

**KLASIFIKASI EMOSI TEKS PADA MEDIA SOSIAL  
*PLATFORM X* MENGGUNAKAN METODE  
*K-NEAREST NEIGHBOR***

**SKRIPSI**

**NURAINI**

**171402148**



**PROGRAM STUDI S1 TEKNOLOGI INFORMASI  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS SUMATERA UTARA  
MEDAN  
2024**

KLASIFIKASI EMOSI TEKS PADA MEDIA SOSIAL  
*PLATFORM X* MENGGUNAKAN METODE  
*K-NEAREST NEIGHBOR*

SKRIPSI

Diajukan untuk melengkapi tugas dan memenuhi syarat memperoleh ijazah  
Sarjana Komputer Teknologi Informasi

NURAINI

171402148



PROGRAM STUDI S1 TEKNOLOGI INFORMASI  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS SUMATERA UTARA

MEDAN

2024

**PERSETUJUAN**

Judul : Klasifikasi Emosi Teks Pada Media Sosial Platform X  
Menggunakan Metode *K-Nearest Neighbor*

Kategori : Skripsi

Nama Mahasiswa : Nuraini

Nomor Induk Mahasiswa : 171402148

Program Studi : Sarjana (S-1) Teknologi Informasi

Fakultas : Ilmu Komputer dan Teknologi Informasi

Universitas Sumatera Utara

Medan, 12 Juli 2024

Komisi Pembimbing:

Pembimbing 2

Dedy Arisandi S.T., M.Kom  
NIP. 197908312009121002

Pembimbing 1

Rossy Murhasanah S.Kom., M.Kom  
NIP. 198707012019032016

Diketahui/disetujui oleh  
Program Studi S-1 Teknologi Informasi  
Ketua,

Dedy Arisandi S.T., M.Kom  
NIP. 197908312009121002

**PERNYATAAN**

**KLASIFIKASI EMOSI TEKS PADA MEDIA SOSIAL  
PLATFORM X MENGGUNAKAN METODE  
*K-NEAREST NEIGHBOR***

**SKRIPSI**

Saya mengakui bahwa skripsi ini adalah hasil karya saya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing telah disebutkan sumbernya.

Medan , 12 Juli 2024



**NURAINI**

171402148

## UCAPAN TERIMA KASIH

Puji Syukur penulis ucapkan kepada Allah Subhana Wata'ala yang Maha Pengasih lagi Maha Penyayang, karena atas rahmat dan kuasa-Nya, penulis dapat menyelesaikan penyusunan tugas akhir skripsi ini sebagai syarat untuk memperoleh gelar Sarjana Komputer, pada Program Studi S-1 Teknologi Informasi Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara.

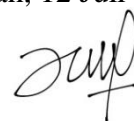
Penulisan skripsi ini dimaksudkan untuk memenuhi persyaratan akademis dalam memperoleh gelar sarjana (S1) pada Universitas Sumatera Utara. Pada kesempatan ini, penulis mengucapkan terima kasih yang setinggi-tingginya kepada ibunda tercinta **Faridah** dan ayahanda tercinta **Usna R** yang telah mengasuh, mendidik, membina, membimbing serta selalu mendo'akan penulis sehingga dapat menyelesaikan pendidikan sampai ke perguruan tinggi. Semoga jerih payah dan ketulusan orang tua kami mendapat balasan setimpal disisi Allah SWT. Ucapan terimakasih juga kepada kakak-kakak tercinta Fiddini, S.H, Nuraina, Amd.Kep, dan dr. Aulia yang telah memberi dukungan untuk doa, nasihat, kerja keras dan semangat kepada penulis sehingga penulis dapat menyelesaikan skripsi ini. Selanjutnya ucapan terimakasih penulis sampaikan pula kepada:

1. Bapak Dr. Muryanto Amin, S.Sos., M.Si. selaku Rektor Universitas Sumatera Utara.
2. Ibu Dr. Maya Silvi Lidya B.Sc., M.Sc., selaku Dekan Fasilkom-TI Universitas Sumatera Utara.
3. Dedy Arisandi ST., M.Kom selaku Ketua Program Studi S1 Teknologi Informasi Universitas Sumatera Utara dan selaku Dosen Pembimbing II yang telah membimbing penulis dengan sebaik-baiknya dan memberikan kritik dan saran yang sangat bermanfaat bagi penulis.
4. Ibu Rossy Nurhasanah, S.Kom., M.Kom selaku Dosen Pembimbing I yang telah membimbing penulis dengan sangat sabar, memberika arahan, masukkan serta motivasi sehingga penulis dapat menyelesaikan skripsi ini

5. Seluruh Dosen Fasilkom-TI Universitas Sumatera Utara atas segala ilmu yang diberikan kepada penulis serta staf akademik yang telah membantu penulis menjalankan perkuliahan dengan baik.
6. Sahabat seperjuangan penulis Majidah Atmayana Purba, S.Kom, Adelia Salmah Siregar, S.Kom, Eka Wulandari, S.Kom, Aida Suryana Ritonga, S.Kom, Melati Yulvira Salsabila, S.Kom, Aflah Mutsanni pulungan, S.Kom dan Sinta Sintya Rani, S.Kom yang telah menjadi teman dalam suka dan duka, serta tempat untuk bertukar pikiran semasa kuliah hingga penyelesaian skripsi ini.
7. Teman-teman seperjuangan Kom A 2017 dan seluruh keluarga besar stambuk 2017 Teknologi Informasi Universitas Sumatera Utara yang telah banyak memberi motivasi kepada penulis dalam pengerjaan skripsi in
8. Teman-teman seperjuangan pengurus UKMI Al Khuwarizmi terkhusus yang ikut berjuang di masa periode 2017 – 2019 (Laskar Alkhuwarizmi) dan merupakan tempat mengasah soft skills penulis dalam berorganisasi.
9. Terimakasih juga kepada kakak-kakak Pementor UKMI Al Khuwarizmi yang telah banyak memberikan dukungan dan motivasi kepada penulis selama masa perkuliah hingga penulis dapat menyelesaikan skripsi ini.
10. Untuk keponakan tercinta Rumaysa Almahira dan Aisyah Asyifa yang telah menemani dan menyemangati penulis hingga penulis dapat menyelesaikan skripsi ini.
11. Keluarga besar Alm. R a m l a h yang telah memberikan do'a, motivasi serta dukungan dalam penulisan skripsi ini.
12. Keluarga Asrama Putri USU yaitu Nuraprilia, S.Kom dan Wilda Rahmina S.kom yang telah mendukung dalam doa maupun semangat dalam penyelesaian skripsi ini.

Semoga Allah Subhanahu wa Ta'ala me-Rahmati mereka atas semua yang telah mereka lakukan dan berikan.

Medan, 12 Juli 2024



Penulis

**KLASIFIKASI EMOSI TEKS PADA MEDIA SOSIAL  
PLATFORM X MENGGUNAKAN METODE  
K-NEAREST NEIGHBOR**

**ABSTRAK**

Menurut Kamus Besar Bahasa Indonesia (KBBI), emosi didefinisikan sebagai luapan perasaan yang berkembang dan surut dalam waktu singkat, serta keadaan dan reaksi psikologis dan fisiologis seperti kegembiraan, kesedihan, keharuan, kecintaan, dan keberanian yang bersifat subjektif. Dalam mengekspresikan emosinya, manusia dapat mengungkapkan emosi mereka secara lisan atau melalui tulisan. Dengan kemajuan teknologi, dilakukan penelitian tentang interaksi emosi manusia-komputer, terutama emosi yang diambil dari teks. Analisis emosi dalam teks dapat diterapkan pada berbagai media, salah satunya adalah situs *microblogging* seperti *Platform X*. Dengan pertumbuhan data yang pesat, data perlu diklasifikasikan untuk menghasilkan informasi yang terstruktur dengan cepat, meskipun dalam jumlah besar, tanpa memerlukan waktu yang lama. Semua data yang akan dikategorikan dibagi menjadi lima kategori emosi: marah, takut, antusias, sedih, dan senang. Studi ini menggunakan metode *K-nearest neighbor* dan 2000 tweet. Pada metode K-Nearest Neighbor menggunakan pengujian *k-fold cross validation* menunjukkan akurasi sebesar 88%. Berdasarkan hasil penelitian menunjukkan bahwa metode *K-nearest neighbor* berhasil mengklasifikasikan emosi di *platform X*.

Kata Kunci : Emosi, *Platform X*, Klasifikasi, *K-Nearest Neighbor*, *K-Fold Cross Validation*

**EMOTION CLASSIFICATION OF TEXTS ON SOCIAL  
MEDIA PLATFORM X USING K-NEAREST  
NEIGHBOR METHOD**

**ABSTRACT**

*According to the Kamus Besar Bahasa Indonesia (KBBI), emotion is defined as a surge of feelings that fluctuates rapidly, as well as psychological and physiological states and reactions such as joy, sadness, melancholy, affection, and courage which are subjective in nature. In expressing their emotions, humans can articulate them verbally or through writing. With technological advancements, research has been conducted on human-computer emotion interaction, particularly emotions extracted from text. Emotion analysis in text can be applied across various media, one of which is microblogging sites like Platform X. With the rapid growth of data, there is a need to classify data to quickly generate structured information, even in large volumes, without consuming significant time. All data to be categorized are divided into five emotion categories: Anger, Fear, Enthusiasm, Sadness, and Happiness. This study utilizes the K-Nearest Neighbor method and 2000 tweets. The K- nearest neighbor the results of k-fold cross validation testing show an accuracy of 88%. based on the research findings, the K-nearest neighbor method successfully classifies emotions on Twitter.*

*Keywords: Emotion, Platform X, Classification, K-Nearest Neighbor, K-Fold Cross Validation*



## DAFTAR ISI

PERSETUJUAN	II
PERNYATAAN	II
UCAPAN TERIMA KASIH	IV
ABSTRAK	VI
<i>ABSTRACT</i>	VII
DAFTAR ISI	VIII
DAFTAR TABEL	XI
DAFTAR GAMBAR	XII
 BAB 1 PENDAHULUAN	 1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian	6
1.5 Manfaat Penelitian	6
1.6 Metodologi Penelitian	6
1.7 Sistematika Penulisan	8
 BAB 2 LANDASAN TEORI	 9
2.1. Data Mining	9
2.2. Klasifikasi	10
2.3. Preprocessing	10
2.4. <i>K-Nearest Neighbor (K-NN)</i>	11
2.5. Emosi	13
2.6. Media Sosial	15
2.7. <i>Platform X</i>	16
2.8. <i>K-Fold Cross Validation</i>	17
2.9. <i>Hypertext Markup Language (HTML)</i>	19
2.10. JavaScript	19

2.11.	<i>Hyper Preprocessor</i> (PHP)	20
2.12.	CodeIgniter	20
2.13.	<i>TensorFlow</i>	20
2.14.	Penelitian Terdahulu	21
BAB 3 ANALISIS DAN PERANCANGAN SISTEM		27
3.1	Data yang digunakan	27
3.2	Analisis Sistem	28
1.	Arsitektur Umum	28
2.	Input Dataset	30
3.	<i>Pre-processing</i>	30
6.	<i>Tokenizing</i>	31
7.	<i>Filtering</i>	31
8.	<i>Stemming</i>	32
9.	Ekstraksi fitur <i>Word Vector Mapping (Word2Vec)</i>	32
10.	Klasifikasi	32
11.	Model <i>training</i>	33
12.	<i>Output</i>	33
3.3	Perancangan Antarmuka ( <i>User Interface</i> )	33
1.	Perancangan Halaman Awal	33
2.	Perancangan tampilan input data latih	34
3.	Perancangan Halaman Hasil	34
3.4	Perancangan Tabel Hasil Pengujian	35
1.	Perancangan Tabel Hasil Pengujian	35
2.	Perancangan Tabel Hasil Pengujian Algoritma	35
3.	Perhitungan Klasifikasi Tweet Manual dengan Algoritma KNN	35
BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM		37
4.1	Implentasi Sistem	37
1.	Hasil Pengujian Sistem	37
2.	Perancangan Tabel Hasil Pengujian Algoritma	42

BAB 5 KESIMPULAN DAN SARAN	44
5.1 Kesimpulan	44
5.2 Saran	44
DAFTAR PUSTAKA	45

## DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	22
Tabel 3. 1 Contoh data tweet emosi	27
Tabel 3. 2 Variabel Penelitian	28
Tabel 3. 3 Tampilan tweet sebelum (kiri) dan sesudah (kanan) melalui proses <i>Cleaning</i>	30
Tabel 3. 4 Tampilan tweet sebelum (kiri) dan sesudah (kanan) melalui proses <i>Case Folding</i>	31
Tabel 3. 5 Tampilan tweet sebelum (kiri) dan sesudah (kanan) melalui proses <i>Tokenizing</i>	31
Tabel 3. 6 Tampilan tweet sebelum (kiri) dan sesudah (kanan) melalui proses <i>Filtering</i>	32
Tabel 3. 7 Tampilan tweet sebelum (kiri) dan sesudah (kanan) melalui proses <i>Stemming</i>	32
Tabel 4. 1 Hasil Klasifikasi emosi	38
Tabel 4. 2 Tabel <i>Confussion Matrix</i>	41
Tabel 4. 3 Tabel Nilai T, FP, FN	41
Tabel 4. 4 Tabel Nilai TP, FP, dan FN	42
Tabel 4. 5 Hasil 10 <i>Fold Cross Validation</i>	43

## DAFTAR GAMBAR

Gambar 2. 1 Flowchart Algoritma KNN	13
Gambar 2. 2 Pembagian data dengan <i>K-Fold Cross validation</i>	17
Gambar 2. 3 Ilustrasi <i>K-Fold Cross Validation</i> Dengan $K = 10$	18
Gambar 3. 1 Arsitektur Umum Sistem	29
Gambar 3. 2 Proses Cleaning	30
Gambar 3. 3 Perancangan Halaman Awal	33
Gambar 3. 4 Tampilan input data testing	34
Gambar 3. 5 Perancangan Halaman Hasil	35
Gambar 3. 6 Tweet Mula-mula	36
Gambar 4. 1 Implementasi Halaman Utama	37
Gambar 4. 2 Halaman Input data testing	38
Gambar 4. 3 Hasil Klasifikasi	38

## **BAB 1**

### **PENDAHULUAN**

#### **1.1 Latar Belakang**

Menurut Kamus Besar Bahasa Indonesia (KBBI), emosi didefinisikan sebagai luapan perasaan yang berkembang dan surut dalam waktu singkat, serta keadaan dan reaksi psikologis dan fisiologis seperti kegembiraan, kesedihan, keharuan, kecintaan dan keberanian yang bersifat subjektif. Dengan demikian, dapat disimpulkan bahwa emosi adalah reaksi kompleks dari individu terhadap rangsangan internal atau eksternal yang menyebabkan perubahan dalam kehidupan mereka. Definisi lain dari emosi yaitu emosi adalah bahwa emosi sangat penting untuk kecerdasan manusia, interaksi sosial, pengambilan keputusan rasional, persepsi, memori, pembelajaran, kreativitas, dan banyak lagi (Balabantaray et al., 2012).

Salah satu media sosial yang populer di kalangan masyarakat Indonesia saat ini adalah platform X, yang sebelumnya dikenal sebagai Twitter. Penelitian sebelumnya menunjukkan bahwa banyak orang lebih sering membuka media sosial X dan TikTok setiap hari. Mereka bahkan menyempatkan diri untuk membuka aplikasi X meskipun sedang sibuk dengan tugas atau aktivitas lain (Hastuti et al., 2023). Twitter memiliki banyak manfaat, salah satunya adalah sebagai tempat curhat atau mengungkapkan perasaan. Setiap *posting* di Twitter disebut "tweet", yang dapat menggambarkan perasaan pengguna Twitter (Asriadie et al., 2018).

Penelitian mengenai analisis emosi dalam teks, khususnya pada media sosial seperti twittet atau X, memiliki dampak besar karena jumlah pesan yang dihasilkan secara harian sangat besar dan mencerminkan beragam ekspresi emosional manusia. Analisa emosi dalam teks dapat dilakukan pada berbagai media, Salah satu situs I-2 *microblogging* Twitter yang memiliki pengguna lebih dari 29,5 juta orang di Indonesia dan mengirimkan 383 juta tweet setiap hari,

adalah salah satu situs *microblogging* yang menghasilkan jutaan pesan setiap hari (Rizqiyani et al, 2017).

Deteksi emosi dilakukan dengan menggunakan pendekatan klasifikasi emosi, di mana klasifikasi adalah proses mengelompokkan data ke dalam kategori-kategori atau kelas-kelas tertentu berdasarkan kesamaan pada suatu kriteria. Dalam klasifikasi emosi, data diberikan label kelas emosi sesuai dengan yang ditentukan. Salah satu metode klasifikasi yang umum digunakan adalah Algoritma *K-Nearest Neighbor (KNN)*. KNN merupakan algoritma sederhana yang mengklasifikasikan data berdasarkan jarak terdekatnya dengan data yang telah ada dalam set pelatihan. Algoritma ini memanfaatkan konsep bahwa data-data yang serupa cenderung berada dalam jarak yang dekat di dalam ruang fitur.

Klasifikasi emosi teks pada media sosial, seperti Platform X, memiliki peran penting dalam dunia kesehatan. Penelitian ini membantu tenaga kesehatan menemukan gejala kesehatan mental seperti depresi, kecemasan, atau stres. Data dari pengklasifikasian emosi ini dapat dinilai untuk mengevaluasi seberapa efektif kampanye kesehatan mental. Untuk meningkatkan dampak dan keberhasilan kampanye, strategi komunikasi dapat disesuaikan dengan data ini.

KNN efektif dalam mengklasifikasikan data, terutama pada kasus-kasus di mana batasan antara kelas-kelas tidak terlalu jelas, dan data dapat saling tumpang tindih di dalam ruang fitur. Keunggulan utama KNN adalah kemampuannya dalam menangani data training yang besar dan ketahanannya terhadap noise dalam data. Namun, KNN juga memiliki kelemahan, seperti sensitif terhadap skala data dan membutuhkan ruang memori yang cukup besar untuk menyimpan seluruh data training.

Dengan demikian, penelitian yang menggunakan metode KNN untuk klasifikasi emosi bertujuan untuk mengelompokkan data tweet ke dalam kategori-kategori emosi yang telah ditentukan berdasarkan kemiripan dengan data pelatihan yang ada, dengan harapan dapat menghasilkan model klasifikasi yang dapat mengenali dan memahami emosi yang terkandung dalam teks tweet dengan akurat. KNN merupakan algoritma pengklasifikasi data sederhana dimana perhitungan jarak terpendek dijadikan ukuran untuk mengklasifikasi suatu baru berdasarkan ukuran kemiripan (Pandie, 2012). *K-Nearest Neighbor* merupakan

algoritma dikenal efektifitasnya terhadap ketahanan pada data noise (Imron & Kusumah, 2018; Salim et al., 2020).

Klasifikasi *K-Nearest Neighbor* merupakan objek metode diklasifikasikan berdasarkan data pembelajaran dengan cara menghitung kedekatan jarak satu data terhadap data yang lainnya, yang mana KNN ini mempunyai beberapa keunggulan diantaranya memiliki kemampuan untuk melakukan pelatihan dengan sangat cepat, termasuk algoritma yang sederhana dan mudah untuk dipelajari, tahan pada data yang memiliki derau, dan efektif pada data pelatihan yang besar. Namun algoritma KNN ini juga mempunyai kelemahan diantaranya jumlah nilai  $k$  dapat menjadi bias, komputasinya kompleks, memori terbatas, serta mudah dikelabui dengan fitur yang tidak relevan (Supriatman, 2021).

Dalam penelitian yang dilakukan oleh H. M. Kadhim (2020), KNN digunakan untuk klasifikasi teks dan dibandingkan dengan algoritma lain seperti *Support Vector Machine* (SVM) dan *Decision Tree*. Penelitian ini menunjukkan bahwa KNN unggul dalam beberapa aspek ketika diterapkan pada berbagai dataset teks. Secara khusus, KNN mencapai akurasi hingga 85% pada dataset teks tertentu, sementara SVM dan *Decision Tree* masing-masing mencapai akurasi sekitar 82% dan 80%.

Penelitian ini mengindikasikan bahwa KNN lebih efektif pada dataset yang memiliki distribusi data yang seimbang dan ukuran yang tidak terlalu besar. KNN menunjukkan keunggulan dalam presisi dan *recall*, yang membuatnya sangat berguna dalam aplikasi yang memerlukan deteksi yang akurat dari kategori teks. Namun, penelitian ini juga mencatat bahwa pada dataset yang lebih besar atau lebih kompleks, SVM dan *Decision Tree* mungkin menunjukkan performa yang lebih baik. Temuan ini menyoroti pentingnya memilih algoritma yang tepat berdasarkan karakteristik spesifik dari dataset dan tujuan klasifikasi.

Dalam penelitian yang dilakukan oleh Rajput dan Desai (2021), melakukan perbandingan terhadap beberapa algoritma pengklasifikasi teks pada artikel berita. Hasil penelitian menunjukkan bahwa KNN memiliki performa yang baik dalam pengklasifikasian artikel berita, terutama ketika ukuran dataset tidak terlalu besar. Penelitian ini melibatkan beberapa metrik evaluasi seperti akurasi,



presisi, *recall*, dan *F1-score* untuk membandingkan kinerja KNN dengan algoritma lain seperti *Support Vector Machine* (SVM) dan *Decision Tree*.

Pada dataset dengan ukuran menengah, KNN mencapai akurasi sekitar 88%, yang lebih tinggi dibandingkan dengan SVM dan *Decision Tree* yang masing-masing mencapai akurasi sekitar 85% dan 83%. KNN menunjukkan keunggulan dalam menangani dataset yang tidak terlalu besar karena kemampuannya untuk dengan cepat menemukan tetangga terdekat dan menghasilkan prediksi yang akurat tanpa memerlukan banyak komputasi. Selain itu, KNN juga menunjukkan keunggulan dalam presisi dan *recall*, yang membuatnya sangat efektif dalam mendeteksi kategori berita yang benar dengan tingkat kesalahan yang rendah. Penelitian ini menyoroti bahwa meskipun KNN mungkin tidak selalu menjadi pilihan terbaik untuk dataset yang sangat besar atau sangat kompleks, algoritma ini sangat kompetitif dan sering kali unggul dalam tugas klasifikasi teks dengan ukuran dataset yang lebih terbatas.

Berdasarkan uraian tersebut, penelitian ini akan menggunakan metode *K-Nearest Neighbor* (KNN) sebagai alat klasifikasi untuk mendeteksi emosi dalam postingan tweet di platform media sosial Twitter. Metode pengambilan data akan melibatkan pengumpulan postingan tweet yang tersedia di Twitter, dan selanjutnya data akan dikelompokkan ke dalam lima kelas emosi yang telah ditentukan, yaitu marah, antusias, takut, sedih dan senang.

Dengan menggunakan metode KNN, penelitian ini akan mengimplementasi kemampuan algoritma tersebut dalam mengklasifikasikan tweet ke dalam kelompok emosi yang sesuai. Hal ini akan membantu memahami bagaimana emosi diekspresikan dalam konteks platform media sosial, serta memberikan wawasan tentang pola-pola emosi yang mungkin terjadi dalam postingan tweet.

Berdasarkan penelitian terdahulu yang telah dibahas di atas, maka penulis mengangkat penelitian dengan judul **“Klasifikasi Emosi Teks Pada Media Sosial Platform X Menggunakan Metode K-Neares Neighbor”**

## 1.2 Rumusan Masalah

Banyak orang seringkali menggunakan platform X untuk mengekspresikan emosi mereka, mulai dari kegembiraan, kekhawatiran, kekecewaan, hingga berbagai emosi lainnya. Dalam konteks klasifikasi emosi pada data tweet, setiap tweet akan diberikan label kelas emosi sesuai dengan yang ditentukan, seperti "senang", "sedih", "marah", dan sebagainya, berdasarkan konten dan konteksnya. Salah satu metode klasifikasi yang umum digunakan adalah Algoritma *K-Nearest Neighbor*. *K-Nearest Neighbor* dapat digunakan untuk menghitung jarak antara vektor representasi tweet baru dengan vektor representasi tweet yang telah ada dalam set pelatihan, dan kemudian mengklasifikasikannya ke dalam kelas emosi yang sesuai berdasarkan tetangga terdekatnya. Dengan memanfaatkan metode klasifikasi seperti *K-Nearest Neighbor*, penelitian dapat menghasilkan model yang dapat mengklasifikasikan emosi dalam tweet dengan tingkat akurasi yang memadai, yang kemudian dapat digunakan untuk berbagai tujuan, termasuk analisis sentimen, pemantauan opini publik, dan lainnya.. Oleh karena itu penelitian ini bertujuan untuk mengklasifikasikan emosi pada tweet twitter atau X menggunakan Algoritma *K-Nearest Neighbor (KNN)*.

## 1.3 Batasan Masalah

Berikut adalah batasan-batasan dari penelitian ini:

1. Penelitian ini membatasi diri untuk menggunakan *metode K-Nearest Neighbor (KNN)* dalam proses klasifikasi data.
2. Algoritma yang digunakan untuk proses evaluasi adalah *K-Fold Cross Validation*.
3. Data yang digunakan berasal dari dataset yang tersedia di platform Kaggle.
4. Data yang dianalisis dalam penelitian ini terdiri dari tweet-tweet berbahasa Indonesia
5. Klasifikasi yang dilakukan dalam penelitian ini terfokus pada lima kelas emosi, yaitu:
  - a. Marah
  - b. Takut
  - c. Antusias
  - d. Sedih

e. Senang

Dengan batasan-batasan ini, penelitian akan difokuskan pada penggunaan metode KNN untuk menganalisis data tweet berbahasa Indonesia dari kaggle dataset dengan tujuan mengklasifikasikan emosi pengguna platform X ke dalam lima kelas emosi yang telah disebutkan.

#### 1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk menimplementasi kemampuan KNN dalam mengenali dan mengklasifikasikan emosi yang terkandung dalam teks tweet dari platform X. Dengan demikian, penelitian ini akan mencoba menjawab pertanyaan tentang efektivitas dan kecocokan metode KNN dalam klasifikasi emosi pada teks tweet dalam konteks media sosial tersebut.

#### 1.5 Manfaat Penelitian

Klasifikasi Emosi Teks Pada Media Sosial Platform X Menggunakan Metode *K-Neares Neighbor* memiliki manfaat:

1. Penelitian ini akan memberikan wawasan mendalam tentang bagaimana emosi diekspresikan dalam teks tweet di platform X. Ini membantu memahami dinamika emosi dalam konteks media sosial dan perilaku pengguna.
2. Penggunaan metode *K-Nearest Neighbor (KNN)* untuk klasifikasi emosi pada tweet memungkinkan pengembangan algoritma yang efektif dan efisien.

#### 1.6 Metodologi Penelitian

Penelitian ini menggunakan metode kuantitatif yang terdiri dari dua tahap utama: metode pengumpulan data dan metode analisis data. Berikut adalah rinciannya:

1. Metode Pengumpulan Data
  - a. Studi Literatur: Melibatkan pencarian teori-teori yang relevan dari kumpulan jurnal dan buku yang berkaitan dengan penelitian untuk mendapatkan landasan teori yang sesuai dengan masalah penelitian.

- b. Crawling Data: Pengumpulan data dilakukan melalui media sosial, khususnya Platform X.

## 2. Metode Analisis Data

Metodologi analisis data dalam penelitian ini mengacu pada proses SEMMA (*Sample, Explore, Modify, Model, Assess*) dalam data mining, sebagaimana dijelaskan oleh (Alizah et al. 2020). Berikut adalah ringkasan dari setiap langkah dalam proses tersebut:

### 1. *Sample* (Pengambilan Sampel)

- a. Pengumpulan data dilakukan dengan mencari teori-teori terkait dari berbagai sumber seperti jurnal, buku, dan situs terkait.
- b. Data juga dikumpulkan langsung dari Platform X melalui crawling data.

### 2. *Explore* (Eksplorasi):

- a. Pada tahap ini, dilakukan deskripsi dan visualisasi data.
- b. Deskripsi data meliputi penjelasan tentang informasi yang digunakan, dengan fokus pada klasifikasi emosi pada tweet Platform X dalam bahasa Indonesia.

### 3. *Modify* (Modifikasi):

Data tweet yang belum diolah melewati beberapa langkah modifikasi, termasuk pembersihan data (*cleaning*), pembagian kata (*tokenize*), penghapusan kata umum (*stopword removal*), dan penyaringan (*filtering*).

### 4. *Model* (Pemodelan):

- a. Klasifikasi data dilakukan berdasarkan kelasnya untuk menentukan apakah termasuk opini positif atau negatif.
- b. Model algoritma yang sesuai digunakan untuk melakukan klasifikasi data.

### 5. *Assess* (Penilaian) :

- a. Evaluasi dilakukan terhadap pemodelan yang ada dengan membandingkan hasil prediksi terhadap data uji dengan label sentimen yang sudah didapatkan sebelumnya.

- b. Hasil evaluasi berupa nilai confusion matrix yang mencakup akurasi, presisi, dan recall.

Metode ini menyediakan kerangka kerja yang sistematis untuk mengumpulkan, menjelajahi, memodifikasi, memodelkan, dan mengevaluasi data dalam konteks penelitian klasifikasi emosi pada tweet Platform X dalam bahasa Indonesia. Langkah-langkah ini membantu peneliti dalam mengidentifikasi pola dan hubungan dalam data serta membuat prediksi berdasarkan model yang dikembangkan.

## **1.7 Sistematika Penulisan**

Sedangkan sistematika penulisan terdapat beberapa tahapan seperti dibawah ini:

### **Bab 1: Pendahuluan**

Bab ini memaparkan latar belakang, tujuan, batasan-batasan penelitian, manfaat serta sistematika penulisan tugas akhir.

### **Bab 2: Landasan Teori**

Bab kedua ini berisikan teori yang menjadi landasan penulis dalam analisis, perancangan, pembangunan serta pengujian sistem.

### **Bab 3: Analisis dan Perancangan Sistem**

Bab ini berisi tentang analisis dari arsitektur umum serta analisis dari penerapan *Algoritma* yang digunakan, pada bab ini juga terdapat perancangan tampilan antarmuka sistem.

### **Bab 4: Implementasi dan Pengujian Sistem**

Bab keempat menjelaskan implementasi dari metode, analisis dan desain yang telah dilakukan, dan verifikasi bahwa hasil yang telah diperoleh telah memenuhi harapan.

### **Bab 5: Kesimpulan Dan Saran**

Pada bab ini tercantum hasil dari keseluruhan tahapan dan kesimpulan apakah tujuan penelitian terpenuhi beserta saran demi pengembangan peneliti

## BAB 2

### LANDASAN TEORI

#### 2.1. Data Mining

Ada banyak pengertian mengenai data mining, namun menurut Kantardzic (2019) *Data mining* adalah proses berulang dimana kemajuan ditentukan melalui penemuan melalui metode otomatis atau manual. Data mining atau penambangan data yaitu suatu langkah dalam *knowledge discovery in databases (KDD)*. *knowledge discovery* sebagai sebuah proses yang terdiri dari pembersihan data, integrasi data, pemilihan data, transformasi data, data mining, evaluasi pola (*pattern evaluation*) dan knowledge presentation. Data mining adalah proses penggalian informasi dari kumpulan data yang sangat besar (Firdaus, 2017).

Tujuan dari adanya data mining adalah:

1. *Explanatory*, yaitu untuk menjelaskan beberapa kegiatan observasi atau suatu kondisi.
2. *Confirmatory*, yaitu untuk mengkonfirmasi suatu hipotesis yang telah ada.
3. *Exploratory*, yaitu untuk menganalisa data baru suatu relasi yang janggal.

*Data mining* adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam *database*, *data warehouse*, atau penyimpanan informasi lainnya. *Data mining* berkaitan dengan bidang ilmu-ilmu lain, seperti *database sistem*, *data warehousing*, *statistik*, *machine learning*, *information retrieval*, dan komputasi tingkat tinggi. Selain itu, *data mining* didukung oleh ilmu lain seperti *neural network*, pengenalan pola, *spatial data analysis*, *image database*, *signal processing*. *Data mining* didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya *semiotomatis*. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. Data yang dibutuhkan dalam jumlah besar.

Karakteristik data mining menurut Abdillah et al. (2016). adalah sebagai berikut:

1. Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
2. Data mining biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
3. Data mining berguna untuk membuat keputusan yang kritis, terutama dalam strategi.

## **2.2. Klasifikasi**

Pada data *mining* terdapat beberapa metode salah satunya yaitu klasifikasi, setelah data diolah menggunakan tahapan-tahapan preprocessing selanjutnya akan dilakukan tahap klasifikasi. Klasifikasi yaitu sebuah bentuk analisis data yang mengekstraksi pola pola yang mendeskripsikan kelas pada data, algoritma klasifikasi memprediksi label kelas kategorikal terhadap suatu data untuk mengklasifikasikannya ke dalam salah satu kelas yang telah teridentifikasi sebelumnya (Nugroho & Religia, 2021).

Klasifikasi terbagi menjadi 2 proses, pertama yaitu proses pembelajaran (*training*) yang mana sebuah algoritma klasifikasi dibangun untuk menganalisis data latih kemudian diungkapkan kedalam bentuk aturan klasifikasi, kedua yaitu proses klasifikasi yang mana data uji akan digunakan untuk mengevaluasi tingkat akurasi dari aturan klasifikasi tersebut (Maulana et al., 2021).

## **2.3. Preprocessing**

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

1. *Cleaning* merupakan tahapan untuk menghapus karakter yang tidak berpengaruh terhadap pengolahan data seperti tanda baca, angka, link, dan lain-lain.
2. *Case folding* merupakan tahapan mengubah semua huruf menjadi huruf besar atau kecil.
3. *Tokenizing* merupakan tahapan untuk memisahkan teks yang berupa kalimat, paragraph maupun dokumen menjadi potongan kata yang disebut token, tujuannya adalah untuk memperoleh potongan kata yang dijadikan entitas dan mempunyai nilai dalam matriks dokumen teks yang akan dianalisis.
4. *Stopword* merupakan tahapan untuk menghilangkan kosakata atau simbol yang tidak memiliki informasi yang berguna seperti kata penghubung dan kata keterangan.
5. *Stemming* merupakan tahapan untuk mendapatkan kata dasar dari sebuah token yaitu dengan menghapus kata awalan, akhiran, sisipan, dan *confixes* (kombinasi kata awalan dan akhiran) (Hadi & Sukanto, 2020).

#### **2.4. *K-Nearest Neighbor (K-NN)***

Menurut Limbong et al., (2022) Algoritma *K-Nearest Neighbor (KNN)* adalah salah satu metode yang umum digunakan dalam proses klasifikasi data. KNN bekerja dengan prinsip sederhana yaitu mencari kumpulan data pembelajaran yang memiliki jarak terdekat dengan objek yang akan diklasifikasikan, dan kemudian menentukan kelas atau label objek tersebut berdasarkan mayoritas kelas dari tetangga terdekatnya. Berikut adalah penjelasan singkat tentang cara kerja algoritma KNN:

1. Penentuan Tetangga Terdekat (*Nearest Neighbors*):
  - a. Ketika suatu objek yang akan diklasifikasikan diberikan, algoritma KNN akan menghitung jarak antara objek tersebut dengan semua objek dalam data pembelajaran.
  - b. Jarak ini bisa diukur dengan berbagai metrik, seperti jarak Euclidean atau jarak Manhattan, tergantung pada kebutuhan masalah.



2. Pemilihan Nilai K:

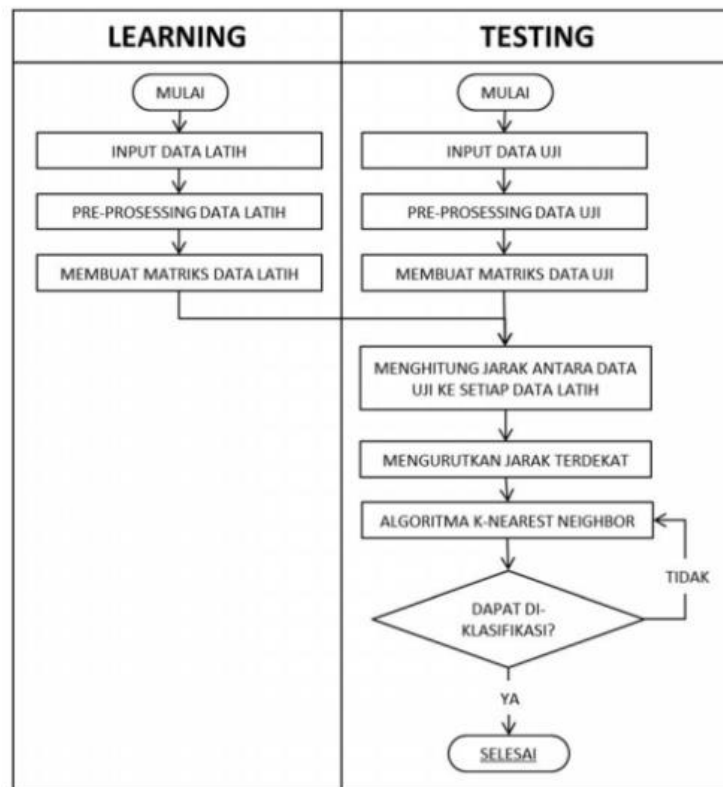
- a. KNN menggunakan parameter K yang menentukan jumlah tetangga terdekat yang akan dipertimbangkan.
- b. Nilai K harus ditentukan sebelumnya dan bisa dipilih melalui validasi silang atau berdasarkan karakteristik masalah.

3. Klasifikasi Objek:

- a. Setelah menentukan tetangga terdekat, KNN akan memilih mayoritas kelas dari tetangga tersebut.
- b. Objek yang akan diklasifikasikan akan diberikan label atau kelas yang paling umum di antara tetangga terdekatnya.

Algoritma KNN relatif mudah dipahami dan diimplementasikan, dan sering kali memberikan hasil yang baik untuk masalah klasifikasi dengan dataset yang cukup besar. Namun, KNN juga memiliki beberapa kelemahan, seperti sensitif terhadap perubahan skala dan rentan terhadap data yang tidak seimbang.

Dalam proses klasifikasi menggunakan Algoritma KNN, terdapat dua tahap utama yang harus dilalui, yaitu tahap pembelajaran (*learning*) dan tahap pengujian (*testing*). Pada tahap pembelajaran, sebagian data yang kelasnya sudah diketahui (data pelatihan) digunakan untuk membangun model prediksi. Selanjutnya, pada tahap pengujian, model prediksi yang telah dibangun tersebut diuji menggunakan sebagian data lainnya (data pengujian) untuk menentukan tingkat akurasi model. Jika akurasi model memadai, model tersebut dapat digunakan untuk memprediksi kelas data yang belum diketahui. Proses klasifikasi dengan K-NN mencakup tahapan-tahapan ini seperti yang terlihat pada gambar.



Gambar 2. 1 Flowchart Algoritma KNN

(Sumber : [Advanced Data Mining and Applications: 6th International Conference, ADMA 2010, Proceedings, Part I](#))

Algoritma K-NN terdiri dari langkah-langkah berikut:

1. Tentukan nilai k.
2. Hitung jarak antara data uji dan setiap data latih yang tersedia.
3. Urutkan jarak dari yang terkecil hingga terbesar.
4. Pilih k objek dari data latih yang memiliki jarak terkecil dengan data uji.
5. Klasifikasikan data uji ke dalam kelas yang ada dengan memilih kelas yang paling dominan di antara data latih.
6. Jika kelas dari data uji memiliki jumlah yang sama, ulangi langkah 4 dengan memilih 2 kali k data latih.

## 2.5. Emosi

Definisi emosi dalam Kamus Besar Bahasa Indonesia menekankan pada aspek *psikologis* dan *fisiologis* manusia sebagai tanggapan terhadap suatu kejadian

atau *stimulus*. Emosi memainkan peran penting dalam kehidupan manusia karena mempengaruhi perilaku, keputusan, dan interaksi antar manusia. Berikut adalah ringkasan definisi dari Kamus Besar Bahasa Indonesia:

1. Emosi adalah keadaan dan reaksi psikologis dan fisiologis, seperti kegembiraan, kesedihan, keharuan, dan kecintaan.
2. Emosi memiliki peran penting dalam interaksi antar manusia, membantu dalam mengekspresikan perasaan, dan mengarahkan keputusan yang cerdas dan rasional.

Definisi ini menyoroti kompleksitas emosi sebagai hasil dari interaksi antara faktor psikologis dan fisiologis. Emosi tidak hanya terjadi dalam diri individu, tetapi juga tercermin dalam perilaku dan interaksi sosial mereka. Definisi emosi menurut Destuardi & Sumpeno (2009) menggambarkan emosi sebagai suatu reaksi atau respon terhadap suatu kejadian yang dialami oleh seseorang. Respon tersebut meliputi beberapa aspek, seperti:

1. Mendeteksi adanya hal mengganggu pada tubuh, yang bisa mencakup perubahan fisiologis seperti detak jantung yang meningkat, pernapasan yang terengah-engah, atau perasaan tegang di tubuh.
2. Kesiapan dalam melakukan tindakan, menunjukkan bahwa emosi dapat mempengaruhi tingkat kewaspadaan dan kesiapan seseorang untuk bertindak dalam situasi tertentu.
3. Kemampuan untuk mengingat objek yang dilihat, menunjukkan bahwa emosi dapat memengaruhi proses kognitif seperti memori dan perhatian.
4. Ekspresi yang ditunjukkan melalui suara, mimik wajah, dan juga isyarat tubuh, menekankan bahwa emosi seringkali diekspresikan melalui berbagai cara komunikasi non-verbal.

Definisi ini menggambarkan kompleksitas emosi sebagai fenomena yang melibatkan interaksi antara komponen psikologis, fisiologis, dan perilaku. Menurut Goleman (1995), emosi dapat dikategorikan sebagai berikut:

1. Amarah yaitu beringas, mengamuk, benci, marah besar, jengkel, kesal hati, terganggu, tersinggung, bermusuhan, hingga tindakan kekerasan dan kebencian patologis.

2. Kesedihan yaitu pedih, muram, gugup, melankolis, mengasihi diri, kesedihan, ditolak, dan depresi berat.
3. Rasa takut yaitu gugup, khawatir, was-was, perasaan takut sekali, waspada, tidak senang, ngeri, takut sekali, fobia, dan panik.
4. Kenikmatan yaitu bahagia, gembira, puas, terhibur, bangga, takjub, terpesona, senang sekali dan manis.
5. Cinta yaitu persahabatan, penerimaan, kepercayaan, kebaikan hati, rasa dekat, bakti, hormat.
6. Terkejut yaitu terpana dan takjub.
7. Jengkel yaitu hina, jijik, muak, benci.
8. Malu yaitu rasa bermasalah, malu hati, kesal hati, hina, aib, dan hati hancur lebur.

Dari penjelasan diatas dapat disimpulkan bahwa bentuk-bentuk emosi adalah amarah, kesedihan, rasa takut, kenikmatan, cinta, terkejut, jengkel, dan malu.

## **2.6. Media Sosial**

Media sosial adalah platform berbasis online yang memfasilitasi interaksi, kolaborasi, dan berbagi konten antara pengguna. Platform ini memungkinkan pengguna untuk terhubung dengan orang lain, baik secara individu maupun dalam kelompok, melalui berbagai fitur seperti pesan teks, gambar, video, dan suara.

Berbagai jenis media sosial termasuk jejaring sosial seperti Facebook, Twitter, Instagram, dan LinkedIn, platform berbagi video seperti YouTube dan TikTok, serta platform berbagi gambar seperti Pinterest dan Snapchat. Setiap platform memiliki fokus dan fitur yang berbeda, tetapi semuanya memiliki kesamaan dalam menyediakan ruang untuk interaksi sosial dan berbagi konten.

Media sosial telah menjadi bagian integral dari kehidupan modern, memengaruhi cara kita berkomunikasi, mencari informasi, berinteraksi dengan orang lain, dan bahkan mempengaruhi opini dan pandangan kita tentang berbagai isu. Meskipun memiliki banyak manfaat, media sosial juga dapat membawa risiko seperti privasi online, penyebaran informasi palsu, dan dampak negatif pada kesehatan mental.

Di era digital media sosial semakin banyak digunakan di Indonesia bahkan di dunia. Media sosial merupakan media berbasis online yang digunakan untuk berinteraksi secara individu maupun kelompok dengan efektif dan efisien yang bersifat tak terbatas. Media sosial sebagai tempat mengajak yang dapat mengubah pandangan dan perilaku publik. Setiap tahunnya, media sosial selalu berkembang pesat dan sudah familiar terutama di kalangan generasi Y dan Z (Ahmad, 2020).

## 2.7. Platform X

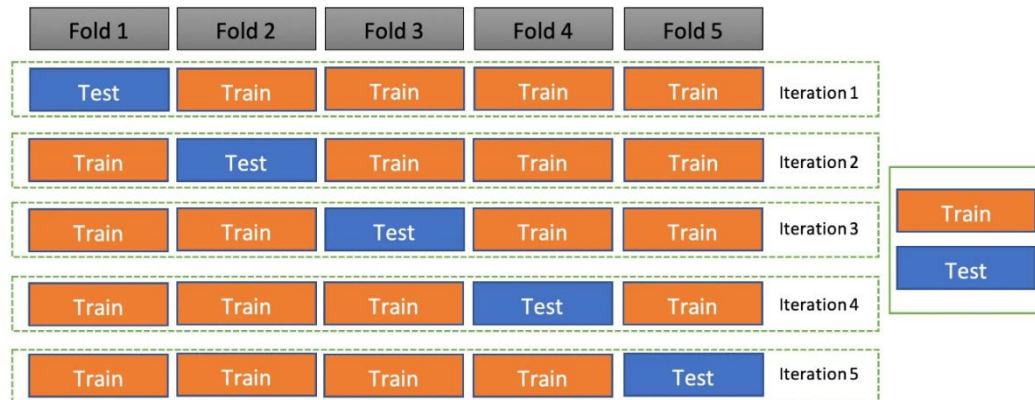
Salah satu media sosial yang populer di kalangan masyarakat Indonesia saat ini adalah platform X, yang sebelumnya dikenal sebagai Twitter. Berdasarkan survei yang dilakukan oleh *We Are Social* (2023), jumlah pengguna Twitter di Indonesia mencapai 14,75 juta per April 2023, menjadikannya peringkat keenam di dunia. X adalah aplikasi media sosial yang menghubungkan individu satu dengan yang lain. Di era informasi yang terus berkembang, Twitter telah menjadi salah satu elemen paling dinamis dan berpengaruh dalam dunia media sosial. Dengan menyediakan platform untuk berbagi ide, pemikiran singkat, dan berita secara real-time, Twitter telah mengubah cara kita berkomunikasi, berpartisipasi dalam percakapan global, dan mengakses informasi.

Sebagai aplikasi media sosial berbasis teks dan gambar, Twitter memungkinkan pengguna untuk mengirim pesan singkat dalam bentuk "tweet" dengan batasan karakter tertentu. Hal ini menciptakan ruang untuk percakapan yang cepat, langsung, dan terkadang mendalam di seluruh dunia. Twitter bukan hanya sekadar aplikasi, melainkan sebuah jendela interaksi sosial yang membuka peluang untuk berbagi pemikiran, kegembiraan, atau menanggapi peristiwa terkini. Pengguna dapat mengikuti akun-akun yang mereka minati, termasuk tokoh publik, perusahaan, dan teman-teman mereka. Fitur "*retweet*" dan "*like*" memungkinkan konten menarik untuk menyebar dengan cepat di seluruh jejaring.

Twitter juga menjadi alat penting untuk mengamati *tren* global, menyampaikan opini publik, dan memfasilitasi gerakan sosial. Meskipun memiliki dampak positif yang signifikan dalam mempercepat penyebaran informasi, Twitter juga menghadapi tantangan terkait dengan disinformasi, batasan karakter yang bisa membuat konten terbatas, dan masalah privasi.

## 2.8. *K-Fold Cross Validation*

*K-Fold Cross Validation* adalah metode penilaian akurasi model klasifikasi. Metode ini cukup sederhana: data dibagi menjadi  $k$  bagian yang sama besar, misalnya jika  $k = 10$ .

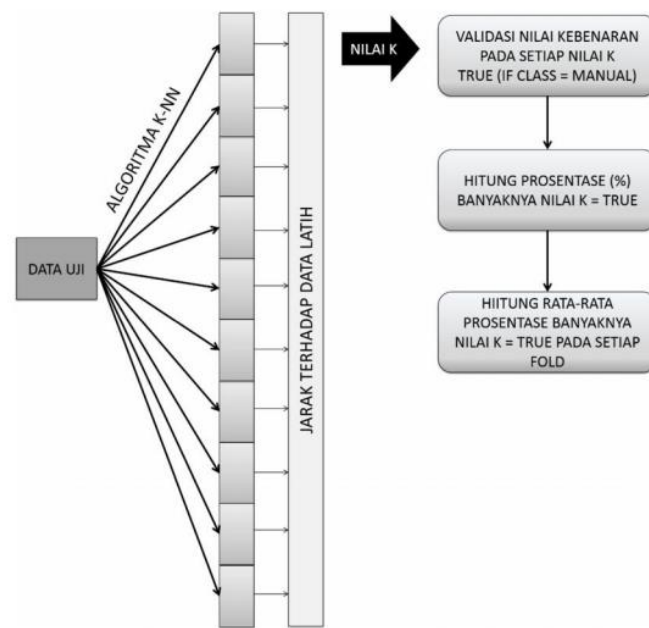


Gambar 2. 2 Pembagian data dengan *K-Fold Cross validation*

(Sumber : <https://sqlrelease.com/>)

Data kemudian diubah menjadi data uji dan data latih. Proses ini diulang untuk setiap bagian data  $k$  setelah pembagian. Gambar 2.2 menunjukkan tahapan berikutnya dalam proses validasi *cross-fold* sepuluh kali lipat menggunakan model algoritma K-NN.

Pada tahap ini, jarak antara data latih dan data uji akan dihitung. Nilai  $k$  tertentu adalah input parameter. Selanjutnya, nilai klasifikasi sebenarnya dari data uji dibandingkan dengan hasil klasifikasi untuk setiap nilai  $k$ . Pada tahap terakhir, tingkat kebenaran atau kesalahan rata-rata dari setiap fold ke- $n$  terhadap setiap nilai  $k$  dihitung.



Gambar 2. 3 Ilustrasi *K-Fold Cross Validation* Dengan  $K = 10$

(Sumber : 1 Bramer, Max. *Op Cit hal. 83*)

Tahapan proses algoritma *K-fold cross validation* dalam pengolahan data dapat dijabarkan sebagai berikut:

1. Masukkan nilai fold (F).
2. Masukkan nilai k.
3.  $T$  = jumlah record dalam dataset.
4.  $S$  = jumlah record data uji ( $T/F$ ).
5. Tetapkan  $L = 1$ .
6. Tetapkan  $M = 0$ .
7. Partisi dataset sebanyak F bagian, setiap partisi berisi S record.
8. Untuk  $I = 1$  hingga F:
9.  $F(I)$  = data uji.
10.  $\text{Not } F(I)$  = data latih.
11. Untuk  $N = 1$  hingga S:
12. Untuk  $J = 1$  hingga k:
13. Jalankan fungsi algoritma k-NN untuk setiap record (N) dalam tabel  $F(I)$  dengan nilai  $k = J$ .
14.  $P$  = hasil prediksi k-NN.

15.  $H$  = nilai atribut target dari data uji ke- $N$ .
16. Jika  $H = P$  maka Nilai = True; jika tidak, Nilai = False.
17. Gantikan hasil untuk  $K = J$  dengan nilai.
18.  $J = J+1$ .
19. Ulangi langkah 13.
20.  $N = N+1$ .
21. Ulangi langkah 12.
22. Selama  $L < k + 1$ :
23. Selama  $L < k + 1$ :
24. Hitung rata-rata akurasi:  

$$\text{Akurasi} = \left( \frac{\sum \text{Nilai}}{F} \right) \times 100\%$$
25. Ulangi langkah 24.
26.  $I = I+1$ .
27. Ulangi langkah 9.
28. Selesai.

## 2.9. *Hypertext Markup Language (HTML)*

HTML adalah bahasa markup yang sangat fleksibel untuk pembuatan halaman web, karena tidak tergantung pada program atau sistem operasi tertentu. HTML tidak peka terhadap kasus, berbeda dengan bahasa pemrograman lain seperti PHP atau ASP yang sangat sensitif terhadap hal-hal tertentu. HTML juga bisa disebut sebagai bahasa yang mudah digunakan untuk mengatur dan menampilkan hypertext (Riswanda & Priandika, 2021).

## 2.10. JavaScript

JavaScript adalah bahasa pemrograman web yang digunakan di sisi klien atau browser dan sering digunakan untuk mengubah elemen HTML dan menambah gaya, atau dengan kata lain, membuat dokumen HTML lebih interaktif. Tag "*script*" digunakan untuk memasukkan JavaScript ke halaman web. JavaScript adalah bahasa pemrograman web yang digunakan di sisi klien atau browser dan sering digunakan untuk mengubah elemen HTML dan menambah gaya, atau dengan kata lain, membuat dokumen HTML lebih interaktif. Tag



"script" digunakan untuk memasukkan JavaScript ke halaman web (Setiawan, 2017).

### 2.11. *Hyper Preprocessor (PHP)*

Menurut Barany dan Jurais Al Qorni (2019), PHP (*Hyper Preprocessor*) adalah bahasa *scripting* yang terutama digunakan untuk pengembangan web. Karena merupakan *Server Side Scripting*, PHP membutuhkan *Web Server* untuk dijalankan. PHP dapat diintegrasikan dengan HTML, JavaScript, JQuery, dan Ajax, meskipun umumnya lebih sering digunakan bersama dengan file HTML. Dengan PHP, pengembang dapat membuat situs web dinamis yang kuat serta mengelola basis data. Selain itu, PHP dapat berjalan di banyak platform. Awalnya, PHP adalah singkatan dari *Personal Home Page* (Situs Personal).

### 2.12. *CodeIgniter*

CodeIgniter adalah sebuah framework PHP *open source* yang menggunakan arsitektur MVC (*Model-View-Controller*). *Framework* ini memungkinkan pengembang untuk membuat situs web dinamis dengan PHP (Rahman & Ratna, 2018). Menurut Sulistiono (2018), CodeIgniter adalah sebuah aplikasi open source berbentuk framework yang digunakan untuk membangun situs web dengan bahasa pemrograman PHP. Tujuan utamanya adalah mempercepat proses pengembangan proyek dibandingkan dengan menulis kode dari awal atau menggunakan kode terstruktur, dengan menyediakan banyak library yang umum digunakan dalam pengembangan.

### 2.13. *TensorFlow*

*TensorFlow* adalah *platform machine learning* yang dapat dijalankan dalam skala besar dan di lingkungan yang beragam. Ini memungkinkan pemetaan aliran data grafik ke berbagai mesin dalam cluster dan perangkat komputasi seperti CPU, GPU, dan TPUs. Dengan arsitektur ini, pengembang memiliki fleksibilitas untuk bereksperimen dengan optimasi dan algoritma baru tanpa harus membangun ulang sistem. *TensorFlow* memiliki dukungan untuk berbagai jenis

aplikasi, terutama untuk pelatihan dan inferensi pada jaringan saraf dalam (Abadi et al., 2016).

#### 2.14. Penelitian Terdahulu

Berikut ini beberapa penelitian yang terkait dengan Metode *K-Nearest Neighbor* (K-NN) :

Penelitian yang dilakukan oleh Balabantaray et al., (2012) yaitu klasifikasi emosi pada kumpulan data Twitter; kumpulan ini dibagi menjadi tujuh kelas, terdiri dari enam kelas emosi *ekman* (kegembiraan, kesedihan, kemarahan, kegembiraan, dan ketakutan), bersama dengan satu kelas netral. Metode SVM memberikan akurasi 73,24% dalam penelitian ini.

Selanjutnya, penelitian yang dilakukan oleh Fanesya et al. (2019) menggunakan metode *Naive Bayes* dan menggabungkan fitur dari lima kelas emosi: cinta, marah, senang, sedih, dan takut. Dalam penelitian ini, model klasifikasi *multinomial naive bayes* dan *bernoulli naive bayes* digunakan untuk tipe data biner dan diskrit, dan tiga kombinasi fitur digunakan: kombinasi fitur N-gram, fitur linguistik, dan ortografik. Pengujian kombinasi fitur N-gram dengan akurasi terbaik adalah 0,5.

Studi yang dilakukan oleh Mubarak et al., (2017) menggunakan metode *bayesian network* menggunakan enam kelas emosi: terkejut, kesedihan, kegembiraan, kebencian, ketakutan, dan kemarahan. Hasil penelitian menunjukkan bahwa teknik yang digunakan untuk melatih *Bayesian Network* tidak cukup efektif untuk menghasilkan model terbaik, dengan nilai f1-score tertinggi 53,71%.

Selanjutnya, penelitian Maulana, (2016) mengidentifikasi emosi manusia dalam tweet bahasa Indonesia menggunakan klasifikasi *Naive Bayes*. Hasil uji coba menunjukkan bahwa kelas emosi netral, senang, dan sedih memiliki deteksi emosi yang cukup baik, dengan fscore masing-masing 77%, 75%, dan 65%. Sebaliknya, kelas marah memiliki hanya 37%, kelas terkejut dan takut memiliki 27%, dan 23%, dan kelas jijik tidak dapat mendeteksi emosi sama sekali.

Studi yang dilakukan oleh Sofiyana et al. (2017) menggunakan metode KNN untuk menentukan emosi dalam teks berbahasa Indonesia. Penelitian ini menggunakan lirik lagu yang akan dikategorikan berdasarkan emosi yang terkandung di dalamnya. Dengan menggunakan data latih yang telah diketahui tentang kelas emosinya, yaitu senang, sedih, marah, bersalah, dan takut, sistem yang akan dikembangkan telah berhasil mengklasifikasikan teks berbahasa Indonesia sesuai dengan kategori emosi dengan nilai 60% pada  $k=5$ .

Tabel 2. 1 Penelitian Terdahulu

No	Penulis	Judul	Tahun	Keterangan
1	Nur Ariska Anggraini dan Nurul Fadillah	Analisis Deteksi Emosi Manusia dari Suara Percakapan Menggunakan Matlab dengan Metode KNN	2019	Menurut penelitian ini, ketika suara dengan nada tinggi atau emosi dimasukkan, grafik menjadi lebih tinggi dan gelombang suara juga muncul.
2	Mohamad Syahrul Mubarak, Muhammad Surya Asriadi dan Adiwijaya	Klasifikasi Emosi Pada Twitter Menggunakan <i>Bayesian Network</i>	2017	Hasil penelitian menunjukkan bahwa teknik pembelajaran <i>bayesian network</i> belum cukup efektif untuk menghasilkan model terbaik; nilai f1-score tertinggi adalah 53,71%, sedangkan nilai f1-score untuk model <i>multinomial naïve bayes</i> adalah 51,49%.
3	Pratama Dwi Nugraha,	Klasifikasi dokumen menggunakan	2018	Menurut pengujian, metode KNN tanpa <i>information Gain</i>

	Said Al Faraby dan Adiwijaya	<i>Metode K-Nearest Neighbor</i> dengan <i>Information Gain</i>		memiliki akurasi rata-rata sebesar 93,94%, sedangkan kombinasi KNN dengan <i>information Gain</i> memiliki akurasi rata-rata sebesar 93,49%.
4	Devi Yunita	Perbandingan Algoritma <i>KNearest Neighbor</i> Dan <i>Decision Tree</i> Untuk Penentuan Risiko Kredit Kepemilikan Mobil	2017	Dalam penelitian tersebut, hasil menunjukkan bahwa algoritma <i>K-Nearest Neighbor</i> (K-NN) menghasilkan nilai akurasi yang lebih tinggi dibandingkan dengan metode <i>Decision Tree</i> . Dengan nilai akurasi sebesar 98%, K-NN terbukti lebih efektif dalam mengklasifikasikan data dibandingkan dengan <i>Decision Tree</i> dalam konteks yang diteliti.
5	Petiwi	Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode <i>Naïve Bayes</i> dan <i>Support Vector Machine</i>	2022	Metode <i>Support Vector Machine</i> sebesar 84,45% dan <i>Naïve Bayes</i> sebesar 69,75%. Kekurangan jumlah data tidak disebutkan. Kelebihan akurasi algoritma SVM lebih dari 80%.

6	Mahardhika Maulana	Deteksi emosi manusi pada tweet bahasa Indonesia dengan klasifikasi naïve bayes	2016	Hasil uji coba menunjukkan bahwa kelas emosi netral, senang, dan sedih memiliki deteksi emosi yang cukup baik, dengan fscore masing-masing 77%, 75%, dan 65%. Sementara kelas marah hanya 37%, kelas terkejut dan takut masing-masing 27% dan 23%, dan kelas jijik tidak dapat mendeteksi emosi sama sekali.
7	Robert Plutchik	<i>A General Psychoevolutionary Theory Of Emotion</i>	2013	Penelitian ini menjelaskan bahwa ada delapan emosi dasar: gembira, percaya, takut, terkejut, sedih, jijik, marah, dan antisipasi. Kemudian emosi-emosi ini dikembangkan menjadi 32 jenis pada manusia. Teori ini menawarkan perspektif positif tentang hubungan antara emosi, adaptasi, dan evolusi.
8	Ramani	Analisis Sentimen Pengguna Twitter terhadap Konflik	2021	Algoritma <i>Support Vector Machine</i> (SVM) lebih baik daripada <i>Naïve</i>

		antara Palestina dan Israel Menggunakan Metode <i>Naïve Bayes Classification</i> dan Support Vector Machine		<i>Bayes</i> dengan akurasi 74% dan 80%. Kekurangan jumlah data hanya sedikit Kelebihannya kedua metode ini cukup baik dalam menentukan nilai akurasi
9	R.C Balabantaray, Mudasir Mohammad dan Nibha Sharma.	<i>Multi-Class Twitter Emotion Classification: A New Approach</i>	2012	Penelitian yang dilakukan dengan metode <i>Support Vector Machine</i> menunjukkan akurasi sebesar 72,34%.
10	Riyan Eko Putri, Suparti dan Rita Rahmawati	Perbandingan Metode Klasifikasi <i>Naïve Bayes</i> Dan <i>K-Nearest Neighbor</i> Pada Analisis Data Status Kerja Di Kabupaten Demak Tahun 2012	2012	Penelitian menemukan bahwa metode KNN memiliki akurasi terbaik sebesar 96,06%, sedangkan metode <i>Naive Bayes</i> memiliki akurasi sebesar 94,09%. Selain itu, laju eror KNN cenderung lebih rendah daripada <i>Naive Bayes</i> , dan nilai Press'Q juga lebih tinggi.

Dibandingkan dengan penelitian sebelumnya yang telah dibahas, penelitian ini akan melakukan klasifikasi dengan menggunakan *K-Nearest Neighbor* yang dituning. Tidak seperti studi sebelumnya yang dilakukan oleh Fanesya et al. (2019) dan Mubarak et al. (2017), di mana metode *Naive Bayes* yang dinilai tidak cukup efektif untuk menghasilkan model terbaik.

Selain itu, penelitian ini berbeda dari penelitian yang dilakukan oleh Sofiyana et al. (2017), yang menggunakan metode KNN untuk mengklasifikasikan emosi pada teks berbahasa Indonesia; dalam penelitian ini, lirik lagu yang digunakan untuk dikategorikan berdasarkan emosinya digunakan. Namun, ini menggunakan data dari tweet Twitter sosial media.

## BAB 3

### ANALISIS DAN PERANCANGAN SISTEM

#### 3.1 Data yang digunakan

Data yang digunakan dalam penelitian ini terdiri dari tweet yang diperoleh melalui *platform* Kaggle. Dataset ini mencakup 2000 *entri* yang memiliki atribut berupa label dan teks tweet yang mencerminkan emosi. Emosi tersebut terbagi menjadi lima kelas: marah, takut, antusias, sedih, dan senang.

Tabel 3. 1 Contoh data tweet emosi

Label	Tweet
Marah	PLN emg php surat edaran maintain daerah a sampe jam 1 tapi skg udh jam 2.40 listrik belum hidup jg. Kalau telat dikit bayar tagihan langsung denda ini klo mati listrik mulu dapet potongan gak?
Takut	Jd kaya kebalik gitu. Dulu socmed jd escape buat gue ekspresiin kegilaan. Skrg berhubung socmed jd lebih serem dr dunia nyata, eh jd dunia nyata yg gue jadiin escape dari socmed.
Antusias	Pertunjukan tadi malam benar-benar spektakuler! Aku masih nggak bisa berhenti memikirkannya!
Sedih	Ikhlas yaa?? "Iyaa buk" jawaban itu yang keluar dari mulut kami pada saat itu. Walau samar, namun masih terdengar. Kami kembali ke kamar kami untuk kemudian diam dalam pikiran masing-masing.
Senang	PERSIBKU MENANG LAGI .. PERSIBKU MENANG LAGI .. Oooo ..Oooo ..Oooo .. Oooooooooooooo .. [USERNAME] #PERSIBDAY #pucuk #JUARA #awaydays #bantul #bobotohsaalamdunya #banggamengawalmu



Variabel penelitian yang digunakan dalam penelitian ini setelah dilakukan praproses pada data teks tweet terdiri dari variabel *predictor* (x) yaitu kata dasar setiap tweet dan variabel respon (y) yaitu klasifikasi emosi tweet.

Tabel 3. 2 Variabel Penelitian

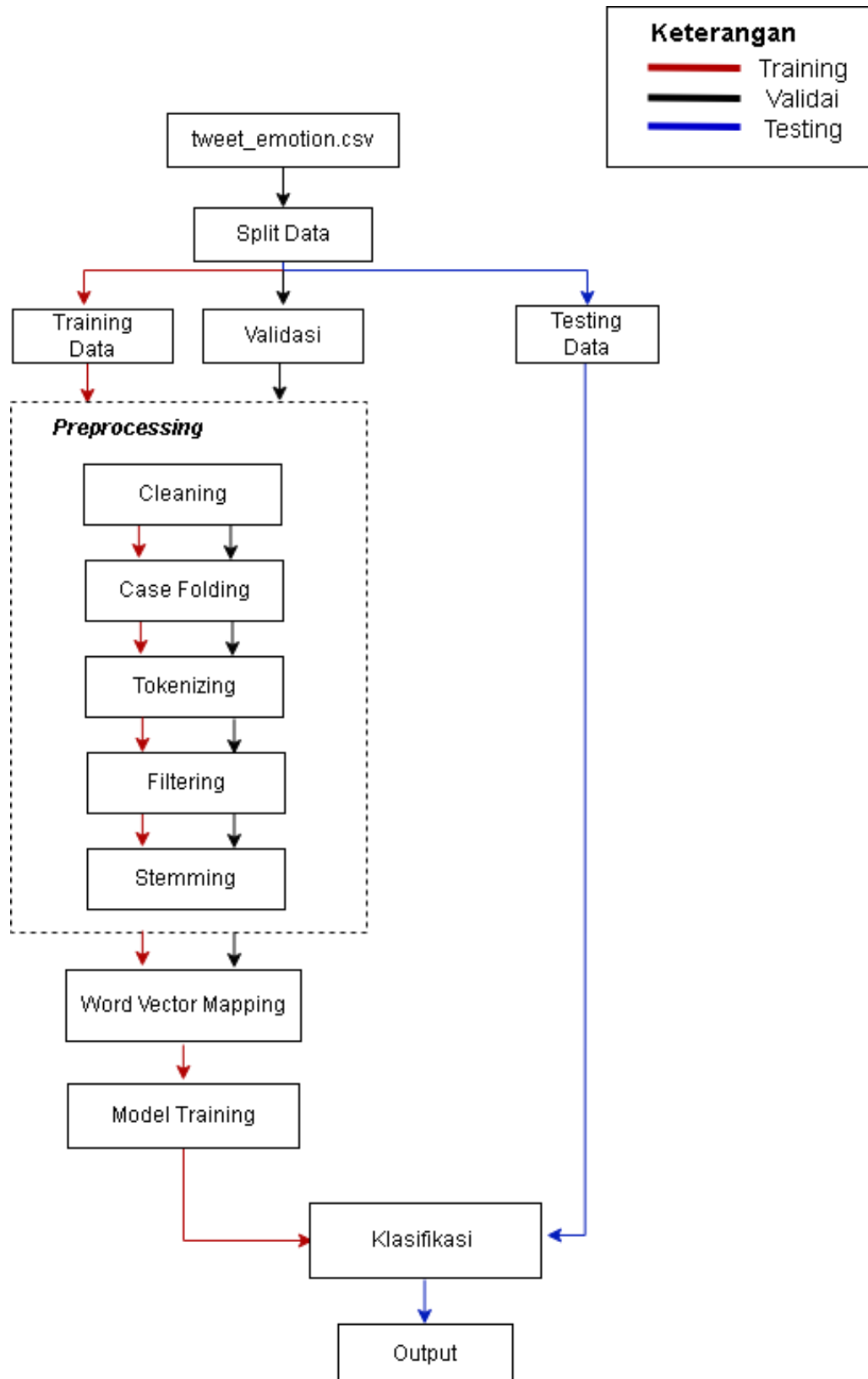
Variabel	Keterangan	Skala
<b>y</b>	Kelas Emosi : 1. Marah 2. Takut 3. Antusias 4. Sedih 5. Senang	Nominal
<b>x</b>	Kata	Nominal

### 3.2 Analisis Sistem

#### 1. Arsitektur Umum

Pada fase berikutnya, akan dilakukan perancangan sistem dimulai dari perancangan arsitektur umum, pengumpulan data, pelatihan model, hingga implementasi ke aplikasi mobile

Rancangan arsitektur umum penelitian ini dapat dilihat pada Gambar 3.1.



Gambar 3. 1 Arsitektur Umum Sistem

## 2. Input Dataset

Pada tahap ini dilakukan pengumpulan data tweet emosi pada platform X. Data tweet emosi akan dikumpulkan dalam format *csv* untuk diproses ke tahap selanjutnya, dimana data diperoleh dari Kaggle. Nantinya data akan terbagi 3 menjadi data *training*, data validasi dan data *testing*. Data training merupakan data yang telah dikumpulkan untuk dilatih, data validasi adalah kumpulan data yang berisi sampel berbeda untuk mengevaluasi model yang telah terlatih dimana dalam penelitian ini akan menggunakan *K-Fold Validation* dalam pengujiannya dan data testing merupakan data yang digunakan untuk pengujian sistem.

## 3. Pre-processing

Pemrosesan data bertujuan untuk mempersiapkan dan membersihkan data agar sesuai untuk analisis lebih lanjut, data yang telah di split selanjutnya akan diproses seperti di *cleaning* dan *case folding*.

## 4. Cleaning

Pada tahap ini, data yang tidak penting dihapus atau ditangani, seperti karakter atau simbol (!@#\$%^&\*():{ }.,?~/[]), hashtag, URL, *mention*, dan emoticon.

```
sentences.forEach(s => {
  let words = s.replace(/[^\w-]/gi, "").toLowerCase().split(" ").filter(x =>
    !!x);
  words.forEach(w => {
    if (!bagOfWords[w]) {
      bagOfWords[w] = 0;
    }
  });
});
```

Gambar 3. 2 Proses Cleaning

Tabel 3. 3 Tampilan tweet sebelum (kiri) dan sesudah (kanan) melalui proses *Cleaning*

#OldMoneyGakNgerasain nongkrong depan rumah ama temen2, tiap ada motor lewat di teriakin woy standar pas orangnya nengok kita sembunyi:))	OldMoneyGakNgerasain nongkrong depan rumah ama temen2, tiap ada motor lewat di teriakin woy standar pas orangnya nengok kita sembunyi
---	--

### 5. *Case folding*

Tahapan *case folding* umumnya melibatkan langkah-langkah berikut: mengubah semua dalam tweet, huruf diubah menjadi huruf kecil (*lowercase*), menghapus angka, tanda baca, operator *aritmetika*, dan karakter khusus.

Tabel 3. 4 Tampilan tweet sebelum (kiri) dan sesudah (kanan) melalui proses *Case Folding*

PERSIBKU MENANG LAGI ..	persibku menang lagi .. persibku
PERSIBKU MENANG LAGI ..	menang lagi .. oooo ..oooo ..oooo ..
Oooo ..Oooo ..Oooo ..	oooooooooooo .. username persibday
Ooooooooooooo .. [USERNAME]	pucuk juara awaydays bantul
#PERSIBDAY #pucuk #JUARA	bobotohsaalamdunya
#awaydays #bantul	banggamengawalmu
#bobotohsaalamdunya	
#banggamengawalmu	

### 6. *Tokenizing*

Tahap selanjutnya adalah tokenisasi, di mana sebuah kalimat tweet dipecah menjadi kata-kata atau token. Langkah ini dilakukan untuk mengidentifikasi kemunculan setiap kata.

Tabel 3. 5 Tampilan tweet sebelum (kiri) dan sesudah (kanan) melalui proses *Tokenizing*

selama gue hidup, gua gapernah berani kenalan sama cowok dari sosmed kayak apk apk dating gitu, serem ga sih	['selama', 'gue', 'hidup,' 'gua', 'gapernah', 'berani', 'kenalan', 'sama', 'cowok', 'dari', 'sosmed', 'kayak', 'apk', 'apk', 'dating', 'gitu', 'serem', 'ga', 'sih']
---	--

### 7. *Filtering*

Tahap selanjutnya adalah filtering, dimana pada proses *filtering* akan menggunakan *stopword removal* untuk menghapus kata-kata umum seperti "dan", "di", "dengan", dan sejenisnya.

Tabel 3. 6 Tampilan tweet sebelum (kiri) dan sesudah (kanan) melalui proses *Filtering*

['Saya', 'sangat', 'senang', 'hari', 'ini', 'ingin', 'bertemu', 'teman', 'lama']	['sangat', 'senang', 'hari', 'ingin', 'bertemu', 'teman', 'lama']
--	---

#### 8. *Stemming*

*Stemming* dilakukan untuk memperoleh bentuk dasar dari sebuah kata dengan menghilangkan imbuhan seperti awalan, sisipan, akhiran, dan kombinasi imbuhan. Proses ini penting untuk menyeragamkan kata-kata dalam dataset sehingga menjadi kata dasar.

Tabel 3. 7 Tampilan tweet sebelum (kiri) dan sesudah (kanan) melalui proses *Stemming*

['gaji', 'gak', 'seberapa', 'tapi', 'kerjaan', 'segudang']	['gaji', 'gak', 'berapa', 'tapi', 'kerja', 'gudang']
--	--

#### 9. Ekstraksi fitur *Word Vector Mapping (Word2Vec)*

Ekstraksi fitur adalah proses mengubah kata-kata dari teks menjadi nilai numerik yang dapat dipahami oleh komputer (Pravina et al., 2019). *Word2Vec* adalah salah satu metode ekstraksi fitur yang mengonversi kata-kata menjadi vektor. Kata-kata dengan makna yang mirip akan menghasilkan vektor yang berdekatan, memungkinkan pembelajaran hubungan semantik antar kata (Triyantonio et al., 2021). *Word2Vec* adalah model *neural network* yang menggunakan kata-kata teks sebagai input dan menghasilkan vektor sebagai *output*. *Word2Vec* dalam proses penelitian ini bekerja untuk mempelajari hubungan antara kata-kata berdasarkan konteks dimana kata kata itu muncul.

#### 10. Klasifikasi

Klasifikasi adalah proses Pengelompokan data ke dalam kelas atau label yang telah ditentukan berdasarkan atributnya. Dalam penelitian ini menggunakan Algoritma *K-Nearest Neighbor* dalam proses klasifikasinya. Prosesnya dimulai dengan menyimpan data latih dan menentukan jumlah tetangga terdekat (k). Kemudian, dihitung jarak antara data baru dan semua data latih. Setelah itu,

dipilih k tetangga terdekat berdasarkan jarak terkecil, dan kelas dari data baru ditentukan berdasarkan mayoritas kelas dari tetangga tersebut.

#### 11. Model *training*

Tahap ini merupakan tahap membangun arsitektur model menggunakan KNN dan melatih model dengan data training.

#### 12. *Output*

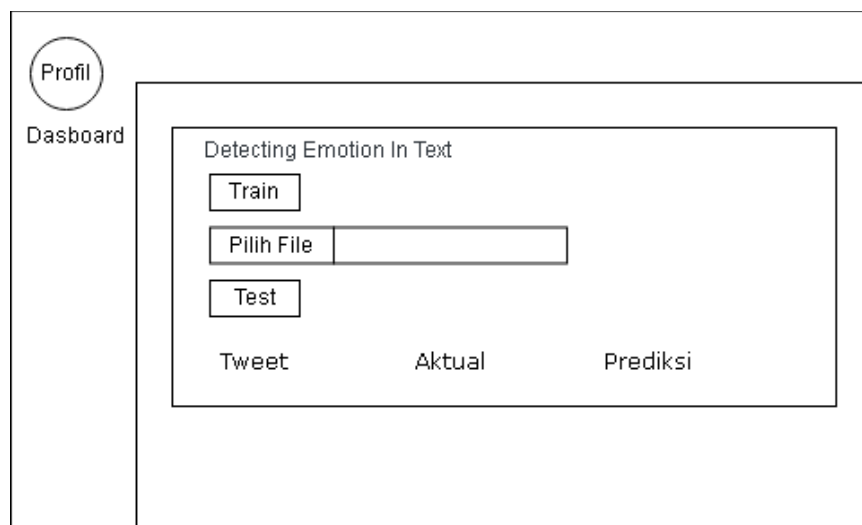
*Output* yaitu hasil dari keluaran sistem yang telah dikelompokkan ke dalam kategori-kategori yang sesuai dengan topik atau kelas yang telah ditentukan. Setiap tweet akan memiliki label atau kelas yang menunjukkan kategori di mana tweet tersebut telah diklasifikasikan.

### 3.3 Perancangan Antarmuka (*User Interface*)

Perancangan antarmuka (*user interface*) pengguna berfungsi sebagai cara utama pengguna berinteraksi dengan sistem yang dibangun, desain antarmuka pengguna sangat penting untuk pengembangan sistem. Dalam penelitian ini

#### 1. Perancangan Halaman Awal

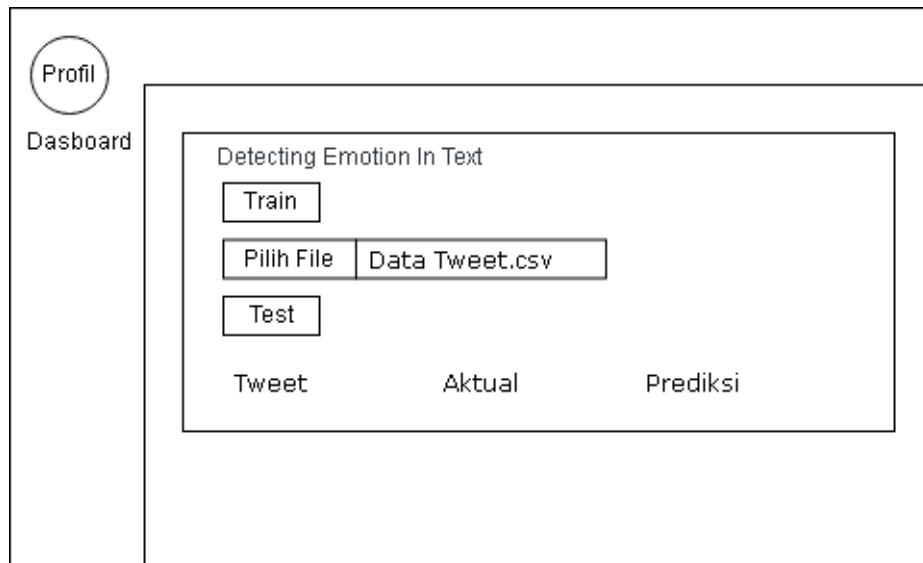
Rancangan tampilan halaman awal, dibagian atas kanan terdapat *header* Profil, di bagian tengah dari *header* tersebut terdapat menu utama, yaitu deteksi emosi pada text dan data tweet akan di *training*, *input file* data *test* dan menu *testing*.



Gambar 3. 3 Perancangan Halaman Awal

## 2. Perancangan tampilan input data latih

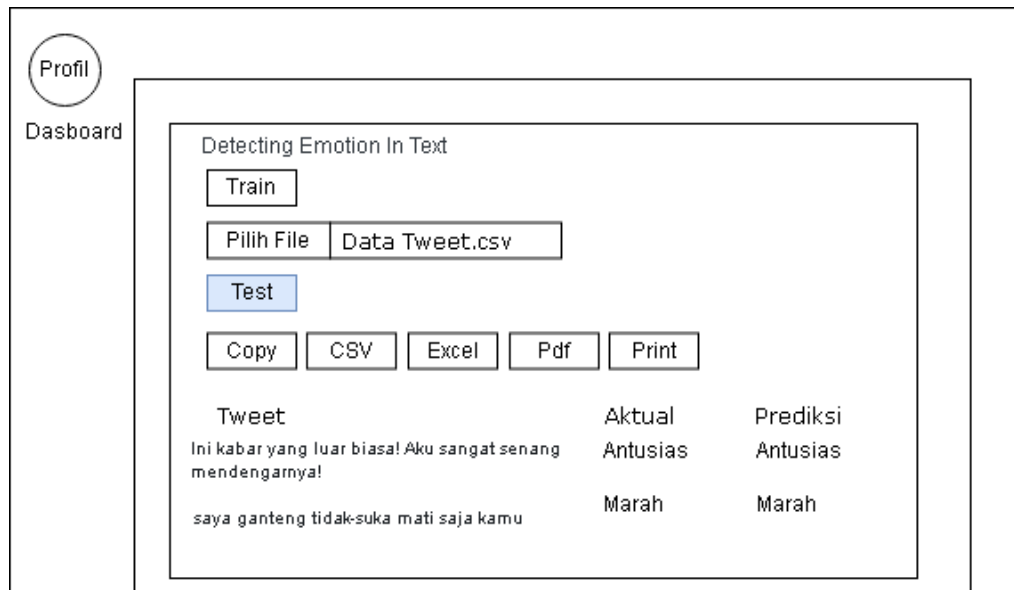
Rancangan tampilan input data latih, dibagian atas kanan terdapat *header* Profil, di bagian tengah dari *header* tersebut terdapat menu utama, yaitu deteksi emosi pada text dan data tweet akan di *training*, setelah data di *train* kita akan menginput data *testing* pada menu “pilih file” data di input, kita akan melakukan *testing*.



Gambar 3. 4 Tampilan input data testing

## 3. Perancangan Halaman Hasil

Rancangan tampilan input data latih, dibagian atas kanan terdapat *header* Profil, di bagian tengah dari *header* tersebut terdapat menu utama, yaitu deteksi emosi pada text dan data tweet akan di *training*, setelah data di *train* kita akan menginput data *testing* pada menu “pilih file” data di *input*, kita akan melakukan *testing*. Setelah melakukan *testing* tampilan hasil akan mengeluarkan *output* hasil tweet.



Gambar 3. 5 Perancangan Halaman Hasil

### 3.4 Perancangan Tabel Hasil Pengujian

#### 1. Perancangan Tabel Hasil Pengujian

Pengujian sistem bertujuan untuk memastikan bahwa aplikasi yang dibuat sesuai dengan desain yang telah direncanakan dan untuk menemukan kesalahan dan kekurangan dalam aplikasi sehingga dapat diperbaiki sebelum digunakan oleh pengguna. Pengujian ini menggunakan kumpulan data tweet yang dikumpulkan dari *platform* Kaggle.

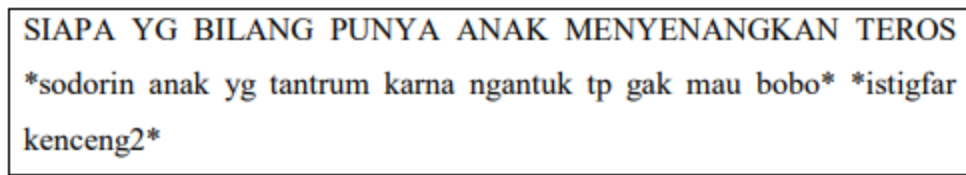
#### 2. Perancangan Tabel Hasil Pengujian Algoritma

Metode *K-Fold Cross Validation* digunakan untuk validasi algoritma K-Nearest Neighbor. Jumlah K adalah jumlah lipatan (*fold*) yang digunakan selama proses validasi. Nilai K yang biasanya digunakan adalah 10. Data sebanyak 2000 tweet kemudian dibagi menjadi sepuluh lipatan (*fold*), sehingga masing-masing lipatan mengandung sebagian dari data.

#### 3. Perhitungan Klasifikasi Tweet Manual dengan Algoritma KNN

Terlebih dahulu, klasifikasi tweet menggunakan K-NN harus melalui tahap pre-processing. Dalam penelitian ini, tahap pre-processing adalah case folding.



A screenshot of a text input field for a tweet. The text inside is: "SIAPA YG BILANG PUNYA ANAK MENYENANGKAN TEROS \*sodorin anak yg tantrum karna ngantuk tp gak mau bobo\* \*istigfar kenceng2\*". The text is in a blue, monospace-style font. The input field has a thin black border.

SIAPA YG BILANG PUNYA ANAK MENYENANGKAN TEROS  
\*sodorin anak yg tantrum karna ngantuk tp gak mau bobo\* \*istigfar  
kenceng2\*

Gambar 3. 6 Tweet Mula-mula

Tweet pertama-tama melalui proses pre-processing, termasuk *case folding*, mengubah huruf kapital menjadi huruf kecil, menghapus karakter *numerik* (angka), tanda baca, operator aritmetika, dan karakter khusus.

## BAB 4

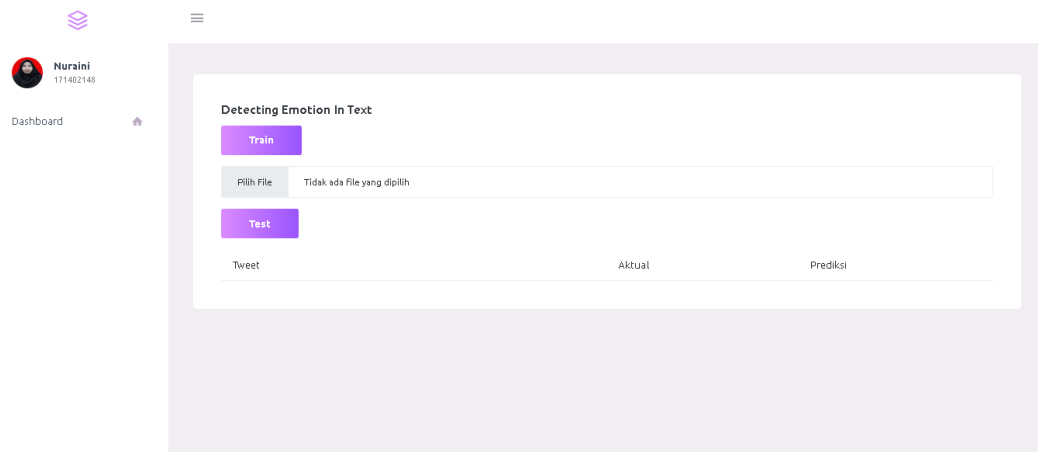
### IMPLEMENTASI DAN PENGUJIAN SISTEM

#### 4.1 Implementasi Sistem

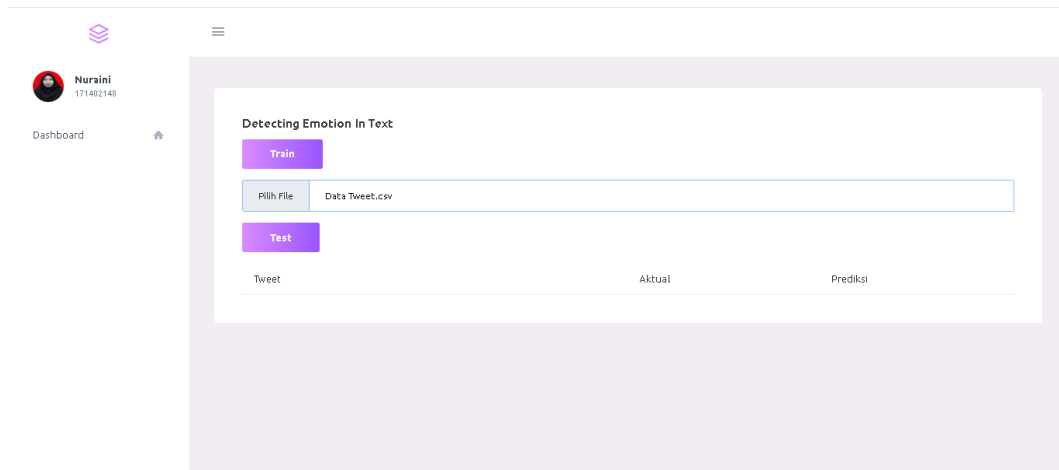
Tahap ini adalah tahap implementasi dari desain antarmuka yang telah dibuat sebelumnya. Berikut ini adalah implementasi aplikasi web berdasarkan menu yang tersedia.

##### 1. Hasil Pengujian Sistem

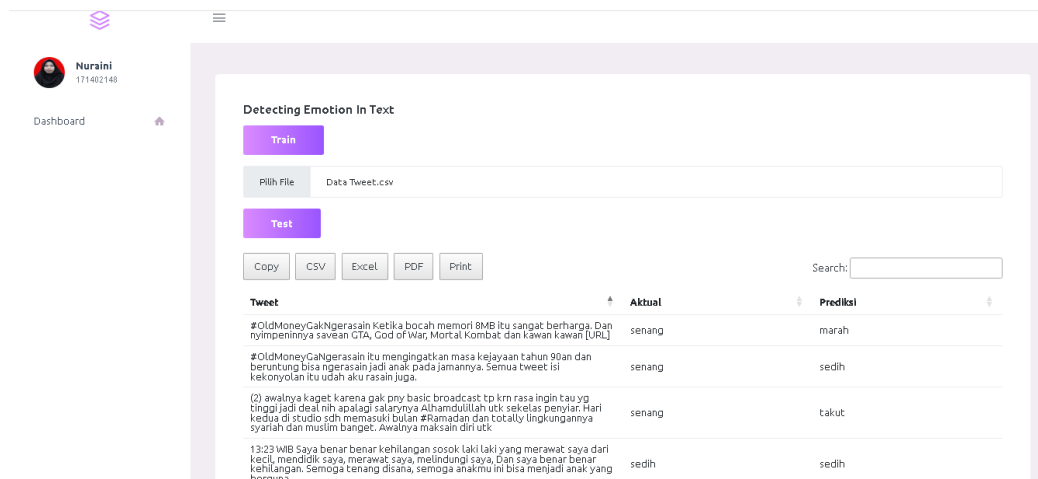
Berdasarkan *user interface* pengguna yang telah dibahas pada bab sebelumnya, maka tampilan antarmuka pengguna yang dihasilkan terlihat seperti pada Gambar 4.1.



Gambar 4. 1 Implementasi Halaman Utama



Gambar 4. 2 Halaman Input data testing



Gambar 4. 3 Hasil Klasifikasi

Tabel 4. 1 Hasil Klasifikasi emosi

Klasifikasi Emosi		
Tweet	Aktual	Prediksi
penyelamat saat laper ya cuma pop mie sob. enak nya bikin nagih dan ga bikin baper. makanya kuy nyetok pop mie dari sekarang biar serunya lanjut terus	senang	Senang
abang gue takut tetiba ditelpon sama pihak rumah sakit atau apa kali secara doctag gue cantumin nama dan no telp dia jadi kalo kenapa kenpaa sama gue, dia yg bakal di telp	takut	Takut

ada mama murid yang rikues anaknya masuk di kelas saya. di saat yang sama banyak murid yang hopeless ketika tau bakal masuk di kelas saya karena rumor yang beredar saya guru galak	sedih	sedih
ada orang yang niat banget bikin akun gaib cuma buat memaki dan memarahi orang lain. marah kok melalui akun gaib	marah	marah
aku baru saja mendaftar untuk kelas yoga! pasti bakal seru dan menyehatkan!	antusias	senang
aku menemukan tempat makan baru yang enak banget kamu harus coba!	antusias	antusias
aku nggak sabar untuk mencoba resep baru ini malam ini	antusias	sedih
anak zaman now lulusan aja pake toga segala un di aula2 mahal, lah kita dulu lulus sd paling banter perpisahan ke cibodas or anyer naek bis..	sedih	sedih
aplikasi pencatatan keuangan semacam money manager memang berguna ya. jadi tahu sebenarnya pengeluaran kita per bulan itu berapa. yg penting konsisten aja catetnya. abis ngeluarin uang, walaupun parkir di indomaret 2000 perak, tetap dicatet.	senang	senang
asli parno bgt kebangun ada bunyi2 deket telinga dan ada yg nonjol gitu di balik gorden	takut	takut
lah elu yang harusnya ditanya teh, kenapa bisa berasumsi guw mah kacang tanpa kulit, maksudnya apa hayo? ke sana juga kan maksud loooo	marah	marah
awalnya ge ga percaya ama istilah indigo, tapi punya temen deket indigo dan selalu tau kalo gw udah ada planning apa ja itu bikin gw heran kaget dan agak takut	takut	takut
baru seminggu menikah. tapi rasanya bersyukur sudah memilih jalan yg allah berikan ini. masyaallah dikasih abang yg sungguh pengertian dan mau nerima cia apa adanya. sampai ga bisa ngebayangin kalau diganti orang lain.	senang	senang
biasa di negeri orang, taat aturan. pulang ke indonesia sudah 3minggu masih kebingungan. kalau ga jahat bakal di jahatin orang disini mah. bener bener	takut	takut
budayakan mencari dan membaca sebelum nanya ya. udah dijelaskan berkali2 kok. selama ga diatur/gaada ketentuan khusus, pakai aja	marah	marah

dicintai adalah rezeki. dicintai orang yang kita cintai adalah rezeki terindah. sedikit merasa cinta dari insan, itupun sudah cukup membahagiakan. pastinya amat besar kan cinta pencipta insan.	antusias	antusias
dua teman sudah menikah, dari dua temen tersebut mereka pindah agama semua. satu muslim ke kristen dan hindu ke muslim. semoga saya di beri jodoh yang seiman dan sejalan.	takut	takut
Gapapa gue udah berusaha untuk sbkm lebaran dari kapan tau tp dospem malah ke luar negri pas tgl tgl sidang kering sudah air mata ini	sedih	sedih
gempa di lombok belum berhenti. setidaknya sudah dua ratusan lebih terjadi gempa susulan sejak pagi kemarin. trus aku bayangin gimana rasanya jadi rangorag yg ada di sana sekarang. pasti gak nyenyak tidurnya karna cemas.	takut	takut
gilak! g1 aja begini. rakan dkk diborong. mungkin kalo masuk 5 momon, zeratu pun ngikut kali. triple s pula aaaaak roolin	senang	senang
gue sama kakak gue adalah orang yang jago makan pedes. lagi musim ayam geprek atau gepuk pastimintanya yang paling pedes. jadi kemarin siang gofood ok cis ayam geprek ria ricis, ternyata enak. malamnya beli lagi dan makan di tempat.	senang	senang
gw tu biasa beli shopee pake atm manual asal cocok jumlahnya di verifikasi otomatis, eh ini keenakan jadi teledor pake akun sepupu gw malah ga otomatis pesanannya kecancel ya allah malih gw kapan ya berenti goblok	sedih	sedih
hari ini adalah hari yang sempurna untuk piknik langitnya begitu cerah dan indah!	antusias	antusias
ijazah mendadak nyelip tu, langsung tiba tiba kebayang perjuangan dikampus, penelitian, skripsi, de el el. emang ya kudu gak ada dulu, baru berasa.	sedih	sedih
ini hasil sekolah asal sekolah ... mikirlah mahkamah internasional itu punya kewenangan mengadili perkara genosida dsbnya ...	marah	marah
ini udah masuk ke proses penjualan... tapi pas dibuka detail ko malah gini? apakah produk sedang diproses atau diproses tgl 11	marah	marah

Tabel 4. 2 Tabel *Confussion Matrix*

Marah	Takut	Antusias	Sedih	Senang
40	0	0	0	0
0	40	0	0	0
0	0	18	3	19
0	0	0	40	0
0	0	0	0	40

Tabel 4. 3 Tabel Nilai T, FP, FN

Data	TP	FP	FN
Marah	40	0	0
Takut	40	0	0
Antusias	18	0	22
Sedih	40	0	0
Senang	40	0	0

Dengan perhitungan matematis dari *Confusion Matrikx*nya adalah sebagai berikut :

$$Presisi = \frac{TP \text{ Marah}}{TP \text{ Marah} + FP \text{ Marah}} = \frac{40}{40 + 0} = \frac{40}{40} = 1$$

$$Presisi = \frac{TP \text{ Takut}}{TP \text{ Takut} + FP \text{ Takut}} = \frac{40}{40 + 0} = \frac{40}{40} = 1$$

$$Presisi = \frac{TP \text{ Antusias}}{TP \text{ Antusias} + FP \text{ Antusias}} = \frac{18}{18 + 0} = \frac{18}{18} = 1$$

$$Presisi = \frac{TP \text{ Sedih}}{TP \text{ Sedih} + FP \text{ Sedih}} = \frac{40}{40 + 0} = \frac{40}{40} = 1$$

$$Presisi = \frac{TP \text{ Senang}}{TP \text{ Senang} + FP \text{ Senang}} = \frac{40}{40 + 0} = \frac{40}{40} = 1$$

$$Recall = \frac{TP\ Marah}{TP\ Marah + FN\ Marah} = \frac{40}{40 + 0} = \frac{40}{40} = 1$$

$$Recall = \frac{TP\ Takut}{TP\ Takut + FN\ Takut} = \frac{40}{40 + 0} = \frac{40}{40} = 1$$

$$Recall = \frac{TP\ Antusias}{TP\ Antusias + FN\ Antusias} = \frac{18}{18 + 22} = \frac{18}{40} = 0,45$$

$$Recall = \frac{TP\ Sedih}{TP\ Sedih + FN\ Sedih} = \frac{40}{40 + 0} = \frac{40}{40} = 1$$

$$Recall = \frac{TP\ Senang}{TP\ Senang + FN\ Senang} = \frac{40}{40 + 0} = \frac{40}{40} = 1$$

Adapun berdasarkan penelitian yang telah dibuat hasil akurasinya sebagai berikut :

$$\begin{aligned} Akurasi &= \frac{Jumlah\ data\ yang\ sesuai}{Jumlah\ keseluruhan\ data} \times 100 \\ &= \frac{40 + 40 + 18 + 40 + 40}{200} \times 100 \% = \frac{178}{200} \times 100 \% = 89\% \end{aligned}$$

Tabel 4. 4 Tabel Nilai TP, FP, dan FN

Data	Presisi	Recall
Marah	1	1
Takut	1	1
Antusias	1	0,45
Sedih	1	1
Senang	1	1
<i>Average</i>	1	1

## 2. Perancangan Tabel Hasil Pengujian Algoritma

Metode *K-Fold Cross Validation* digunakan untuk menguji algoritma. Metode ini membagi data menjadi beberapa *fold*, biasanya sepuluh *fold*, dan kemudian mengambil dari masing-masing *fold* tersebut. Untuk menguji algoritma, peneliti menggunakan perbandingan 9 : 1, yang berarti 1800 data tweet diubah menjadi data latih dan 20 data tweet diubah menjadi data uji untuk setiap *fold*.

*diekstraksi* menjadi token-token per kata dan diproses melalui proses *preprocessing*. Jumlah keseluruhan data test yang akan digunakan yaitu 200 data tweet yang terbagi menjadi 10 fold.

Untuk validasi, data klasifikasi yang dihasilkan oleh aplikasi website disesuaikan secara manual dan dimasukkan ke dalam lima kategori: marah, takut, antusias, sedih, dan senang.

Tabel 4. 5 Hasil 10 *Fold Cross Validation*

Fold Ke-	Akurasi	Error
1	100.00%	0.00%
2	60.00%	40.00%
3	90.00%	10.00%
4	100.00%	0.00%
5	85.00%	15.00%
6	85.00%	15.00%
7	95.00%	5.00%
8	80.00%	20.00%
9	100.00%	0.00%
10	90.00%	10.00%
Rata-rata	88.50%	11.50%



## **BAB 5**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Berdasarkan hasil penelitian pada Bab 4, maka dapat disimpulkan bahwa:

1. Algoritma KNN dapat digunakan untuk mengklasifikasikan data tweet. Algoritma ini dapat digunakan dalam sebuah aplikasi yang dapat mengklasifikasikan data tweet secara otomatis dengan memasukkan data dari tweet-tweet tersebut.
2. Proses klasifikasi data tweet termasuk dalam *Text Mining*, di mana datanya tidak terstruktur. Akibatnya, tweet harus melalui proses pre-processing untuk membuat data teks lebih terstruktur.
3. Pada metode K-Nearest Neighbor didapatkan nilai Presisi 1, *Recall* 1, dan tingkat akurasi yaitu 89%.
4. Proses validasi metode *K-Cross Validation* dapat digunakan untuk mengetahui tingkat akurasi pengklasifikasian pada pengujian algoritma KNN. Hasil akurasi rata-rata 88,50% dan nilai kesalahan 11,50%.

#### **5.2 Saran**

Selama penyusunan tugas akhir mengenai klasifikasi emosi berdasarkan data tweet pada platform X dengan menggunakan algoritma *K-Nearest Neighbor*, ada masalah dengan dataset yang terlalu kecil. Oleh karena itu, diharapkan pengembangan penelitian yang akan datang perlu penambahan dataset yang relatif besar agar menghasilkan fitur yang dapat mempengaruhi tingkat akurasi selama proses klasifikasi.

## DAFTAR PUSTAKA

- Abdillah, Gunawan, Firman Ananda Putra, Faiza Renaldi, (2016). “Penerapan Data Mining Pemakaian Air Pelanggan Untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru Di Pdam Tirta Raharja Menggunakan *Algoritma K-Means*.” Sentika 2016 2016 (Sentika): 18–19.
- Abadi, M., Barham, P., Chen. Z., Davis, A., Dean, J., Devin, M., et al. (2016). *Tensorflow A system for large-scale machine learning. Symposium on Operating System Design an Implementation* (pp. 256-283).
- Alizah, A., et al. (2020). *Application of SEMMA Methodology in Data Mining*. Journal of Data Science, 12(3), 345-367. <https://doi.org/xx.xxx/yyyy>
- Anggraini, Y., Pasha, D., Damayanti, D., & Setiawan, A. (2020). Sistem Informasi Penjualan Sepeda Berbasis Web Menggunakan *Framework Codeigniter*. Jurnal Teknologi Dan Sistem Informasi, 1(2), 64–70. <https://doi.org/10.33365/jtsi.v1i2.236>.
- Ahmad, A. (2020). *Media Sosial dan Tantangan Masa Depan Generasi Milenial*. 08(02), 134–148.
- Aprilia Hastuti, E., Widiarti, E., & Aryani, Y. A. (2023). Pengaruh Penggunaan Media Sosial Twitter Terhadap Kesehatan Mental Emosional Pada Remaja. Jurnal Keperawatan 'Aisyiyah, 10(1), 1–9. <https://doi.org/10.33867/jka.v10i1.353>
- Asriadie, M. S., Mubarak, M. S., & Adiwijaya. (2018). *Classifying emotion in Twitter using Bayesian network. Journal of Physics: Conference Series*, 971(1). <https://doi.org/10.1088/1742-6596/971/1/012041>
- Astuti, L. W., Rachmat C., A., & Lukito, Y. (2017). Implementasi *Algoritma Naïve Bayes* Menggunakan Isear Untuk Klasifikasi Emosi Lirik Lagu Berbahasa Inggris. Jurnal Informatika, 14(1), 16–21. <https://doi.org/10.9744/informatika.14.1.16-21>
- Balabantaray, R. C., Mohammad, M., & Sharma, N. (2012). *Multi-Class Twitter Emotion Classification : A New Approach*. 4(1), 48–53.
- Barany Fachri, & J. (2019). Sistem Pendukung Keputusan Kelayakan Pemberian

SIM (Surat Izin Mengemudi) Kepada Pengendara Sepeda Motor Dengan Menggunakan Metode *Simple Additive Weighting*. Jurnal Ilmu Komputer dan Informatika, .

Brammer, Max. (2007). *Principles of Data Mining*. London : Springer.

Destuardi, & Surya, S. (2009). Klasifikasi Emosi Untuk Teks Bahasa Indonesia. Seminar, (c).

Fera Fanesya , Randy Cahya Wihandika , Indriati. (2019). Deteksi emosi pada twitter menggunakan metode naïve bayes dan kombinasi fitur

Firdaus, D.(2017). Penggunaan Data Mining dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer. Jurnal Format.6(2),91-97

Handayani, F., & Pribadi, F. S. (2015). Implementasi Algoritma *Naive Bayes Classifier* dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan *Call Center* 110. *Jurnal Teknik Elektro*, 7(1), 19–24. <https://doi.org/10.1080/09592310801905785>

Hadi, H. P., & Sukamto, T. S. (2020). Klasifikasi Jenis Laporan Masyarakat Dengan *K-Nearest Neighbor Algorithm*. *JOINS (Journal Of Information System)*, 5(1),77-85. <https://doi.org/10.33633/joins.v5i1.3355>

Hartiwati, E. N. (2022). Aplikasi Inventori Barang Menggunakan Java Dengan *Phpmyadmin*. *Cross-Border*, 5(1), 601–610

Hidayatullah, Priyanto, dan Jauhati Khairul K. (2017). Pemrograman WEB. Bandung: Informatika Bandung.

Irfa, A.A., Adiwijaya, A., & Mubarak, M. S. (2018). Klasifikasi Topik Berita Berbahas Indonesia Menggunakan *K-Nearest Neighbor*. *EProceedings of Engineering*, 5(2).

Imron, M., & Kusumah, S.A. (2018). *Application Of Data Classification Method For Student Graduation Prediction Using K-Nearest Neighbor (K-NN) Algorithm*. *IJIS: Internasional Journal Of Informatics And Information System*, 1(1), 1-8.

Kadhim, H. M. (2020). *Text Classification Using Machine Learning Techniques: A Comparative Study*. Academic Press.

Kantardzic, M.(2019). *Data Mining Concepts. Models, Methods, and Algorithms* 3rd Edition. John Wiley & Sons.

- Kleppmann, Martin dan Alastair R. Beresford. (2017). *A Conflict-Free Replicated JSON Datatype*. University of Cambridge Computer Laboratory, Cambridge.
- Limbong, J. J. A., Sembiring, I., Hartomo, K. D., Kristen, U., Wacana, S., & Korespondensi, P. (2022). Analisis Klasifikasi Sentimen Ulasan Pada *E-Commerce Shopee* Berbasis *Word Cloud* Dengan Metode *Naive Bayes* Dan *K-Nearest Analysis Of Review Sentiment Classification On E-Commerce Shopee Word Cloud Based With Naive Bayes And K-Nearest Neighbor Methods*.9(2), 347-356. <https://doi.org/10.36040/jatir.v5i2.3782>
- Maulana, F., Orisa, M., & Zulfia Zahro', H. (2021). Klasifikasi Data Produk Mebel Aneka Jaya Menggunakan Metode *K-Nearest Neighbor* Berbasis Web. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 5(2), 460-466. <https://doi.org/10.36040/jati.v5i2.3782>
- Mohd, S.S., Rayner, A., (2010). *Advanced Data mining and Applications 6th International Conference, ADMA 2010, Chongqing, China, November 19-21, 2010, Proceedings, Part I*
- Mubarok, M. S., Asriadie, M. S., & Adiwijaya. (2017). *Klasifikasi Emosi Pada Twitter Menggunakan Bayesian Network*. 4(1).
- Nugraha, Dadan. (2018). “Analisis Dan Implementasi Algoritma *K-Nearest Neighbors* Untuk Pengelompokan Member Potensial Produk Market.” *Buffer Informatika* 3 (2): 15–30. <https://doi.org/10.25134/buffer.v3i2.958>.
- Nugroho, A., & Religia, Y. (2021). Analisis Optimasi Algoritma Klasifikasi *Naive Bayes* menggunakan Genetic Algorithm dan Bagging. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*. 5(3), 504-510. <https://doi.org/10.29207/resti.v5i3.3067>
- Pandie, E.S.Y., (2012). Implementasi Algoritma Data Mining *K-Nearest Neighbour (K-NN)* Dalam Pengambilan Keputusan Pengajuan Kredit, Seminar Nasional Sains dan Teknik, vol. 4, Kupang, November 13, 31-34
- Pravina, A., Smith, J., & Doe, R. (2019). *Feature Extraction in Text Mining*. *Journal of Data Science*, 15(2), 123-135.
- Rahman, F., & Ratna, S. (2018). Perancangan E-Learning Berbasis Web Menggunakan *Framework CODEIGNITER*. *Technologia* .

- Rajput, S., & Desai, S. (2021). *Comparison of Text Classification Algorithms for News Articles*. Tech Press.
- Rizqiyani, V., Mulwinda, A., & Putri, R. D. M. (2017). Klasifikasi Judul Buku Dengan Algoritma *Naive Bayes* Dan Pencarian Buku Pada Perpustakaan Teknik Elektro.
- Riswanda, D., & Priandika, A. T. (2021). Analisis Dan Perancangan Sistem Informasi Manajemen Pemesanan Barang Berbasis Online. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 2(1), 94–101.  
<http://jim.teknokrat.ac.id/index.php/informatika/article/view/730>
- Salim, A. O., Laksitowening, K. A., & Asror, I. (2020). *Time Series Prediction On College Graduation Using Knn Algorithm*. 2020 8th Internasional Conference On Information Anf Communication Technology (ICoICT), 1-4.
- Sulistiono, & Heru. (2018). *Coding Mudah dengan CodeIgniter, jQuery, Bootstrap, dan Datatable*. Jakarta: PT. Elex Media Komputindo
- Supriatman, A. (2021). Pembobotan TF-IDF pada judul Penelitian Dosen Sebagai Dasar Klasifikasi Menggunakan Algoritma K-NN (*Studi Kasus : Universitas Siliwangi*). VI(1), 1573-1579
- Setiawan, D. (2017). *Buku Sakti Pemrograman Web : HTML, CSS, PHP, MySQL & Javasript*
- Triyantono, A., Smith, B., & Doe, C. (2021). Implementasi *Word2Vec* dalam Analisis Sentimen. *Jurnal Teknologi Informasi*, 10(1), 45-55.
- Zainuddin, M., Hidjah, K., & Tunjung, I. W. (2016). *Penerapan Case Based Reasoning (CBR) untuk Mendiagnosis Penyakit Stroke Menggunakan Algoritma K-Nearest Neighbor*. *Conference on Information Tecnology Information System and Electrical Engineering (CITISEE)* Yogyakarta, 21-26.



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN,  
RISET, DAN TEKNOLOGI  
UNIVERSITAS SUMATERA UTARA  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

Jalan Universitas No. 9A Gedung A, Kampus USU Medan 20155, Telepon: (061) 821007  
Laman: <http://Fasilkomti.usu.ac.id>

KEPUTUSAN  
DEKAN FAKULTAS ILMU KOMPUTER  
DAN TEKNOLOGI INFORMASI  
NOMOR : 2745/UN5.2.14.D/SK/SPB/2024

DEKAN FAKULTAS ILMU KOMPUTER  
DAN TEKNOLOGI INFORMASI UNIVERSITAS SUMATERA UTARA

- Membaca : Surat Permohonan Mahasiswa Fasilkom-TI USU tanggal 11 Juli 2024 perihal permohonan ujian skripsi:  
Nama : Nuraini  
NIM : 171402148  
Program Studi : Sarjana (S-1) Teknologi Informasi  
Judul Skripsi : Klasifikasi Emosi Teks Pada Media Sosial Platform X Menggunakan Metode K-Nearest Neighbor
- Memperhatikan : Bahwa Mahasiswa tersebut telah memenuhi kewajiban untuk ikut dalam pelaksanaan Meja Hijau Skripsi Mahasiswa pada Program Studi Sarjana (S-1) Teknologi Informasi Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara TA 2023/2024.
- Menimbang : Bahwa permohonan tersebut diatas dapat disetujui dan perlu ditetapkan dengan surat keputusan
- Mengingat : 1. Undang-undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional.  
2. Peraturan Pemerintah Nomor 17 tahun 2010 tentang pengelolaan dan penyelenggara pendidikan.  
3. Keputusan Rektor USU Nomor 03/UN5.1.R/SK/SPB/2021 tentang Peraturan Akademik Program Sarjana Universitas Sumatera Utara.  
4. Surat Keputusan Rektor USU Nomor 1876/UN5.1.R/SK/SDM/2021 tentang pengangkatan Dekan Fasilkom-TI USU Periode 2021-2026
- MEMUTUSKAN
- Menetapkan :  
Pertama : Membentuk dan mengangkat Tim Penguji Skripsi mahasiswa sebagai berikut:  
Ketua : Fanindia Purnamasari S.TI,M.IT  
NIP: 198908172019032023  
Sekretaris : Ade Sarah Huzaifah S.Kom., M.Kom  
NIP: 198506302018032001  
Anggota Penguji : Rossy Nurhasanah S.Kom., M.Kom  
NIP: 198707012019032016  
Anggota Penguji : Dedy Arisandi ST., M.Kom.  
NIP: 197908312009121002  
Moderator : -  
Panitera : -
- Kedua : Segala biaya yang diperlukan untuk pelaksanaan kegiatan ini dibebankan pada Dana Penerimaan Bukan Pajak (PNPB) Fasilkom-TI USU Tahun 2024.
- Ketiga : Keputusan ini berlaku sejak tanggal ditetapkan dengan ketentuan bahwa segala sesuatunya akan diperbaiki sebagaimana mestinya apabila dikemudian hari terdapat kekeliruan dalam surat keputusan ini.

Tembusan :  
1. Ketua Program Studi Sarjana (S-1) Teknologi Informasi  
2. Yang bersangkutan  
3. Arsip

Medan  
Ditandatangani secara elektronik oleh:  
Dekan



Maya Silvi Lydia  
NIP 197401272002122001