi data di mana tahap pertama yang dilakukan adalah *stopword removal*. Setiap kata akan diperiksa. Jika ada kata sambung, kata depan, kata ganti, atau kata-kata yang tidak ada hubungannya dengan analisis sentimen, maka kata-kata tersebut akan dihapus. Pada penelitian ini menggunakan kamus *stopword* Tala dimana data berjumlah 758 data [41]. Kamus *stopword* milik Tala ini telah banyak banyak digunakan oleh para peneliti untuk menghapus kata-kata yang tidak penting dalam bahasa Indonesia. Untuk *stopword* bahasa inggris dari Stopwords ISO di github. Contoh *stopword* yang digunakan dapat dilihat pada Tabel 4.5.

Table 4.5 Kamus *Stopword*

|  |  |  |  |
| --- | --- | --- | --- |
| *Stopword* | | | |
| ada | hal | ‘ll | e |
| bagai | ia | 10 | f |
| cara | jadi | a | g |
| dahulu | kala | b | h |
| empat | lagi | c | i |
| guna | macam | d | j |

Diagram alir untuk proses *stopword removal* dapat dilihat pada Gambar 4.5.



Gambar 4.5 Diagram Alir *Stopword Removal*

Penjelasan dari diagram alir di atas adalah sebagai berikut:

1. Kata yang telah diberi token akan dibandingkan dengan daftar *stopword*.
2. Memeriksa apakah kata tersebut sama dengan daftar *stopword* atau tidak.
3. Jika kata tersebut sama dengan kata yang ada pada daftar *stopword*, maka kata tersebut maka kata tersebut akan dihapus.

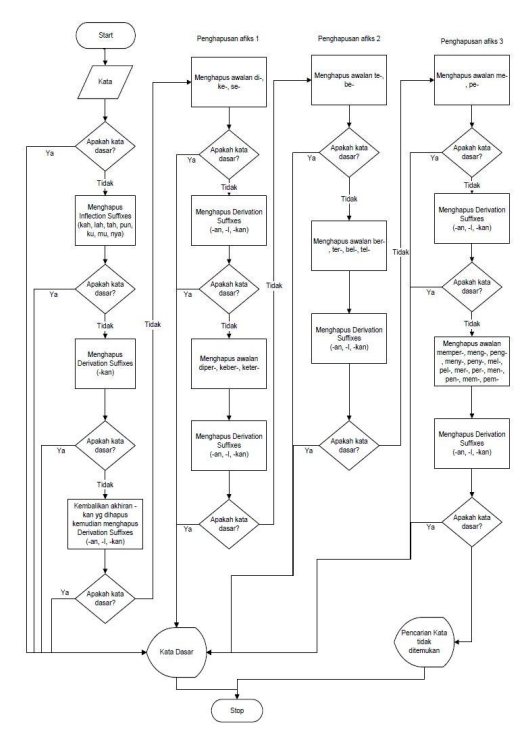
Hasil *stopword removal* dapat dilihat pada Tabel 4.6.

Table 4.6 Hasil *Stopword Removal*

|  |  |  |
| --- | --- | --- |
| Kode | *Input* Tokenisasi | *Output Stopword Removal* |
| 1 | ['habis', 'paper', 'dah', 'eh', 'tepp', 'mantappp', 'assignment', 'aku', 'pon', 'baru', 'settle'] | ['habis', 'paper', 'dah', 'tepp', 'mantappp', 'assignment', 'pon', 'settle'] |
| 1 | ['tolong', 'periksa', 'sikit', 'line', 'di', 'batu', 'satu', 'temerloh', 'saya', 'nak', 'buat', 'assignment', 'pun', 'terbantut', 'ni', 'terima', 'kasi'] | ['tolong', 'periksa', 'sikit', 'batu', 'temerloh', 'nak', 'assignment', 'terbantut', 'terima', 'kasi'] |
| 0 | ['ya', 'faham', 'faham', 'aja', 'pak', 'dosen'] | ['ya', 'faham', 'faham', 'aja', 'dosen'] |
| 1 | ['i', 'belajar', 'dulu', 'makan', 'tidur', 'berak', 'muntah', 'semua', 'i', 'fikir', 'pasal', 'assignment', 'takde', 'nak', 'pakai', 'socmed', 'lang'] | ['belajar', 'makan', 'tidur', 'berak', 'muntah', 'fikir', 'pasal', 'assignment', 'takde', 'nak', 'pakai', 'socmed', 'lang'] |
| 0 | ['langsung', 'wae', 'bu', 'dosen'] | ['langsung', 'wae', 'bu', 'dosen'] |
| 1 | ['nak', 'madah', 'im', 'not', 'used', 'to', 'being', 'alone', 'dari', 'zaman', 'study', 'aku', 'biasa', 'sorg2', 'waima', 'g', 'makan', 'sekalipun', 'molah', 'assignment', 'k'] | ['nak', 'madah', 'zaman', 'study', 'sorg2', 'waima', 'makan', 'molah', 'assignment'] |

Tahap selanjutnya dalam normalisasi data adalah *stemming. Stemming* adalah konversi kata menjadi kata dasar atau penghilangan imbuhan. Stemming di sini menggunakan daftar kamus kata berimbuhan yang memiliki kata dasar dengan membandingkan kata-kata dalam komentar dengan daftar kamus kata dasar. *Stemming* dilakukan dengan bantuan *library online* yaitu Sastrawi 1.0.1. dan NLTK.

Diagram alir untuk proses *stemming* dapat dilihat pada Gambar 4.6.



Gambar 4.6 Diagram Alir *Stemming*

Hasil dari *stemming* dapat dilihat pada Tabel 4.7.

Table 4.7 Hasil *Stemming*

|  |  |  |
| --- | --- | --- |
| Kode | *Input Stopword Removal* | *Output Stemming* |
| 1 | ['habis', 'paper', 'dah', 'tepp', 'mantappp', 'assignment', 'pon', 'settle'] | ['habis', 'paper', 'dah', 'tepp', 'mantappp', 'assign', 'pon', 'settl'] |
| 1 | ['tolong', 'periksa', 'sikit', 'batu', 'temerloh', 'nak', 'assignment', 'terbantut', 'terima', 'kasi'] | ['tolong', 'periksa', 'sikit', 'batu', 'temerloh', 'nak', 'assign', 'bantut', 'terima', 'kasi'] |
| 0 | ['ya', 'faham', 'faham', 'aja', 'dosen'] | ['ya', 'faham', 'faham', 'aja', 'dosen'] |
| 1 | ['belajar', 'makan', 'tidur', 'berak', 'muntah', 'fikir', 'pasal', 'assignment', 'takde', 'nak', 'pakai', 'socmed', 'lang'] | ['ajar', 'makan', 'tidur', 'berak', 'muntah', 'fikir', 'pasal', 'assign', 'takde', 'nak', 'pakai', 'socmed', 'lang'] |
| 0 | ['langsung', 'wae', 'bu', 'dosen'] | ['langsung', 'wae', 'bu', 'dosen'] |
| 1 | ['nak', 'madah', 'zaman', 'study', 'sorg2', 'waima', 'makan', 'molah', 'assignment'] | ['nak', 'madah', 'zaman', 'studi', 'sorg2', 'waima', 'makan', 'molah', 'assign'] |

1. Ekstraksi Fitur

Penelitian ini dilakukan secara daring dengan sumber data utama yang berasal dari platform Twitter/X. Penelitian ini akan mencakup tweet yang relevan yang diposting dalam rentang waktu tertentu untuk mendapatkan gambaran yang representatif tentang pandangan mahasiswa terhadap tugas kuliah.

Twitter/X dipilih juga karena memiliki berbagai fitur yang mendukung analisis data, seperti hashtag, mention, dan berbagai ekspresi dalam bentuk teks yang memudahkan pengumpulan data yang terstruktur dan relevan [9].

1. Waktu Penelitian

Table 3.1 Waktu Penelitian

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Kegiatan | November 2024 | Desember 2024 | Desember 2024 | Januari 2025 | Februari 2025 |
| Pengumpulan Data |  |  |  |  |  |
| Pengembangan Model |  |  |  |  |  |
| Implementasi |  |  |  |  |  |
| Analisa Hasil |  |  |  |  |  |
| Pembuatan Laporan |  |  |  |  |  |

## C. Populasi dan Sampel

Populasi dalam penelitian ini adalah semua mahasiswa yang aktif memberikan opini atau komentar terkait tugas kuliah di platform Twitter/X. Populasi ini terdiri dari berbagai kalangan mahasiswa dari berbagai universitas dan disiplin ilmu yang menggunakan Twitter/X untuk berbagi pengalaman akademik mereka.

Sampel yang digunakan dalam penelitian ini adalah data 1000 tweet yang relevan dengan topik tugas kuliah. Pemilihan sampel ini bertujuan untuk mendapatkan representasi yang cukup luas mengenai pandangan mahasiswa terhadap tugas kuliah. Sampel akan dipilih secara acak dari tweet yang memenuhi kriteria tertentu, seperti penggunaan kata kunci yang relevan dan tingkat relevansi tweet dengan topik penelitian.

## E. Instrumen Penelitian

Instrumen penelitian utama dalam studi ini adalah dataset tweet yang dikumpulkan dari Twitter/X menggunakan teknik web scraping melalui API Twitter. Instrumen-instrumen yang digunakan dalam penelitian ini adalah sebagai berikut:

1. **Dataset**: Dataset penelitian terdiri dari tweet yang relevan dengan topik tugas kuliah, yang dikumpulkan menggunakan API resmi dari Twitter/X. Data akan diperoleh melalui proses *scraping* yang mencakup pengumpulan tweet berdasarkan kata kunci yang telah ditentukan. Proses ini akan dilakukan dengan memperhatikan etika penggunaan data dan batasan yang ditetapkan oleh platform Twitter/X.
2. **Validasi Dataset**: Setelah pengumpulan data, dataset yang terkumpul akan divalidasi oleh ahli bahasa dan pakar dalam bidang data science untuk memastikan akurasi dan relevansi dalam klasifikasi sentimen. Validasi ini dilakukan agar hasil analisis dapat dipertanggungjawabkan dan memiliki keandalan yang tinggi. Validasi bertujuan untuk memastikan bahwa tweet yang dikumpulkan relevan dengan topik tugas kuliah dan bahwa teks yang diproses telah disaring dengan benar dari tweet yang tidak relevan [10].
3. **Alat Analisis**: Untuk menganalisis sentimen, penelitian ini akan menggunakan perangkat lunak dan pustaka pemrograman seperti Python dengan pustaka *scikit-learn* untuk implementasi algoritma KNN. Selain itu, pustaka lainnya seperti *NLTK* (Natural Language Toolkit) dan *spaCy* akan digunakan untuk pemrosesan bahasa alami (NLP) dan klasifikasi sentimen.
4. **Instrumen Evaluasi**: Evaluasi model sentimen akan dilakukan menggunakan metrik *precision*, *recall*, *accuracy*, dan *F1-score* untuk mengukur kinerja model KNN dalam mengklasifikasikan tweet berdasarkan sentimen yang terkandung di dalamnya. Metrik evaluasi ini digunakan untuk memastikan bahwa hasil analisis sentimen dapat dipercaya dan menggambarkan pandangan mahasiswa secara akurat.

## F. Teknik Pengumpulan Data

Start

Scraping Data

Preprocessing Data

End

Gambar 3. 2 Teknik Pengumpulan Data

Terdapat beberapa teknik dalam pengumpulan data yang dilakukan oleh peneliti. Teknik pengumpulan data yang ada di dalam pengumpulan data ini didapatkan dari berbagai sumber yang berkaitan dengan penelitian yang sedang dilakukan oleh peneliti. Teknik pengumpulan data merupakan langkah yang paling strategis dalam penelitian, karena tujuan utama dari penelitian adalah mendapatkan data untuk diteliti lebih lanjut [40]. Teknik pengumpulan data yang dilakukan peneliti dalam penelitian ini adalah sebagai berikut:

* 1. *Scraping Data*:

Start

Access Twitter API

Set Parameters: Menentukan kata kunci, rentang tanggal (1 Desember 2023 - 30 November 2024), dan jumlah tweet (1000 tweet).

Fetch Data: Mengambil data sesuai parameter.

Preprocess Data:

Menghapus duplikasi dan teks yang tidak relevan.

Save Data:

Menyimpan data

End

Gambar 3. 3 Scraping Data

Data tweet akan diambil menggunakan API resmi Twitter/X. Menurut Cai (2023, p. 3), API Twitter merupakan metode yang dapat diandalkan untuk mengumpulkan data teks dari platform sosial media seperti Twitter, yang sangat berguna dalam analisis sentimen dan pemrosesan bahasa alami [8].

* 1. *Preprocessin*g Data:

Start

**Tokenisasi**: Memecah teks menjadi kata-kata.

**Penghapusan Stopwords**: Menghapus kata tidak penting.

**Stemming:** Mengubah kata ke bentuk dasar.

****

**Pelabelan:** Memberi label sentimen (positif, negatif, netral)

End

Gambar 3. 4 Preprocessing Data

1. *Tokenization*:

Memecah data teks menjadi token-token (kata atau frasa) yang lebih kecil untuk memudahkan analisis lebih lanjut [11].

1. Penghapusan *Stopwords*:

Stopword seperti kata penghubung atau kata yang sering muncul namun tidak memberikan banyak informasi (seperti "dan", "di", "itu") akan dihapus agar analisis lebih fokus pada kata yang lebih bermakna.

1. *Stemming*:

Mengubah kata menjadi bentuk dasarnya (misalnya "berlari" menjadi "lari") [9].

1. Pelabelan Data (*Labeling*):

Pelabelan sentimen pada tweet yang telah diproses menjadi kategori seperti positif, negatif, atau netral untuk memudahkan analisis lebih lanjut.

## G. Teknik Analisis Data

* 1. Pembobotan Fitur:

Metode **Term Frequency-Inverse Document Frequency (TF-IDF)** digunakan untuk memberikan bobot pada setiap kata dalam tweet, yang akan membantu dalam menentukan pentingnya kata tersebut dalam konteks analisis sentimen. TF-IDF memperhitungkan seberapa sering kata muncul dalam tweet tertentu serta seberapa jarang kata tersebut muncul secara keseluruhan dalam dataset, yang membantu untuk menyoroti kata-kata yang lebih informatif [38].

* 1. Klasifikasi dengan K-Nearest Neighbor (KNN):

Setelah pembobotan fitur dilakukan, algoritma K-Nearest Neighbor (KNN) akan digunakan untuk mengklasifikasikan tweet ke dalam tiga kategori sentimen utama: positif, negatif, dan netral. Dalam KNN, setiap tweet yang memiliki kata atau konteks serupa dengan tweet lainnya akan dikelompokkan dalam kategori yang sama. Metode ini terbukti efektif dalam analisis sentimen karena kemampuannya untuk memahami hubungan antara kata-kata dan konteks dalam data teks [12]. Selain itu, metode KNN juga memungkinkan evaluasi menggunakan teknik k-fold*cross-validation* untuk mengukur akurasi dan menghindari overfitting [12].

* 1. Evaluasi Model dengan K-Fold *Cross Validation*:

Model yang dibangun menggunakan algoritma KNN akan dievaluasi dengan teknik k-fold cross-validation. Teknik ini membagi data menjadi beberapa subset, dan model akan dilatih serta diuji pada masing-masing subset untuk memastikan generalisasi model yang lebih baik dan meningkatkan akurasi [10]. Ini adalah langkah penting untuk memvalidasi ketepatan klasifikasi sentimen yang dilakukan oleh model.

* 1. Visualisasi Hasil:

Setelah klasifikasi sentimen dilakukan, hasil analisis akan disajikan dalam bentuk visual seperti grafik batang, diagram lingkaran, atau tabel distribusi sentimen untuk memudahkan interpretasi. Visualisasi ini akan memperlihatkan perbandingan antara jumlah tweet yang termasuk dalam kategori positif, negatif, dan netral, sehingga hasil penelitian dapat dipahami dengan lebih jelas [1].

## H. Blok Diagram Penelitian

Proses penelitian ini akan diilustrasikan dalam bentuk blok diagram sebagai berikut:

1. **Input:** Data tweet yang diperoleh melalui API resmi Twitter/X.
2. **Preprocessing:** Data yang terkumpul akan diproses melalui tahap pembersihan (cleaning), tokenisasi, penghapusan stopword, stemming, dan pelabelan data untuk mempersiapkan data bagi analisis lebih lanjut.
3. **Proses:** Penggunaan algoritma K-Nearest Neighbor (KNN) untuk mengklasifikasikan tweet ke dalam kategori sentimen (positif, negatif, atau netral), dengan evaluasi menggunakan teknik k-fold cross-validation.
4. **Output:** Klasifikasi sentimen tweet menjadi tiga kategori utama: positif, negatif, atau netral.



Gambar 3.5 Blok Diagram Penelitian

* + 1. Validasi dan Keterkaitan Data

1. **Validasi Instrumen**:

Untuk memastikan akurasi hasil klasifikasi, proses validasi instrumen dilakukan dengan melibatkan ahli bahasa dan ahli data science. Validasi ini bertujuan untuk menilai kualitas instrumen yang digunakan dalam penelitian, seperti proses pembersihan data (preprocessing), pelabelan sentimen, serta pengaplikasian algoritma K-Nearest Neighbor (KNN). Ahli bahasa memberikan masukan terkait ketepatan pemahaman konteks linguistik dalam tweet, sementara ahli data science memastikan bahwa pendekatan teknis, seperti pemilihan fitur dan algoritma yang digunakan, sudah sesuai dan efektif untuk analisis sentimen [12]. Validasi yang komprehensif ini penting untuk memastikan bahwa data yang digunakan relevan, akurat, dan sesuai dengan tujuan penelitian [9].

1. **Penyajian Data:**

Data yang diperoleh akan disajikan dengan jelas menggunakan tabel, grafik, dan visualisasi lainnya untuk menggambarkan proses analisis yang dilakukan. Tabel akan digunakan untuk menyajikan distribusi sentimen dalam setiap kategori (positif, negatif, dan netral), sedangkan grafik batang dan diagram lingkaran akan memvisualisasikan perbandingan frekuensi setiap sentimen. Penyajian ini bertujuan untuk memudahkan pembaca dalam memahami proses dan hasil penelitian secara lebih sistematis dan [11]. Selain itu, persamaan matematika yang menggambarkan perhitungan bobot fitur dalam metode TF-IDF dan evaluasi model menggunakan k-fold cross-validation akan disertakan untuk menunjukkan langkah-langkah teknis yang digunakan dalam analisis data [38].

Pentingnya penyajian data yang jelas dan terstruktur dapat membantu meningkatkan pemahaman tentang hasil analisis, terutama dalam penelitian yang melibatkan data besar dan kompleks [10].

# DAFTAR PUSTAKA

[1] A. Tanggu Mara, E. Sediyono, and H. Purnomo, “Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Metode Pembelajaran Dalam Jaringan (DARING) Di Universitas Kristen Wira Wacana Sumba,” *Jointer - Journal of Informatics Engineering*, vol. 2, no. 01, pp. 24–31, Jun. 2021, doi: 10.53682/jointer.v2i01.30.

[2] Husdi and Andi Kamaruddin, “Auto Labeling Untuk Analisis Sentimen Opini Mahasiswa Baru Terhadap Pembelajaran Mata Kuliah Dengan Algoritma K-Nearest Neighbor,” *Jurnal RESTIKOM : Riset Teknik Informatika dan Komputer*, vol. 6, no. 1, pp. 148–157, Apr. 2024, doi: 10.52005/restikom.v6i1.311.

[3] A. Halimi, K. Kusrini, and M. R. Arief, “ANALISIS SENTIMEN MASYARAKAT INDONESIA TERHADAP PEMBELAJARAN ONLINE DARI DI MEDIA SOSIAL TWITTER MENGGUNAKAN LEXICON DAN K-NEAREST NEIGHBOR,” *COREAI: Jurnal Kecerdasan Buatan, Komputasi dan Teknologi Informasi*, vol. 2, no. 1, pp. 18–28, Aug. 2021, doi: 10.33650/coreai.v2i1.2283.

[4] R. Gunawan, R. Septiadi, F. Apri Wenando, H. Mukhtar, and Syahril, “K-Nearest Neighbor (KNN) untuk Menganalisis Sentimen terhadap Kebijakan Merdeka Belajar Kampus Merdeka pada Komentar Twitter,” *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 2, pp. 152–158, Aug. 2022, doi: 10.37859/coscitech.v3i2.3841.

[5] S. Arikunto, *Prosedur Penelitian*. Jakarta: Rineka Cipta, 2019.

[6] S. Arikunto, *Prosedur Penelitian*. Rineka Cipta, 2019.

[7] M. Furqan, S. Sriani, and S. M. Sari, “Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia,” *Techno.Com*, vol. 21, no. 1, pp. 51–60, Feb. 2022, doi: 10.33633/tc.v21i1.5446.

[8] S. J. Cai, “Sentiment analysis using natural language processing and machine learning,” *Journal of Data Acquisition and Processing*, vol. 38, no. 2, 2023, doi: https://doi.org/10.5281/zenodo.7766376.

[9] C. C. Aggarwal, “Machine Learning for Text: An Introduction,” in *Machine Learning for Text*, Cham: Springer International Publishing, 2018, pp. 1–16. doi: 10.1007/978-3-319-73531-3\_1.

[10] J. Yadav, “Sentiment Analysis on Social Media,” Jan. 09, 2023. doi: 10.32388/YF9X04.

[11] J. Silge and D. Robinson, *Text mining with R*. O’Reilly Media, 2017.

[12] J. Supriyanto, D. Alita, and A. R. Isnain, “Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring,” *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 4, no. 1, pp. 74–80, Mar. 2023, doi: 10.33365/jatika.v4i1.2468.

[13] F. A. Juliasari and D. A. Sihombing, “Analisis Persepsi Kemudahan Penggunaan Dan Persepsi Manfaat Terhadap Minat Pembelian Konsumen Gofood Di Kalangan Mahasiswa Ibm Asmi,” NCBMA (Universitas Pelita Harapan Indonesia).

[14] S. Yusuf and Nurihsan, *Pengembangan Program Bimbingan dan Konseling di Sekolah*. Bandung: Remaja Rosdakarya. Remaja Rosdakarya, 2019.

[15] A. R. Safira, H. K. Sirajuddin, A. Khairan, and A. Mubarak, “Penerapan algoritma K-nearest neighbor dalam rekomendasi keminatan mahasiswa (Studi kasus: Program Studi Teknik Sipil Universitas Khairun),” *JATI (Jurnal Jaringan dan Teknologi Informasi)*, vol. 3, no. 2, pp. 6–12, 2024, doi: https://doi.org/10.0000/jati.

[16] R. F. Tanjung, N. Neviyarni, and F. Firman, “LAYANAN INFORMASI DALAM PENINGKATAN KETERAMPILAN BELAJAR MAHASISWA STKIP PGRI SUMATERA BARAT,” *Jurnal Penelitian Bimbingan dan Konseling*, vol. 3, no. 2, Sep. 2018, doi: 10.30870/jpbk.v3i2.3937.

[17] R. Mustofa, H. Irawadi, H. S. Lemana, and M. Ridwan, “Aktivitas Latihan Mahasiswa Prodi Pendidikan Kepelatihan Olahraga FIK UNP,” *Jurnal Patriot*, vol. 2, no. 3, pp. 743–756, 2020.

[18] R. Mustofa, H. Irawadi, H. S. Lemana, and M. Ridwan, “Aktivitas Latihan Mahasiswa Prodi Pendidikan Kepelatihan Olahraga FIK UNP,” *Jurnal Patriot*, vol. 2, no. 2, pp. 743–756, 2020.

[19] D. Y. Khairun and M. D. Nurmala, “PROGRAM BIMBINGAN DAN KONSELING MAHASISWA FKIP UNTIRTA BERBANTUAN SOFTWARE ANALISIS TUGAS PERKEMBANGAN,” *Jurnal Penelitian Bimbingan dan Konseling FKIP Untirta*, pp. 1–5, 2000.

[20] Nika Sintesa, “Analisis Pengaruh Time Management Terhadap Kedisiplinan dan Akademik Mahasiswa,” *Trending: Jurnal Manajemen dan Ekonomi*, vol. 1, no. 1, pp. 36–46, Dec. 2022, doi: 10.30640/trending.v1i1.465.

[21] T. Rinanda and C. C. Cen, “Study of the Role of Academics in Efforts to Welfare Indonesian Society,” *AFoS J-LAS*, vol. 2, no. 4, pp. 22–28, 2022.

[22] Syamsudin, W. Rachmawanto, and T. W. Astuti, “Dukungan Keluarga Terhadap Prestasi Belajar Mahasiswa Tingkat III Semester VI,” *Jurnal Keperawatan*, vol. 8, no. 2, pp. 27–35, 2022.

[23] C. Leonita and L. Tulistyantoro, “Perancangan Interior Coffee Shop dengan Fasilitas Belajar untuk Mahasiswa di Denpasar,” *INTRA*, vol. 6, no. 1, pp. 15–23, 2017.

[24] R. I. Borman, B. Priyopradono, and A. R. Syah, “Klasifikasi Objek Kode Tangan pada Pengenalan Isyarat Alphabet Bahasa Isyarat Indonesia (BISINDO),” *SNIA*, pp. 1–4, 2018.

[25] M. Marsono, A. H. Nasyuha, S. N. Arif, M. Zunaidi, and N. Y. L. Gaol, “Implementasi Algoritma K-Nearest Neighbor Dalam Mendiagnosis Kurap Pada Kucing,” *Journal of Computer System and Informatics (JoSYC)*, vol. 4, no. 1, pp. 61–65, Nov. 2022, doi: 10.47065/josyc.v4i1.2479.

[26] R. Samuel, R. Natan, and U. Syafiqoh, “Application of Cosine Similarity and K-Nearest Neighbor (K-NN) in Classification and Book Search,” *Journal of Big Data Analytic and Artificial Intelligence*, vol. 1, no. 1, pp. 9–14, 2018.

[27] A. M. Argina, “Penerapan Metode Klasifikasi K-Nearest Neigbor pada Dataset Penderita Penyakit Diabetes,” *Indonesian Journal of Data and Science*, vol. 1, no. 2, pp. 29–33, Jul. 2020, doi: 10.33096/ijodas.v1i2.11.

[28] M. Misdram and A. Cahyono, “Optimasi Komposisi Makanan Untuk Penderita Anemia Menggunakan Metode Variable Neighborhood Search,” *Spirit*, vol. 13, no. 1, pp. 28–34, 2021, doi: 10.53567/spirit.v13i1.201.

[29] D. S. Wisdayani, I. M. Nur, and R. Wasono, “Penerapan Algoritma K-Nearest Neighbor dalam Klasifikasi Tingkat Keparahan Korban Kecelakaan Lalu Lintas di Kabupaten Jawa Tengah,” in *Prosiding Mahasiswa Seminar Nasional Unimus*, 2019, pp. 373–380.

[30] A. Heriyanto, “Penerapan Metode K-Nearest Neighbor (KNN) Untuk Klasifikasi Stanting Pada Balita,” Publikasi Ilmiah Universitas Muhammadiyah Jember.

[31] H. Sulistiani, I. Darwanto, and I. Ahmad, “Penerapan Metode Case Based Reasoning,” *Jurnal Edukasi Dan Penelitian Informatika*, vol. 6, no. 1, pp. 23–38, 2020.

[32] Y. D. Alfiyanti, “KLASIFIKASI FUNGSI SENYAWA AKTIF DATA BERDASARKAN KODE SIMPLIFIED MOLECULAR INPUT LINE ENTRY SYSTEM ( SMILES ) MENGGUNAKAN METODE MODIFIED K - NEAREST,” 2018.

[33] P. Kelulusan, “PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN K-NEAREST NEIGHBOR BERBASIS PARTICLE SWARM OPTIMIZATION Nursetia Wati,” *Jtii*, vol. 6, no. 2, pp. 118–127, 2021.

[34] W. Nugroho, “Optimasi Metode K-Nearest Neighbours dengan Backward Elimination Menggunakan Dataset Software Effort Estimation Bianglala Informatika,” *Bianglala Informatika*, vol. 8, no. 2, pp. 129–133, 2020.

[35] J. Riany, M. Fajar, and M. P. Lukman, “Penerapan deep sentiment analysis pada angket penilaian terbuka menggunakan K-nearest neighbor,” *Jurnal SISFO: Inspirasi Profesional Sistem Informasi*, vol. 6, no. 1, pp. 147–156, 2016, doi: https://doi.org/10.24089/j.sisfo.2016.09.011.

[36] T. W. Putra, Triayudi. A, and Andrianingsih, “Analisis sentimen pembelajaran daring menggunakan metode Naïve Bayes, KNN, dan Decision Tree,” *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi*, vol. 6, no. 1, pp. 20–26, 2022, doi: https://doi.org/10.35870/jtik.v6i1.368.

[37] A. Deviyanto, “Penerapan analisis sentimen pada pengguna Twitter menggunakan metode K-Nearest Neighbor,” *Jurnal Ilmiah Sistem Komputer dan Aplikasi (JISKA)*, vol. 3, no. 1, 2018, doi: https://doi.org/10.14421/JISKA.2018.31-01.

[38] L. F. Narulita, “Analisa sentimen pada tinjauan buku dengan algoritma k-nearest neighbour,” *KONVERGENSI*, vol. 13, no. 2, pp. 76–81, 2017, doi: https://doi.org/10.30996/konv.v13i2.2758.

[39] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif dan R&D*, 4th ed. Alfabeta, 2022.

[40] Sugiyono, *Metode Penelitian Pendekatan Kuantitatif, Kualitatif, dan R&D*. Alfabeta, 2020.

[41] F. Z. Tala, “A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia,” 2003.