



**SEKOLAH TINGGI TEKNOLOGI TERPADU NURUL FIKRI**

**Analisis Sentimen pada Komentar di Platform X menggunakan  
*Support Vector Machine***

**TUGAS AKHIR**

**Fadli Fadillah  
0110220027**

**PROGRAM STUDI TEKNIK INFORMATIKA  
SEKOLAH TINGGI TEKNOLOGI TERPADU NURUL FIKRI  
AGUSTUS 2024**



**STT TERPADU  
NURUL FIKRI**

**SEKOLAH TINGGI TEKNOLOGI TERPADU NURUL FIKRI**

**Analisis Sentimen pada Komentar di Platform X menggunakan  
*Support Vector Machine***

**TUGAS AKHIR**

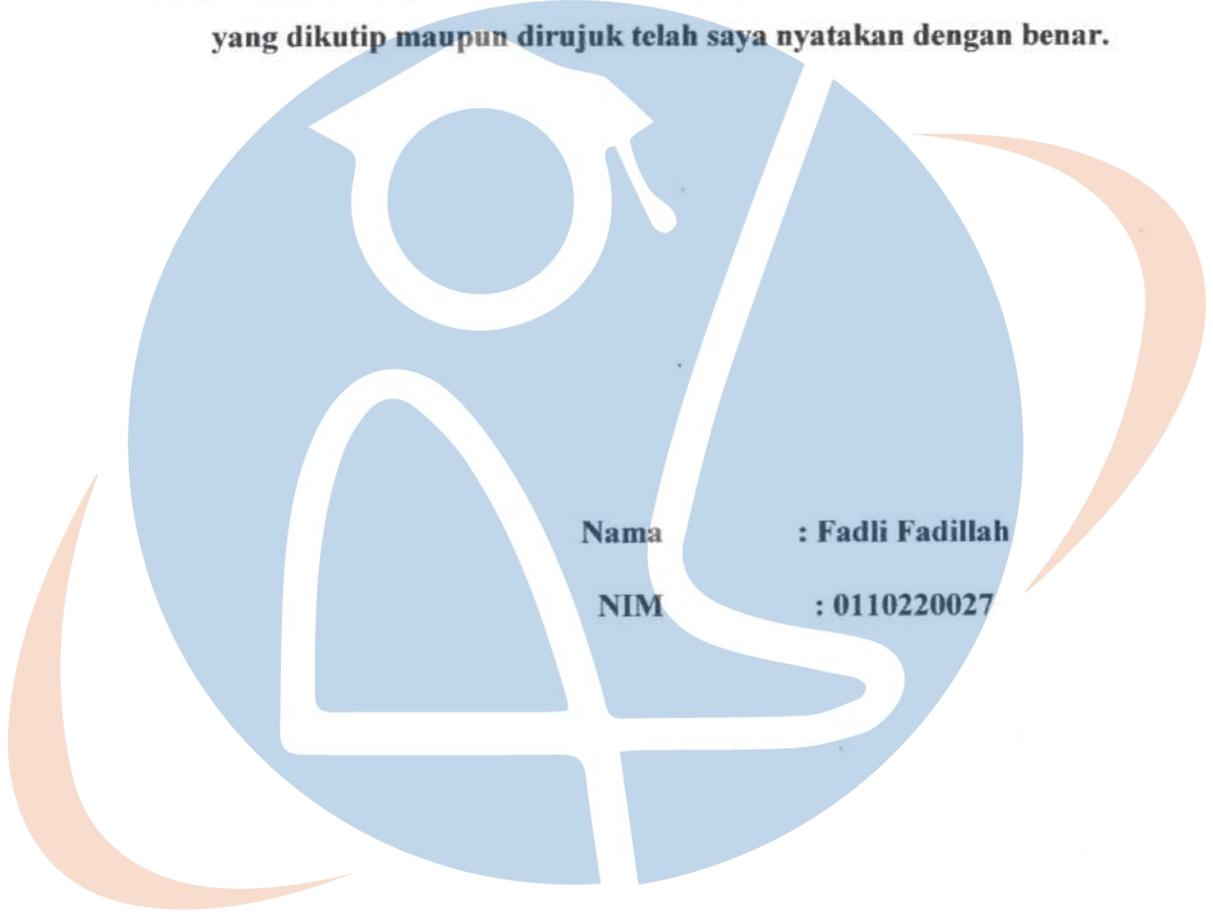
**Diajukan sebagai salah satu syarat untuk memperoleh gelar S.KOM**

**STT** **Fadli Fadillah** **NF**  
**0110220027**

**PROGRAM STUDI TEKNIK INFORMATIKA  
SEKOLAH TINGGI TEKNOLOGI TERPADU NURUL FIKRI  
AGUSTUS 2024**

**HALAMAN PERNYATAAN ORISINALITAS**

**Skripsi/Tugas Akhir ini adalah hasil karya penulis, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar.**



**Nama : Fadli Fadillah**  
**NIM : 0110220027**

**STT - NE** Depok, 13 Agustus 2024  
Tanda Tangan

A handwritten signature in black ink, appearing to be 'Fadli', is written over the 'Tanda Tangan' text.

**Fadli Fadillah**

## HALAMAN PENGESAHAN

Skripsi/Tugas Akhir ini diajukan oleh :

Nama : Fadli Fadillah

NIM : 0110220027

Program Studi : Teknik Informatika

Judul Skripsi : Analisis Sentiment pada Komentar di Platform X menggunakan  
*Support Vector Machine*

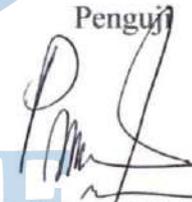
Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Strata 1 pada Program Studi Teknik Informatika, Sekolah Tinggi Teknologi Terpadu Nurul Fikri

DEWAN PENGUJI

Pembimbing

  
Ahmad Rio Adriansyah, S.Si., M.Si.

Penguji

  
Pudy Prima, S.T, M.Kom

Ditetapkan di : Depok

Tanggal : 24 Juli 2024

## KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT, karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan skripsi/Tugas Akhir ini. Penulisan skripsi/Tugas Akhir ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana komputer Program Studi Teknik Informatika pada Sekolah Tinggi Teknologi Terpadu Nurul Fikri. Penulis menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi penulis untuk menyelesaikan skripsi/tugas akhir ini. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Allah SWT.
2. Orang tua dan semua anggota keluarga yang telah memberikan dorongan baik secara moril maupun materil dalam penyelesaian tugas ini.
3. Bapak Dr. Lukman Rosyidi selaku Ketua Sekolah Tinggi Teknologi Terpadu Nurul Fikri.
4. Ibu Tiffany Navarian (S.Kom., M.T.I) selaku Ketua Program Studi Teknik Informatika Sekolah Tinggi Teknologi Terpadu Nurul Fikri.
5. Bapak Lukman Rosidi, S.T., M.M., M.T selaku Dosen Pembimbing Akademik yang telah membimbing penulis selama berkuliah di Sekolah Tinggi Teknologi Terpadu Nurul Fikri.
6. Bapak Ahmad Rio Adriansyah, S.Si., M.Si. selaku Dosen Pembimbing Tugas Akhir penulis dalam menyelesaikan penulisan ilmiah ini.
7. Para Dosen di lingkungan Sekolah Tinggi Teknologi Terpadu Nurul Fikri yang telah membimbing penulis dalam menuntut ilmu yang telah diberikan.

Dalam penulisan ilmiah ini tentu saja masih banyak terdapat kekurangan-kekurangan yang mungkin disebabkan oleh keterbatasan kemampuan dan pengetahuan yang penulis miliki. Walaupun demikian, penulis telah berusaha menyelesaikan penulisan ilmiah ini sebaik mungkin. Oleh karena itu apabila terdapat kekurangan di dalam penulisan ilmiah ini, dengan rendah hati penulis menerima kritik dan saran dari pembaca.

Akhir kata, penulis berharap Allah SWT berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga skripsi ini membawa manfaat bagi pengembangan ilmu.

Depok, 24 Juni 2024



Fadli Fadillah



STT - NF

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI  
TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS**

---

Sebagai sivitas akademik Sekolah Tinggi Teknologi Terpadu Nurul Fikri, saya yang bertanda tangan di bawah ini:

Nama : Fadli Fadillah.

NIM : 0110220027

Program Studi : Teknik Informatika

Jenis karya : Skripsi / Tugas Akhir

demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada STT-NF **Hak Bebas Royalti Noneksklusif (Non-exclusive Royalty - Free Right)** atas karya ilmiah saya yang berjudul :

Analisis Sentimen pada komentar di Platform X menggunakan *Support Vector Machine*

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini STT-NF berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Depok

Pada tanggal : 13 Agustus 2024

STT - NF

Yang Menyatakan



(Fadli Fadillah)

## ABSTRAK

(300 kata )

Nama : Fadli Fadillah  
NIM : 0110220027  
Program Studi : Teknik Informatika  
Judul : Analisis Sentimen pada Komentar di Platform X menggunakan  
*Support Vector Machine*

Media sosial menjadi suatu hal yang sering digunakan oleh kalangan masyarakat di Indonesia, banyak berbagai macam sosial media yang digunakan oleh masyarakat seperti Facebook, Instagram, TikTok, dan X. Pada penggunaannya media sosial seperti X sering digunakan untuk melempar opini terhadap suatu hal dengan fitur cuitannya.

Cuitan yang ditulis di media sosial memiliki kebebasan dalam berekspresi, tak luput dari kebebasan berekspresi tersebut dari dampak Positif dan dampak Negatif yang ada di dalamnya. Cuitan yang dilakukan di sosial media meninggalkan jejak komentar yang ditinggalkan oleh pengguna media sosial, menjadi sumber data yang berharga untuk penelitian dan analisis.

Analisis sentimen pada komentar merupakan sub-bidang dari pemroses bahasa alami atau *Natural Language Processing* yang berkaitan dengan pemahaman dan evaluasi sentimen yang terkandung dalam komentar, ulasan, atau respons yang ditulis oleh pengguna sosial media. Tugas akhir ini melakukan analisa pada cuitan dari platform X menggunakan metode *Support Vector Machine*,

Penelitian ini berusaha memberikan kontribusi terhadap pengembangan analisis sentimen. Hasil uji analisis sentimen pada komentar menggunakan *Support Vector Machine*, menunjukkan hasil akurasi dari analisis sebesar 66% dengan emosi yang diteliti.

Kata kunci : Analisis Sentimen, *Support Vector Machine*, Platform X

## **ABSTRACT**

*Name* : Fadli Fadillah

*NIM* : 0110220027

*Study Program* : *Informatic Engineering*

*Title* : *Sentiment Analysis in Comments on X Platform into Emotions using Support Vector Machine method*

*Social media is one of the things that is often used by people in Indonesia, there are many kinds of social media used by people such as Facebook, Instagram, TikTok, and X Platform. Tweets written on social media have freedom of expression, this freedom of expression has both positive and negative impacts.*

*Tweets made on social media leave a trail of comments left by social media users, providing a valuable source of data for research and analysis.*

*Sentiment analysis in comments is a sub-field of natural language processing which is concerned with understanding and evaluating the sentiment contained in comments, reviews or responses written by social media users.*

*This final project analyzes tweets from the X platform using the Support Vector Machine method, into basic human emotions, namely anger, fear, love, sadness and happiness. This research seeks to contribute to the development of sentiment analysis. The results of the sentiment analysis test on comments using the Support Vector Machine show that the accuracy of the analysis is 66% with the emotions studied.*

*Key words* : *Sentiment Analysis, Support Vector Machine, X Platform*

STI - NF

## DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS.....	i
HALAMAN PENGESAHAN.....	ii
KATA PENGANTAR .....	iii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI.....	v
TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS .....	v
ABSTRAK.....	vi
<i>ABSTRACT</i> .....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR .....	xi
DAFTAR TABEL.....	xii
BAB I PENDAHULUAN .....	1
1.1 Latar belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Tujuan dan Manfaat Penelitian.....	2
1.4 Batasan Masalah.....	3
1.5 Sistematika Penulisan.....	3
BAB II KAJIAN LITERATUR .....	5
2.1 Tinjauan Pustaka .....	5
2.2 Landasan Teori.....	6
2.2.1 Analisis Sentimen.....	6
2.2.2 <i>Support Vector Machine</i> .....	7
2.2.3 <i>Preprocessing</i> .....	8
2.2.4 Model Evaluasi.....	8

2.2.5 Bahasa .....	10
2.2.6 X.....	10
2.2.7 Emosi.....	11
2.2.8 TF – IDF.....	11
2.2.9 <i>Latent Semantic Analysis</i> .....	11
2.2.10 Python .....	11
<b>BAB III.....</b>	<b>13</b>
<b>METODOLOGI PENELITIAN.....</b>	<b>13</b>
3.1 Jenis Penelitian .....	13
3.2 Tahapan Penelitian .....	13
3.2.1 Kajian Literatur.....	14
3.2.2 Perumusan Masalah .....	14
3.2.1 Pengambilan Data .....	14
3.2.2 Perancangan Sistem .....	14
3.3 Rancangan Penelitian .....	15
3.3.1 Pengumpulan data .....	16
3.3.2 Preprocessing .....	16
3.3.3 Seleksi Fitur .....	17
3.3.4 Training dan Testing .....	17
3.3.5 Klasifikasi .....	17
3.3.6 Pengujian.....	17
<b>BAB IV IMPLEMENTASI DAN EVALUASI.....</b>	<b>19</b>
4.1 Perencanaan Sistem.....	19
4.2 Perancangan Sistem .....	20
4.3 Proses Implementasi Sistem.....	21

4.3.1 Import Module .....	22
4.3.2 Pemrosesan dataset.....	22
4.3.3 Preprocessing Data.....	23
4.4 Pengujian Sistem.....	26
4.4.1 Pengujian Model <i>Machine Learning</i> .....	26
<b>BAB V KESIMPULAN DAN SARAN.....</b>	<b>32</b>
5.1 Kesimpulan.....	32
5.2 Saran.....	32
<b>DAFTAR PUSTAKA.....</b>	<b>33</b>



STT - NF

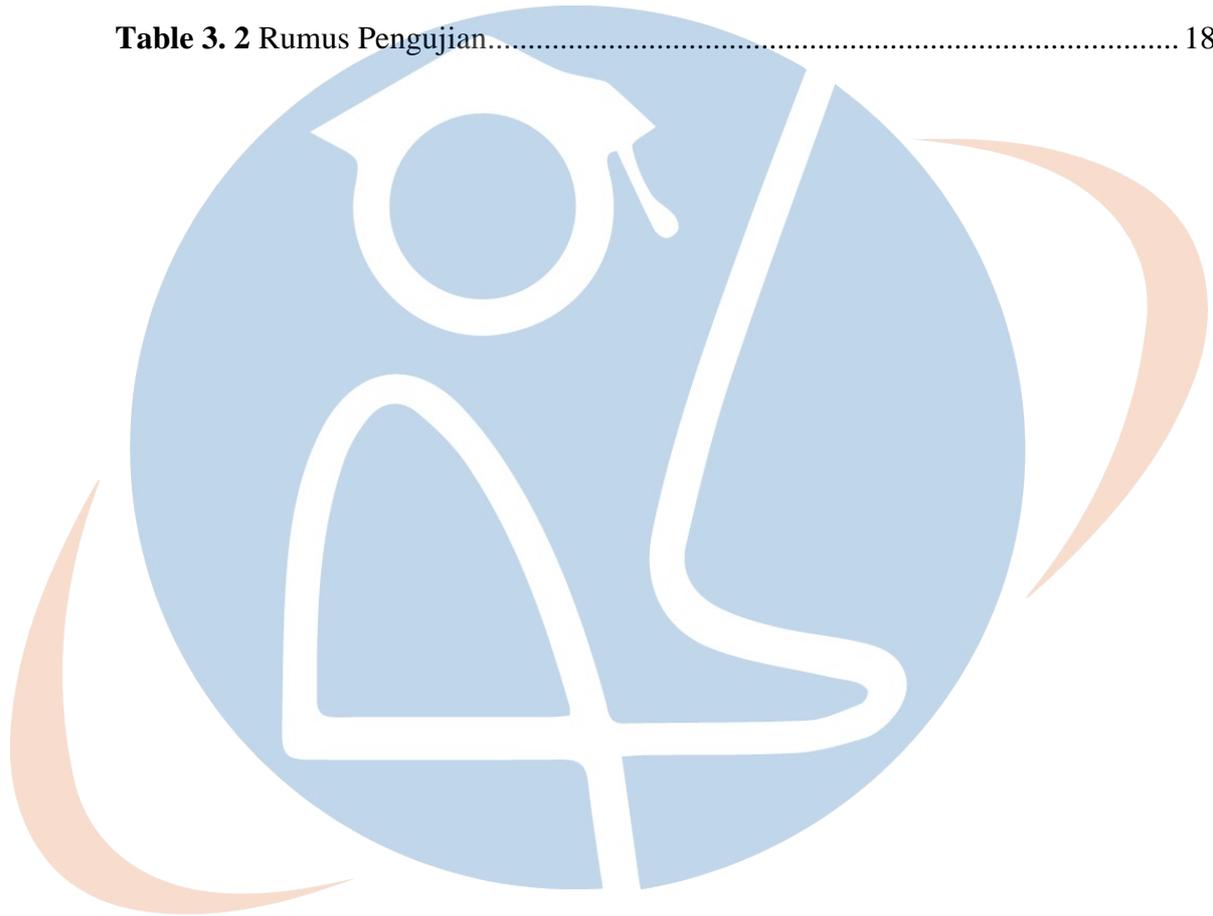
## DAFTAR GAMBAR

<b>Gambar 2. 1</b> Ilustrasi SVM.....	7
<b>Gambar 3. 1</b> Alur Tahapan Penelitian .....	13
<b>Gambar 3. 2</b> Alur Rancangan Penelitian .....	15
<b>Gambar 4. 1</b> Perancangan Sistem.....	20
<b>Gambar 4. 2</b> Sampel Hasil Cuitan .....	23
<b>Gambar 4. 3</b> Confusion Matrix Analisis Sentimen menggunakan TF-IDF.....	29
<b>Gambar 4. 4</b> Confusion Matrix Analisis Sentimen menggunakan LSA .....	29
<b>Gambar 5. 1</b> kata-kata yang terdapat dalam label <i>sadness</i> .....	30
<b>Gambar 5. 2</b> kata-kata yang terdapat dalam label <i>love</i> .....	30
<b>Gambar 5. 3</b> kata-kata yang terdapat dalam label <i>happy</i> .....	31
<b>Gambar 5. 4</b> kata-kata yang terdapat dalam label <i>anger</i> .....	31
<b>Gambar 5. 5</b> kata-kata yang terdapat dalam label <i>fear</i> .....	31

STT - NF

## DAFTAR TABEL

<b>Tabel 2.1</b> Penelitian terkait.....	5
<b>Table 3.1</b> Dataset.....	16
<b>Table 3.2</b> Rumus Pengujian.....	18



STT - NF

# BAB I

## PENDAHULUAN

### 1.1 Latar belakang

Media sosial telah menjadi bagian penting dalam kehidupan sehari-hari banyak orang di seluruh dunia. Platform media sosial seperti Facebook, X, Instagram, dan lainnya memberikan tempat bagi pengguna untuk berinteraksi, berbagi pendapat, dan berpartisipasi dalam diskusi online. Dampak positif media sosial memudahkan transmisi, pengambilan, dan akses informasi[1].

Komunikasi yang dilakukan di sosial media meninggalkan jejak komentar yang ditinggalkan oleh pengguna media sosial, menjadi sumber data yang berharga untuk penelitian dan analisis. Salah satu metode penelitian atau analisis yang dapat digunakan adalah analisis sentimen.

Analisis sentimen pada komentar merupakan sub-bidang dari pemroses bahasa alami atau *Natural Language Processing* yang berkaitan dengan pemahaman dan evaluasi sentimen yang terkandung dalam komentar, ulasan, atau respons yang ditulis oleh pengguna sosial media[2]. Tugas dasar dari metode sentimen analisis terdapat mengklasifikasi beberapa teks, kalimat, atau fitur. Kalimat dari fitur tersebut dapat memiliki berbagai macam emosi yang ditunjukkan. Pada tahapan sentimen analisis terdapat ekstraksi fitur yang berguna mengubah teks menjadi data yang dapat diklasifikasikan.

X merupakan salah satu media sosial yang populer di kalangan pengguna internet secara global. Platform X ini digunakan sebagai sarana komunikasi 2 (dua) arah dan layanan berbagi komentar terhadap sebuah diskusi di internet. Pengguna internet dapat memposting cuitan atau komentar yang dapat berisi gambar, video, tautan dan teks. Mengutip pada artikel yang berjudul “Jumlah Pengguna Twitter Tembus 400 Juta Orang pada Akhir 2022, Ini Trennya Sedekade Terakhir” di website [databoks.katadata.co.id](http://databoks.katadata.co.id) mengutip dari data Business of Apps. Jumlah pengguna X sudah mencapai 401 juta lebih pengguna pada tahun 2022. Di mana pada tahun 2021 pengguna X atau Twitter mencapai 362 juta pengguna, ini berarti ada kenaikan pengguna sekitar 10,77% pengguna X.

Banyaknya pengguna pada platform media sosial X dapat berdampak tidak terkontrolnya cuitan atau komentar yang diupload oleh para pengguna yang dapat menimbulkan berbagai macam reaksi pada setiap komentarnya. Reaksi yang timbul atas setiap komentar memiliki emosi di dalamnya. Emosi yang dianalisa pada penelitian ini merupakan emosi dasar pada manusia[3]. Analisis Sentimen menjadi salah satu metode untuk melakukan Text Processing pada cuitan yang diupload pada X atau Twitter untuk diubah menjadi data yang berguna yang dapat diklasifikasikan. Dengan dilakukannya klasifikasi pada cuitan diharapkan dapat mengidentifikasi berbagai macam reaksi emosi pada cuitan tersebut.

Penelitian ini akan berfokus dalam melakukan ekstraksi menggunakan *TF-IDF* dan *Latent Semantic Analysis* pada cuitan atau komentar yang diupload pada platform X sebagai solusi untuk mengidentifikasi reaksi emosi pengguna platform media sosial tersebut. Tujuan dari penelitian ini adalah untuk mengklasifikasikan reaksi emosi pada cuitan di platform X. Penelitian dalam hal ini menjadi kunci untuk memahami, mengatasi, dan memanfaatkan beragam sentimen yang terdapat dalam cuitan atau komentar pengguna di platform X.

## **1.2 Rumusan Masalah**

Beberapa hal yang perlu dikaji untuk menjawab “Apakah sentimen analisis dapat mendeteksi emosi pada komentar ?” adalah sebagai berikut :

- a. Seberapa akurat metode *Support Vector Machine* mendeteksi ekstraksi emosi di dalam cuitan ?
- b. Bagaimana langkah untuk mendeteksi sebuah emosi pada komentar ?
- c. Bagaimana perbandingan Ekstraksi Fitur yang digunakan ?

## **1.3 Tujuan dan Manfaat Penelitian**

### **1.3.1 Tujuan**

Penelitian ini memiliki tujuan yang mendukung untuk meningkatkan analisis sentimen pada teks, yaitu:

- a. Mendapatkan hasil akurasi ekstraksi emosi pada cuitan.
- b. Mengetahui langkah-langkah dalam analisis sentimen emosi pada komentar

- c. Mengetahui perbandingan proses dan hasil akurasi antara metode ekstraksi fitur terhadap analisis sentimen

### 1.3.2 Manfaat

Penelitian ini memberikan manfaat untuk mendukung analisis sentimen pada teks, yaitu:

- a. Mendapatkan hasil akurasi ekstraksi emosi pada komentar
- b. Mendapatkan metode ekstraksi fitur yang tepat untuk komentar
- c. Memudahkan penyedia layanan untuk mendeteksi emosi pada komentar, dan
- d. Membantu penyedia layanan dalam mengatasi mental pengguna berdasarkan ekstraksi emosi pada komentar pengguna.

### 1.4 Batasan Masalah

Pada penelitian tugas akhir ini terdapat batasan-batasan masalah yang perlu diperhatikan sebagai berikut:

- a. Penelitian ini hanya menguji bahasa Indonesia saja
- b. Penelitian ini hanya menguji 5 (lima) emosi saja, yaitu senang, marah, sedih, takut, dan cinta.
- c. Penelitian ini hanya dapat mengidentifikasi 1 (satu) emosi pada sebuah komentar.
- d. Penelitian ini hanya menggunakan data sekunder
- e. Penelitian ini tidak bisa mendeteksi kata yang memiliki ambiguitas

### 1.5 Sistematika Penulisan

Pada penelitian ini untuk memudahkan dalam memahami penelitian tugas akhir secara menyeluruh diperlukan sistematika penulisan, yaitu:

#### 1. BAB I Pendahuluan

Pada bab ini penulis menceritakan latar belakang dari penelitian yang dilakukan, berdasarkan referensi yang penulis temukan di literatur yang penulis baca, serta merumuskan apa saja masalah yang penulis temukan pada penelitian ini dan menentukan batasan masalah dari penelitian yang penulis lakukan.

#### 2. BAB II Kajian Literatur

Pada bab ini dijelaskan kajian literatur yang penulis temukan untuk melakukan penelitian. Teori-teori terkait serta landasan teori yang penulis temukan sebagai dasar untuk melakukan penelitian

### 3. BAB III Metodologi Penelitian

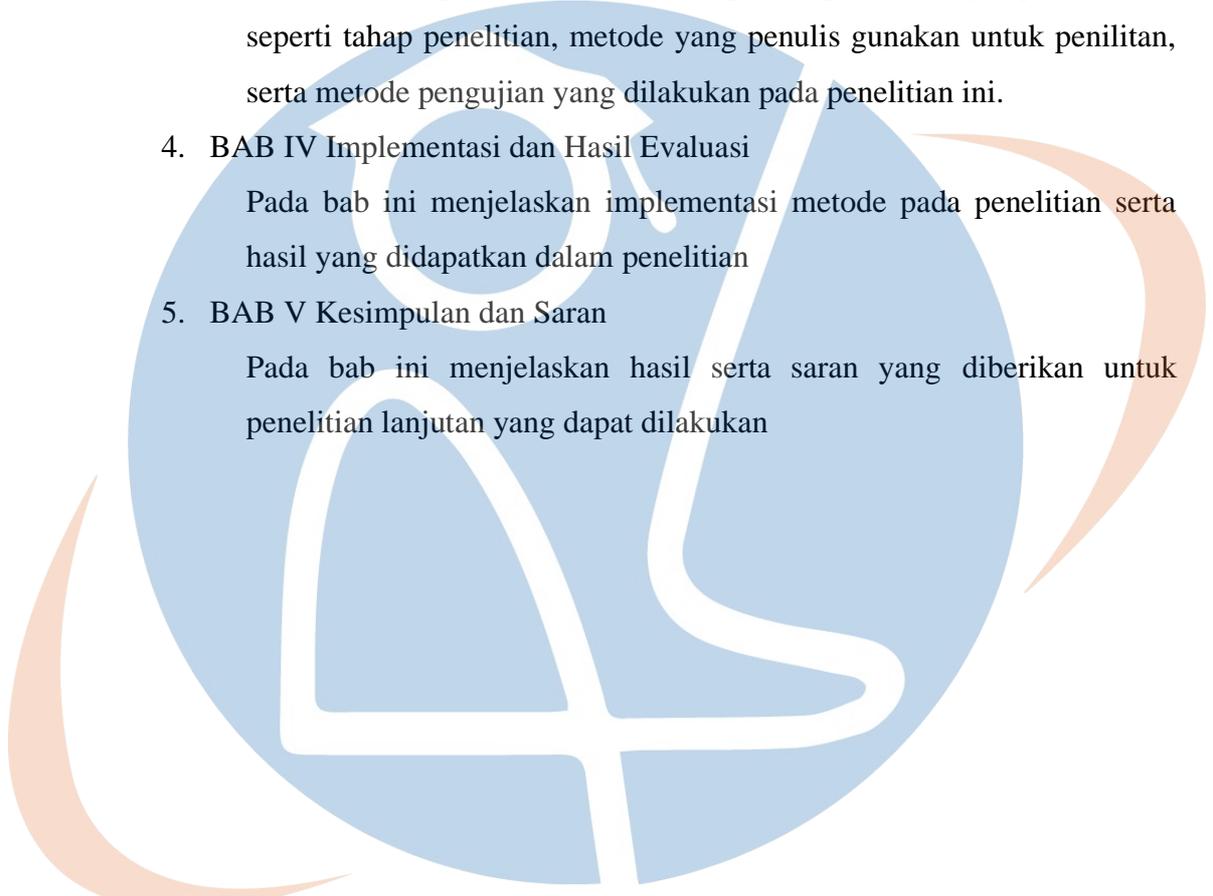
Pada bab ini penulis menjelaskan proses penelitian yang dilakukan seperti tahap penelitian, metode yang penulis gunakan untuk penilitan, serta metode pengujian yang dilakukan pada penelitian ini.

### 4. BAB IV Implementasi dan Hasil Evaluasi

Pada bab ini menjelaskan implementasi metode pada penelitian serta hasil yang didapatkan dalam penelitian

### 5. BAB V Kesimpulan dan Saran

Pada bab ini menjelaskan hasil serta saran yang diberikan untuk penelitian lanjutan yang dapat dilakukan



STT - NF

## BAB II

### KAJIAN LITERATUR

Kajian literatur ini bertujuan untuk memberikan landasan teori yang kuat untuk memahami dan menyelidiki topik penelitian ini, yaitu “Analisis Sentimen pada Komentar di Platform X *Support Vector Machine*”. Melalui kajian literatur ini, penulis berupaya menjelaskan konsep analisis sentimen, metode yang digunakan dalam penelitian sebelumnya, dan temuan signifikan terkait penggunaan analisis sentimen pada media sosial.

#### 2.1 Tinjauan Pustaka

Tabel 2 1 Penelitian terkait

No	Nama dan Tahun	Judul	Topik	Subjek	Hasil
1	Bagus Muhammad Akbar, 2021	Analisis Sentimen dan Emosi Vaksin Sinovac pada Twitter menggunakan metode Naive Bayes dan Valence Shifter	Analisis Sentimen	Komentar atau Teks	Informasi sentimen pada komentar
2	Auliya Rahman Isnain, 2021	Sentimen Analisis Publik terhadap kebijakan lockdown pemerintah Jakarta menggunakan algoritma SVM	Analisis Sentimen	Komentar atau teks	Informasi sentimen pada komentar
3	Yasinta Aulia Nur Jannah, 2022	Analisis sentimen dan emosi publik pada awal pandemi Covid-19 berdasarkan data twitter dengan pendekatan berbasis Leksikon	Analisis Sentimen	Komentar atau teks	Informasi sentimen pada komentar

Penelitian oleh Auliya Rahman Isnain pada tahun 2021, mengeksplorasi sentimen publik terhadap kebijakan lockdown oleh pemerintah Jakarta. Penelitian tersebut memeriksa pergeseran perilaku baik budaya, etika dan norma yang ada, dalam mengeluarkan opini-opini yang mereka miliki[4].

Penelitian oleh Bagus Muhammad Akbar pada tahun 2021, mengeksplorasi sentimen dan emosi pengguna twitter terhadap vaksin sinovac menggunakan metode naive bayes. Informasi mengenai vaksin sinovac memunculkan berbagai tanggapan publik yang tidak luput dari pro dan kontra.

Penelitian oleh Yasinta Amalia Nurjanah pada tahun 2022, dengan topik permasalahan Analisis sentimen dan emosi pada awal pandemi Covid-19 berdasarkan data twitter dengan pendekatan berbasis Leksikon. Dalam penelitiannya dengan menggunakan metode leksikon menunjukkan hasil sentimen negatif yang lebih banyak diekspresikan dibandingkan hasil sentimen positif[5].

Sementara itu pada penelitian lainnya yang diteliti oleh Rizki Wijayatun Pratiwi pada tahun 2021, penelitian ini mengeksplorasi sentimen pada *review* produk *Skincare Female Daily* menggunakan metode *Support Vector Machine*. Metode ini memberikan hasil akurasi yang maksimal dengan tingkat akurasi sebesar 87% dengan *recall* sebesar 90%, *precision* sebesar 84,90%, dan *f1 score* sebesar 87,3% [6].

Sedangkan penelitian yang dilakukan oleh penulis, mengeksplorasi tentang cuitan kalimat yang memiliki ekspresi sentimen emosi dengan menggunakan metode *Support Vector Machine* sebagai *machine learning* dan perbandingan 2 (dua) metode ekstraksi fitur. SVM dipilih sebagai metode pembelajaran mesin karena memproses data secara efektif dan efisien sehingga merupakan metode ini cocok untuk mencapai hasil akurasi yang optimal[6]. Emosi yang ingin diteliti pada kalimat cuitan ini merupakan emosi dasar pada manusia, seperti senang, marah, cinta, takut, dan sedih.

## **2.2 Landasan Teori**

Landasan teori menjelaskan konsep, teori, dan prinsip yang menjadi dasar di dalam penelitian ini. Berikut adalah landasan teori yang penulis susun.

### **2.2.1 Analisis Sentimen**

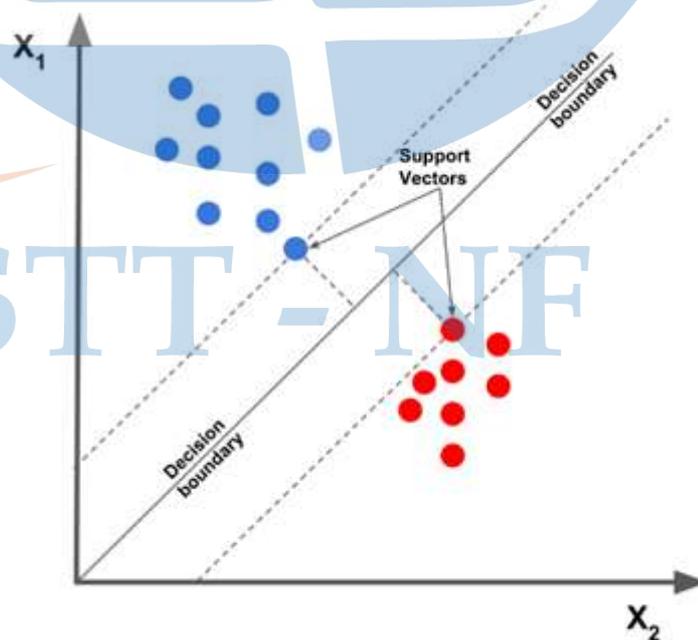
Analisis sentimen adalah bidang penelitian yang bertujuan untuk mengidentifikasi, mengklasifikasi, dan menganalisis sentimen atau opini dalam teks. Sentimen tersebut dapat berupa positif, negatif, netral, dan analisis ini dapat

diterapkan pada berbagai domain, termasuk media sosial, ulasan produk, dan berita. Analisis sentimen dapat dibagi menjadi metode klasifikasi pembelajaran mesin dan metode klasifikasi berbasis aturan[4].

Metode klasifikasi berbasis aturan menggunakan aturan linguistik dan pemrograman komputer untuk mengidentifikasi sentimen dalam teks. Contoh aturan ini termasuk pengenalan kata kunci dan aturan tata bahasa. Sedangkan metode klasifikasi menggunakan pembelajaran mesin merupakan metode klasifikasi sentimen. Algoritma seperti *Support Vector Machine* dan *Naive Bayes* sering digunakan dalam metode ini.

### 2.2.2 *Support Vector Machine*

*Support Vector Machine* (SVM) adalah teknik klasifikasi dan regresi dalam pembelajaran *machine learning* yang efektif di berbagai konteks. SVM digunakan untuk memisahkan kelas atau kategorisasi data dengan mencari *hyperline* terbaik yang memisahkan dua kelas[7]. SVM beroperasi dengan mencari *hyperline* terbaik yang dapat memisahkan dua kelas dalam ruang berdimensi tinggi. *Hyperline* ini dipilih sedemikian rupa sehingga jarak antara *hyperline* dan instance terdekat dari setiap kelas, yang disebut vektor pendukung, memberikan nilai maksimum.



Gambar 2. 1 Ilustrasi SVM

Dalam penelitian ini, SVM digunakan sebagai metode karena dapat memisahkan kelas-kelas pada data yang diinput. Hal ini dilakukan dengan menemukan garis yang dapat memisahkan kelompok data[8].

### **2.2.3 Preprocessing**

*Preprocessing* merupakan tahapan pengolahan data untuk memastikan bahwa data yang digunakan dalam analisis atau pemodelan dapat mencapai bentuk yang sesuai. Tahapan ini mencakup berbagai teknik dalam proses untuk membersihkan, mengubah, dan mempersiapkan data mentah agar siap digunakan dalam metode *machine learning*. Tahapan *preprocessing* yang digunakan untuk penelitian ini sebagai berikut :

a. *Stemming*

*Stemming* merupakan salah satu metode pada *preprocessing data*, yaitu proses mencari kata-kata dasar pada teks dengan menghilangkan imbuhan dari teks yang diproses[9].

b. *Stopwords*

*Stopwords* merupakan salah satu metode pada *preprocessing data*, yaitu proses menghilangkan kata-kata umum pada teks atau dokumen, seperti “dan”, “di”, “atau”, “seperti”, “dengan” dalam bahasa Indonesia[9].

c. *Tokenizing*

*Tokenize* merupakan salah satu metode pada *preprocessing data*, yaitu proses untuk memecah teks-teks menjadi unit-unit yang lebih kecil, yang disebut token. Token dapat berupa kata-kata, frasa, atau karakter, yang tergantung pada tingkat analisis yang diinginkan[10].

### **2.2.4 Model Evaluasi**

Pada *machine learning* metode evaluasi sangat penting digunakan untuk memastikan bahwa model yang dikembangkan dapat berfungsi dengan baik pada data yang diujikan ke dalam model *machine learning*. Pada penelitian ini metode evaluasi yang digunakan adalah *Confusion Matrix*. *Confusion Matrix* adalah alat

evaluasi untuk klasifikasi data yang menunjukkan jumlah prediksi benar dan salah yang dilakukan oleh model dibandingkan dengan label aktual dalam data uji.

Berdasarkan *Confusion Matrix* penulis dapat menentukan nilai performa penelitian, di antara nilai penelitian yang ditentukan adalah sebagai berikut:



STT - NF

a. *Accuracy*

*Accuracy* merupakan kemampuan mesin untuk mengukur rasio bernilai benar dengan keseluruhan jumlah data[11].

$$\frac{TP + TN}{TP + FP + FN + TN}$$

b. *Precision*

*Precision* merupakan kemampuan mesin untuk mengukur kinerja sistem yang dibangun. Dalam hal ini *precision* merupakan perbandingan antara jumlah dokumen relevan dengan jumlah dokumen yang ditemukan melalui pencarian.[12].

$$Precision = \frac{TP}{TP + FP}$$

c. *Recall*

*Recall* adalah kemampuan mesin untuk menentukan apakah sistem berhasil menemukan kembali informasi yang relevan. Dalam hal ini *recall* melakukan perbandingan dokumen relevan pada sistem[12].

$$Recall = \frac{TP}{FP + FN}$$

### 2.2.5 Bahasa

Bahasa merupakan salah satu sistem atau cara manusia untuk berkomunikasi. Bahasa disampaikan dalam bentuk ekspresi sebagai alat komunikasi dalam situasi tertentu dalam berbagai aktivitas[13]. Sistem ini merupakan simbol vokal yang disepakati yang digunakan oleh komunitas pengguna untuk berkomunikasi. Pada penelitian ini bahasa yang digunakan adalah bahasa indonesia, bahasa ini digunakan sebagai data yang dilatih untuk diteliti.

### 2.2.6 X

X atau twitter merupakan salah satu website yang sering dikunjungi oleh pengguna internet, dan X dianggap website *microblogging* terpopuler didunia, menjadi tempat berinteraksi secara daring[14]. X sudah mencapai 373 juta pengguna di seluruh dunia pada tahun 2023 lalu. Dalam menggunakan media sosial X, pengguna yang terdaftar dapat memposting gambar, teks, dan video yang dikenal

sebagai cuitan (*tweet*). Bagi pengguna yang tidak terdaftar di platform X hanya dapat membaca *tweet* melalui *user interface* pada situs web.

### **2.2.7 Emosi**

Menurut artikel “Jenis Emosi dan Reaksi Emosional” di website Verywell Mind, emosi adalah reaksi yang dihasilkan dari pengalaman manusia terhadap suatu peristiwa atau kejadian. Emosi manusia adalah pola perubahan kompleks yang mencakup gairah fisiologis, emosi, proses kognitif, dan respons perilaku sebagai respons terhadap situasi yang dianggap penting secara pribadi[5].

Sebagai contoh adalah ketika seseorang merasa senang ketika mendapatkan kabar baik, atau seseorang akan merasa takut ketika mereka terancam. Pada penelitian ini terdapat 5 emosi yang diteliti, yaitu cinta, senang, sedih, takut, dan marah. Ke-5 emosi tersebut merupakan emosi dasar yang ada pada manusia[3] menurut penelitian berjudul “Lima Emosi dasar Manusia melalui seni new media”.

### **2.2.8 TF – IDF**

*TF-IDF* adalah salah satu metode paling umum untuk menghitung bobot setiap kata. Dalam metode ini, kata-kata diekstraksi menjadi angka-angka berformat vektor, yang digunakan untuk menentukan bobot kata-kata[15]. Dokumen teks yang diekstraksi diubah menjadi vektor dan kumpulan data digunakan untuk membuat fitur baru[16].

### **2.2.9 Latent Semantic Analysis**

*Latent Semantic Analysis (LSA)* adalah teknik pemilihan fitur yang mirip dengan *TF-IDF*. *LSA* adalah metode pengindeksan dan pengambilan otomatis yang menggunakan struktur semantik (hubungan antara istilah dan dokumen) yang tersirat dalam dokumen untuk menemukan dokumen yang terkait dengan istilah dalam kueri.[17].

### **2.2.10 Python**

Pada penelitian ini, penulis melakukan uji terhadap data menggunakan bahasa pemrograman dan modul. Bahasa pemrograman yang penulis gunakan untuk melakukan uji coba adalah pemrograman Python. Pemrograman python merupakan bahasa pemrograman yang kuat, mudah dibaca dan dipahami oleh

penggunanya[18]. Python juga memiliki beberapa modul yang digunakan untuk melakukan *processing* dataset untuk mempermudah penulisan program, modul yang dapat digunakan sebagai *preprocessing* data, yaitu:

a. Modul NLTK

NLTK merupakan platform untuk membangun bahasa pemrograman python agar lebih mudah digunakan dan dapat bekerja dengan data bahasa manusia. NLTK dapat digunakan untuk pemrosesan teks, klasifikasi, tokenisasi, *stemming*, *tagging*, dan *semantic reasoning*[19].

b. Modul *Sklearn*

*Scikit-learn* atau *sklearn* merupakan platform *open-source* untuk menggunakan model *machine learning* dan *modeling library* lainnya untuk pemrograman python[20].

c. Modul *matplotlib*

*Matplotlib* merupakan *library* yang digunakan untuk melakukan visualisasi statis, visualisasi animasi, dan visualisasi interaktif menggunakan python[21].

STT - NF

## BAB III METODOLOGI PENELITIAN

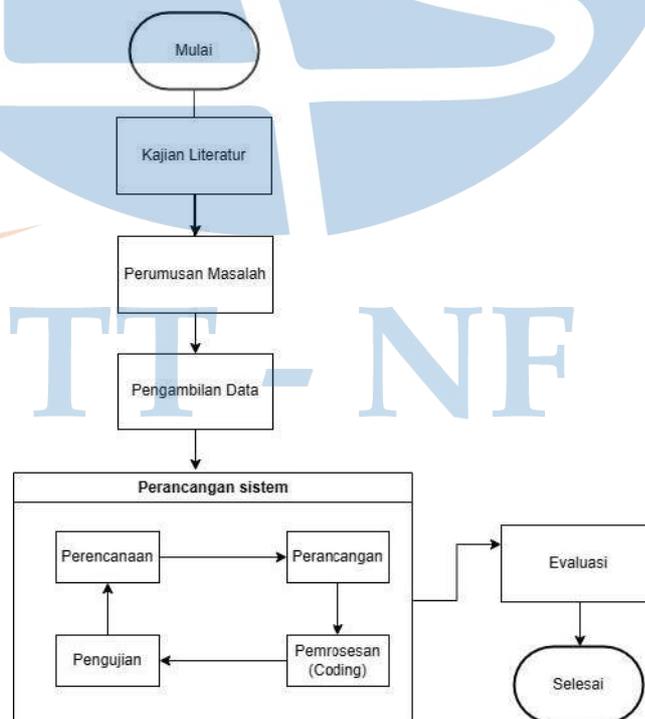
Bab ini menjelaskan rancangan penelitian, langkah-langkah yang digunakan dalam pengumpulan dan analisis data, serta metode yang digunakan dalam penelitian studi tugas akhir ini.

### 3.1 Jenis Penelitian

Jenis penelitian yang digunakan adalah penelitian eksploratif. Jenis penelitian ini bertujuan untuk menyelidiki dan memahami fenomena untuk memperoleh wawasan dasar mengenai suatu topik masalah yang belum banyak diketahui. Tujuan utama dari penelitian ini adalah mengeksplorasi pemahaman mendalam terhadap makna dan konteks di mana suatu fenomena terjadi, dalam hal ini sentimen dan bentuk teks sebuah tweet.

### 3.2 Tahapan Penelitian

Berikut ini merupakan tahapan-tahapan yang akan dilakukan dalam penulisan ini secara umum :



**Gambar 3. 1** Alur Tahapan Penelitian

### **3.2.1 Kajian Literatur**

Tahapan ini dilakukan untuk memahami dan mengevaluasi literatur terdahulu yang relevan dengan topik penelitian terkait. Pada tahapan ini ditemukan beberapa penelitian terdahulu yang dapat dijadikan acuan sebagai sumber referensi penelitian, yaitu penelitian oleh Auliya Rahma Isnain yang berjudul “Sentimen Analisis Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma SVM”[4].

### **3.2.2 Perumusan Masalah**

Tahapan ini dilakukan untuk mengidentifikasi, merumuskan, dan memperjelas masalah yang akan menjadi fokus utama dalam penelitian ini. Dalam tahap ini penulis mengidentifikasi topik penelitian yang relevan berdasarkan observasi, literatur, atau dalam kebutuhan praktis bidang tertentu, dalam hal ini adalah sentimen analisis.

### **3.2.1 Pengambilan Data**

Tahap ini dilakukan untuk memastikan bahwa data yang diperoleh dalam penelitian ini relevan, akurat, dan dapat diandalkan. Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari website *Kaggle*. Data yang dikumpulkan ini terakhir di update pada tahun 2022 lalu.

### **3.2.2 Perancangan Sistem**

Tahapan ini dilakukan untuk pengembangan sistem pemrosesan data menjadi efektif dan efisien. Pada tahapan ini dibagi menjadi beberapa tahap, di antaranya:

#### **a. Perencanaan Sistem**

Pada tahap pertama ini penulis melakukan analisis kebutuhan berdasarkan instrumen pengumpulan data sebelumnya. Sehingga penulis dapat memahami alur sistem yang dibutuhkan untuk mendapatkan gambaran sistem yang jelas mengenai pemrosesan data dan hasil yang dibutuhkan.

b. Perancangan Sistem

Pada tahap kedua ini penulis membuat pemodelan pada sistem pemrosesan berdasarkan analisis kebutuhan. Perancangan ini menggunakan *Flowchart* sebagai bagan alir pemrosesan.

c. Pemrosesan Sistem (Koding)

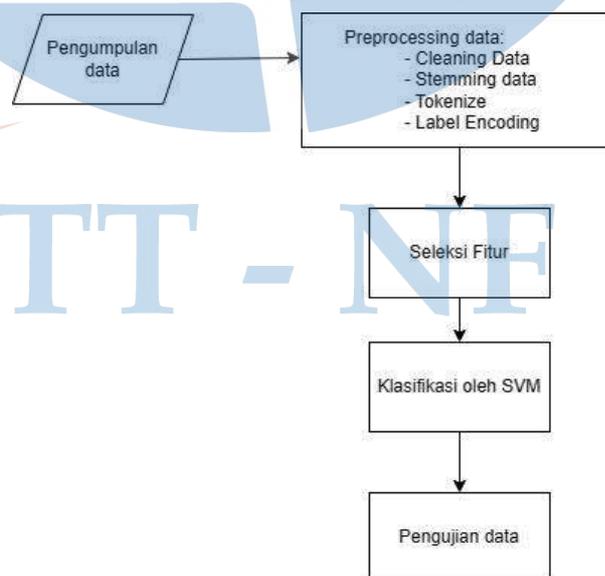
Pada tahap ketiga ini penulis mengimplementasikan pemodelan sistem yang sudah dirancang pada tahap sebelumnya dengan menggunakan bahasa pemrograman yaitu Python dan modul-modul yang dibutuhkan seperti pandas, numpy, nltk, sklearn, dan modul lainnya.

d. Pengujian Sistem

Pada tahap pengujian ini data diuji dengan metode *Support Vector Machine* untuk menguji nilai akurasi data berdasarkan nilai aktual dan prediksi dari metode yang digunakan.

### 3.3 Rancangan Penelitian

Penelitian yang dilakukan oleh penulis memiliki beberapa rancangan penelitian, ini dilakukan agar penelitian berjalan sesuai dengan kaidahnya, berikut beberapa rancangan yang penulis gunakan :



**Gambar 3. 2** Alur Rancangan Penelitian

### 3.3.1 Pengumpulan data

Proses pengumpulan data dan informasi dari berbagai sumber di Internet dilakukan sebagai pengumpulan dataset dan merupakan tahap pertama penelitian ini. Data yang didapatkan untuk penelitian ini merupakan data sekunder yang sudah tersedia di website bernama “kaggle”. Data ini berisi cuitan atau komentar yang ada di platform X secara umum dan sudah diberikan label emosi yang diteliti yaitu *anger* (marah), *fear* (takut), *happy* (senang), *love* (cinta), *sadness* (sedih).

Table 3. 1 Dataset

Cuitan	Emosi
"Soal jln Jatibaru,polisi tdk bs GERTAK gubernur .Emangny polisi tdk ikut pmbhasan? Jgn berpolitik. Pengaturan wilayah,hak gubernur. Persoalan Tn Abang soal turun temurun.Pelik.Perlu kesabaran. [USERNAME] [USERNAME] [URL]"	Anger
"Kepingin gudeg mbarek Bu hj. Amad Foto dari google, sengaja, biar teman-teman jg membayangkannya. Berbagi itu indah."	Happy
"Setiap kesempatan yg pernah hadir tuk dapat membuatmu selalu merasa aman dan nyaman, kini jadi suatu kehormatan yg pernah didapat"	Love
"Separuh hati ini iri jika melihat seorang anak yg masih bisa duduk atau berjalan bersama ibundanya, dan separuh hati ini turut berempati melihat seorang anak yg begitu gundah saat ibundanya sakit. Yang sabar ya kawan, semoga Allah memberikan yg terbaik untuk kita semua."	Sadness
"Hari ini jadwal presentasi proker di LPPM karna ada praktikum jadinya pak ketua di temani sekretaris 1. Tapi selaku penanggung jawab ada di sekretaris 2 jadi deg deg an ini. Bismillah pak ketua (Btw pen Upload foto pak ketua tapi jangan dulu yaa)"	Fear

### 3.3.2 Preprocessing

*Preprocessing* merupakan tahapan yang dilakukan untuk mengolah data agar data dapat diproses untuk ke tahap selanjutnya, dalam tahap *preprocessing* dilakukan dengan beberapa proses, seperti pembersihan data, *Stemming*, Tokenisasi teks, dan Label Encoding. Proses pembersihan data berupa penghilangan simbol-simbol yang ada didalam teks yang dapat mengurangi nilai pada setiap teks.

### **3.3.3 Seleksi Fitur**

Seleksi fitur pada penelitian ini dilakukan untuk mengekstraksi fitur pada cuitan dengan menggunakan metode *TF – IDF* dan *Latent Semantic Analysis (LSA)*. Metode *TF - IDF* merupakan proses seleksi fitur untuk meningkatkan akurasi dan mengurangi waktu komputasi rancangan sistem penelitian sedangkan *LSA* merupakan metode pengindeksan dan pengambilan otomatis yang menggunakan struktur semantik (hubungan antara istilah dan dokumen). Ekstraksi yang dilakukan dapat dilihat pada tabel 3.1.

### **3.3.4 Training dan Testing**

Pada tahap pelatihan dan pengujian, data yang diproses sesuai algoritma SVM diproses ulang hingga mencapai dataset yang ditentukan. Tahap pelatihan membuat model pembelajaran mesin dari data, dan tahap pengujian menguji data dengan algoritma mesin vektor dukungan untuk menentukan kinerja pembelajaran mesin yang dibuat. Data latih yang digunakan adalah data berlabel untuk menentukan nilai setiap kata. Nilai ini menjadi nilai masukan untuk proses pelatihan dan pengujian.

### **3.3.5 Klasifikasi**

Model support vector machine dengan *TF – IDF* digunakan untuk klasifikasi pada penelitian ini. Fase *TF-IDF* melakukan pemilihan fitur dalam teks. Tahap klasifikasi *TF-IDF* diharapkan lebih efisien. Untuk membuat klasifikasi lebih efisien, penulis mengurangi jumlah data analisis dengan mengidentifikasi fitur, yang kemudian diproses berdasarkan model pengklasifikasi yang dibuat selama proses pelatihan penelitian ini. Data untuk klasifikasi ini dibagi menjadi dua bagian yaitu data latih dan data uji.

### **3.3.6 Pengujian**

Pada tahap pengujian, nilai *Accuracy*, *Precision*, dan *Recall*. Pengujian data yang dilakukan meliputi perolehan nilai dari proses pelabelan dan hasil klasifikasi menggunakan model support vector machine. Pengujian dilakukan dengan menggunakan metode *Confusion Matrix* sebagai berikut:

**Table 3. 2** Rumus Pengujian

<b>Jenis</b>	<b>Rumus</b>
<i>Accuracy</i>	$((TP + TN) / (TP + TN + FP + FN)) * 100$
<i>Preciosion</i>	$(TP / (FP + TP)) * 100$
<i>Recall</i>	$(TP / (TP + TN)) * 100$
F1 – Score	$(2 * Recall * Precision) / (Recall + Precision)$



STT - NF

## BAB IV

### IMPLEMENTASI DAN EVALUASI

Pada Bab ini berisi penerapan dari konsep, metode, dan sistem yang dirancang oleh penulis.

#### 4.1 Perencanaan Sistem

Pada tahap ini membahas tentang perencanaan sistem yang kemudian akan digunakan oleh penulis untuk melakukan penelitian. Tahapan ini di antaranya, seperti:

1. Mencari studi literatur terdahulu yang membahas tentang Analisis Sentimen dan penggunaan metode *Support Vector Machine (SVM)*.
2. Menentukan data yang diperlukan untuk penelitian yang akan dilakukan berdasarkan studi literatur yang sudah ditemukan.
3. Mengumpulkan data
4. Menentukan metode pendukung, seperti *preprocessing data*, *processing data*, dan proses evaluasi data, yang nantinya akan digunakan untuk perancangan sistem.
5. Mencari sumber-sumber kode yang dapat membantu pengimplementasian metode, dan
6. Menentukan hasil akhir yang penulis inginkan terhadap penelitian ini.

Setelah penulis melakukan perencanaan untuk pengelolaan sistem, selanjutnya penulis menentukan perencanaan kebutuhan sistem untuk pengelolaan Analisis Sentimen sebagai berikut:

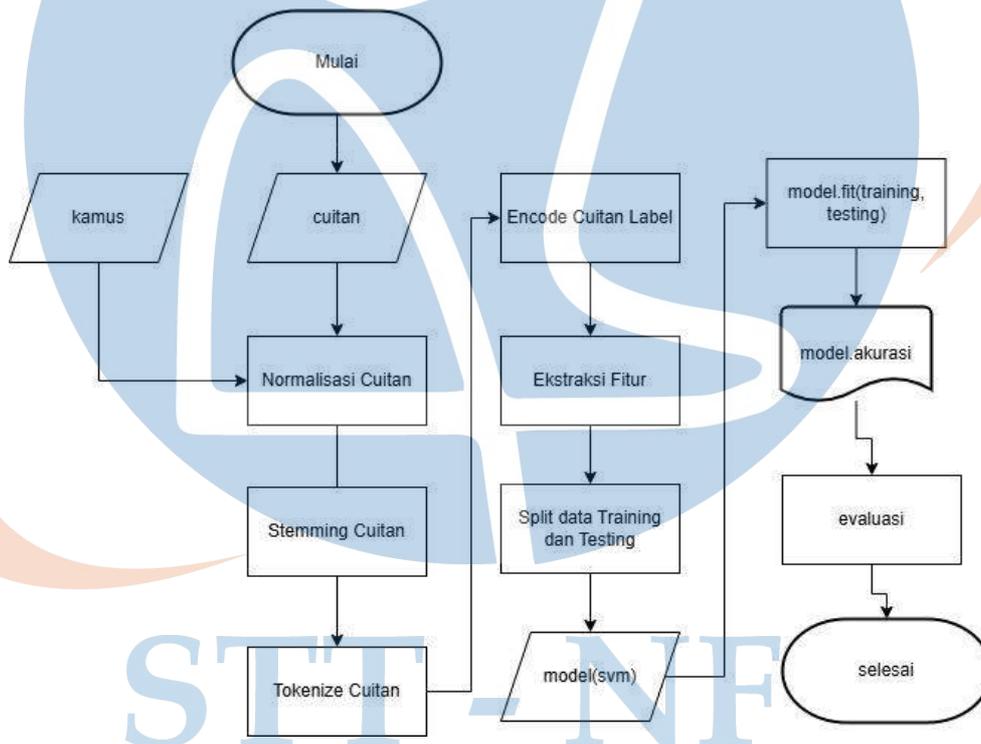
1. *Software* untuk mengelola data, yaitu Google Colab
2. Modul sistem pengelolaan data
  - a. Modul Pandas untuk melakukan proses *filter* dan memilih data berdasarkan kondisi tertentu.
  - b. Modul NLTK untuk melakukan proses klasifikasi pada data, seperti *tokenize*, *stemming*, dan *stopword*.
  - c. Modul Sastrawi untuk melakukan proses *stemming* dengan dasar bahasa Indonesia

- d. Modul *Sklearn* untuk melakukan proses implementasi pada model *machine learning*.
- e. Modul *matplotlib* untuk melakukan proses *visualisasi data*.

#### 4.2 Perancangan Sistem

Pada tahapan ini membahas tentang perancangan sistem yang digunakan pada analisis sentimen pada komentar di X menggunakan metode *Support Vector Machine*.

Penulis melakukan perancangan sistem menggunakan bagan alir untuk mengetahui langkah yang tepat untuk menganalisa sistem yang akan digunakan.



**Gambar 4. 1** Perancangan Sistem

Tahapan analisa sistem yang dilakukan penulis sebagai berikut :

1. Input cuitan dan kamus bahasa pada sistem yang penulis buat langkah pertama penulis memasukkan data komentar yang sudah dikumpulkan yang diproses oleh sistem beserta dengan data kamus.

2. Komentar selanjutnya dilakukan proses normalisasi, dibersihkan setiap kata yang di dalamnya terdapat kata singkatan, seperti “yg” = “yang”, “jl” = “jalan” berdasarkan dengan kamus yang sudah di *input* kedalam sistem.
3. Komentar yang sudah dinormalisasi berdasarkan dari singkatan-singkatan, dilakukan proses stemming untuk membersihkan kata-kata yang memiliki imbuhan, seperti “dilandjutkannya” = “lanjut”, “sepertinya” = “seperti”, “rasain” = “rasa”.
4. Data komentar yang memiliki label, pada tahap selanjutnya label diubah dengan metode encode menjadi nilai *numeric*. Ini dibutuhkan agar komentar dan labelnya dapat dilakukan analisis menggunakan model yang sudah ditentukan.
5. Data komentar dilakukan ekstraksi fitur menggunakan metode *TF-IDF* dan *LSA*
6. Data komentar yang sudah diekstraksi dan label dipisah menjadi nilai x untuk komentar dan nilai y untuk label.
7. Sistem meng*import* model untuk penerapan terhadap data nilai x dan nilai y. Model yang di *import* adalah *Support Vector Machine (SVM)*.
8. Model melakukan proses terhadap nilai x dan nilai y, yang sebelumnya sudah dilakukan berbagai macam proses pembersihan data. Untuk mendapatkan nilai akurasi nilai x dan nilai y diimplementasikan ke dalam model SVM.
9. Setelah melakukan proses implementasi data kepada model SVM didapatkannya nilai akurasi yang dibutuhkan. Nilai akurasi dibutuhkan untuk mengetahui kesesuaian prediksi terhadap nilai x dan nilai y.

Langkah ini dilakukan penulis supaya perancangan dilakukan dengan baik dan dengan mengurangi sedikit kesalahan yang dapat terjadi.

### **4.3 Proses Implementasi Sistem**

Pada tahap ini proses sistem diimplementasikan untuk mengolah data berdasarkan rancangan yang sudah dibuat pada tahap perencanaan sistem.

### 4.3.1 Import Module

Tahapan selanjutnya lakukan import modul yang sudah ditentukan pada tahap perencanaan sistem dengan menggunakan perintah berikut:

```
import pandas as pd
from nltk.stem import PorterStemmer
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from sklearn.preprocessing import LabelEncoder
from sklearn.feature_extraction.text import CountVectorizer,
TfidfTransformer, TfidfVectorizer
from sklearn.metrics import accuracy_score,
classification_report
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
import matplotlib.pyplot as plt
!pip install PySastrawi
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
```

Pada sintaks diatas ini terdapat modul beberapa modul yang diimpor, seperti:

- NLTK sebagai *preprocessing* data
- Sklearn* digunakan untuk melakukan *training* dan *testing*
- Sastrawi* digunakan untuk mengekstraksi kata menjadi kata dasarnya berdasarkan bahasa Indonesia.

### 4.3.2 Pemrosesan dataset

Pada tahapan ini dataset dan kamus yang sudah ada di *import* ke dalam notebook menggunakan modul pandas.

```
twit_emotion = pd.read_csv("Twitter_Emotion_Dataset.csv")
kam_sing = pd.read_csv("kamus_singkatan.csv")
```

Potongan sintaks diatas digunakan untuk mengimpor dataset csv yang digunakan sebagai data uji. Variabel 'twit\_emotion' digunakan untuk menyimpan dataset uji, sedangkan variabel 'kam\_sing' digunakan untuk menyimpan database kamus.

Untuk melihat isi dari dataset yang sudah diimpor dapat menggunakan kode berikut:

```
twit_emotion.head()
```

	label	tweet
0	anger	Soal jln Jatibaru,polisi tdk bs GERTAK gubernu...
1	anger	Sesama cewe lho (kayaknya), harusnya bisa lebi...
2	happy	Kepingin gudeg mbarek Bu hj. Amad Foto dari go...
3	anger	Jln Jatibaru,bagian dari wilayah Tn Abang.Peng...
4	happy	Sharing pengalaman aja, kemarin jam 18.00 bata...

**Gambar 4. 2** Sampel Hasil Cuitan

Setelah dataset di *import*, dataset perlu dilakukan proses pencocokan data secara manual terhadap dataset cuitan dengan dataset kamus, dikarenakan masih ada dataset kamus yang tidak sesuai dengan format karena memiliki spasi di akhir kata maka dilakukan proses pembersihan seperti berikut:

```
fixed_word = kam_sing.copy()
col =
pd.DataFrame({kam_sing.columns[0]:[kam_sing.columns[0]})
fixed_word = pd.concat([col, fixed_word], ignore_index=True)
j = fixed_word['aamiin;amin ']
for x in range(len(fixed_word["aamiin;amin "])):
    fixed_word['aamiin;amin ')[x] = j[x].split(';')
fixed_word =
fixed_word.rename(columns={fixed_word.columns[0]: "kata"})
for y in range(len(fixed_word["kata"])):
    if fixed_word["kata"][y][1][-1] == " ":
        fixed_word["kata"][y][1] = fixed_word["kata"][y][1][:-1]
```

Potongan sintaks diatas digunakan untuk normalisasi database kamus singkatan.

### 4.3.3 Preprocessing Data

Pada tahapan ini data yang sudah di *import* akan dilakukan *preprocessing data* seperti *cleansing data*, *tokenize*, *stemming*, dan seleksi fitur. Adapun tahapan yang dilakukan sebagai berikut:

1. Membersihkan dan normalisasi dataset dari huruf yang disingkat agar data memiliki kesesuaian, untuk membersihkannya dilakukan secara manual berdasarkan dataset cuitan dan dataset kamus

```

df = twit_emotion.copy()
for y in fixed_word["kata"]:
    df['tweet'] = df['tweet'].str.replace(f" {y[0]} ", f"
{y[1]} ")
    df['tweet'] = df['tweet'].str.replace("USERNAME", "")
    df['tweet'] = df['tweet'].str.replace("URL", "")

```

Potongan sintaks diatas digunakan untuk memproses normalisasi atau menyesuaikan nilai-nilai dataset uji, menggunakan metode perulangan

2. Setelah melakukan pembersihan data dilakukan *stemming* untuk melakukan proses pengambilan kata dasar dari kata-kata yang memiliki imbuhan menggunakan modul sastrawi.

```

df_stem = df.copy()
panjang = len(df_stem['tweet'])
factory = StemmerFactory()
stemmer = factory.create_stemmer()
for x in range(panjang):
    kalimat = df_stem['tweet'][x]
    kalimat_stem = stemmer.stem(kalimat)
    df_stem['tweet'][x] = kalimat_stem

```

Potongan sintaks diatas digunakan untuk memproses *stemming* cuitan menggunakan modul Sastrawi.

3. Setelah dataset dilakukan normalisasi, serta *stemming* kata dasar dari imbuhan, dan merubah kata-kata yang disingkat menjadi kata aslinya, lalu dilakukan proses *tokenize* untuk memotong kalimat dan menghilangkan kata-kata umum berdasarkan *stopwords*.

```

# Melakukan seleksi dengan Stop Words
df_filter = []
stopWords = stopwords.words('indonesian')
for x in df_stem['tweet']:
    tokens = word_tokenize(x)
    filtered_tokens = [word for word in tokens if word.lower()
not in stopWords]
    if len(filtered_tokens) == 0:
        df_filter.append(tokens)

```

```

else:
    df_filter.append(filtered_tokens)
# Melakukan merging pada dokumen tweet
documents_merged = [' '.join(doc) for doc in df_filter]

```

4. Tahap selanjutnya setelah melakukan *stemming* maka dilakukan ekstraksi fitur menggunakan *TF-IDF* untuk menentukan frekuensi matrix dari dataset cuitan.

```

tfidf_vector = TfidfVectorizer()
tfidf_matrix = tfidf_vector.fit_transform(df_stem['tweet'])

```

Pada potongan sintaks diatas, untuk mengekstraksi fitur penulis menggunakan modul *TF-IDF* melalui *sklearn* yang sudah diimpor pada sesi awal.

5. Setelah melakukan ekstraksi fitur menggunakan *TF-IDF* buat kembali ekstraksi fitur menggunakan *LSA* seperti pada potongan sintaks berikut:

```

vectorizer = CountVectorizer()
latent_semantic = vectorizer.fit_transform(df_stem['tweet'])
n_components = 100 # Jumlah komponen (dimensi) yang diinginkan
svd = TruncatedSVD(n_components=n_components)
latent_semantic = svd.fit_transform(latent_semantic)

```

Sintaks mengimport modul *vectorize* untuk melakukan vectorisasi terhadap data cuitan yang disimpan pada variabel `df_stem['tweet']` selanjutnya data yang disimpan pada variabel dilakukan ekstraksi fitur menggunakan *LSA*

6. Setelah dataset cuitan diubah menjadi matrix dengan metode *TF-IDF* dan *LSA* langkah selanjut merubah label menjadi angka dengan metode *Encode* menggunakan *mapping* pada python.

```

df_stem.label.map({"anger":0, "fear":1, "love":2,
                  "sadness":3, "happy":4})

```

## 4.4 Pengujian Sistem

Pada tahapan ini sistem akan menguji data yang sudah diproses dengan menggunakan metode machine learning yang sudah ditentukan, berikut tahapan yang dilakukan:

### 4.4.1 Pengujian Model *Machine Learning*

Pada tahap ini penulis menguji dataset yang sudah dilakukan *preprocessing* data dan ekstraksi fitur menggunakan *TF-IDF* untuk diuji oleh *machine learning SVM* berikut langkah yang dilakukan:

1. Memisahkan dataset cuitan dengan label menjadi nilai x dan y, di mana nilai x berisi cuitan dan nilai y berisi label, lalu nilai x dan y diubah menjadi data *training* dan *testing* sebagai bahan uji. Data yang dipisah memiliki parameter *testing* sebesar 20% dari data yang digunakan dan 80% *training* data agar *machine learning* dapat melakukan lebih banyak pelatihan terhadap data.

```
X = tfidf_matrix
y = df_stem.label
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42) # Pemisahan data terhadap
data x & y training dan test
```

2. Tahap selanjutnya dilakukan implementasi data *training* kedalam model *machine learning* untuk kemudian hasilnya dapat dievaluasi

```
# Pelatihan model SVC
model = SVC(kernel = 'linear') # Pengambilan model
model.fit(X_train, y_train) # Implementasi model terhadap
data training
```

Potongan sintaks diatas model SVC menggunakan kernel *linear* untuk melakukan uji terhadap dataset yang sudah dipisah.

3. Setelah dilakukan implementasi data *training* terhadap model *machine learning*, selanjutnya dilakukan evaluasi model terhadap data *testing* untuk mendapatkan nilai akurasi model *machine learning* untuk mendeteksi emosi pada komentar

```
# Evaluasi model
y_pred = model.predict(X_test)
print("Accuracy:", accuracy_score(y_test, y_pred))
print("Classification Report:")
print(classification_report(y_test, y_pred))
```

Pada potongan sintaks diatas, hasil dari model SVC dilakukan evaluasi dan perbandingan terhadap data *training* dan *testing*.

Selanjutnya lakukan pengujian fitur yang diekstraksi menggunakan *LSA* untuk diuji oleh *machine learning* SVM:

1. Memisahkan dataset dengan label menjadi nilai x dan y

```
X = latent_semantic
y = df_stem.label
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42) # Pemisahan data terhadap
data x & y training dan test
```

2. Tahap selanjutnya dilakukan implementasi data *training* kedalam model *machine learning* untuk kemudian hasilnya dapat dievaluasi

```
# Pelatihan model SVC
model = SVC(kernel = 'linear') # Pengambilan model
model.fit(X_train, y_train) # Implementasi model terhadap
data training
```

Potongan sintaks diatas model SVC menggunakan kernel *linear* untuk melakukan uji terhadap dataset yang sudah dipisah.

3. Setelah dilakukan implementasi data *training* terhadap model *machine learning*, selanjutnya dilakukan evaluasi model terhadap data *testing* untuk mendapatkan nilai akurasi model *machine learning* untuk mendeteksi emosi pada komentar

```
# Evaluasi model
y_pred = model.predict(X_test)
print("Accuracy:", accuracy_score(y_test, y_pred))
print("Classification Report:")
print(classification_report(y_test, y_pred))
```

Setelah melakukan evaluasi model *machine learning* terhadap dataset didapatkan hasil evaluasi sebagai berikut:

### 1. Hasil evaluasi SVM menggunakan fitur *TF-IDF*

Accuracy: 0.659477866061294

Classification Report:

	precision	recall	f1-score	support
anger	0.65	0.77	0.71	229
fear	0.90	0.61	0.73	119
happy	0.64	0.63	0.64	214
love	0.77	0.73	0.75	119
sadness	0.53	0.55	0.54	200
accuracy			0.66	881
macro avg	0.70	0.66	0.67	881
weighted avg	0.67	0.66	0.66	881

### 2. Hasil evaluasi menggunakan *LSA*

Accuracy: 0.5732122587968218

Classification Report:

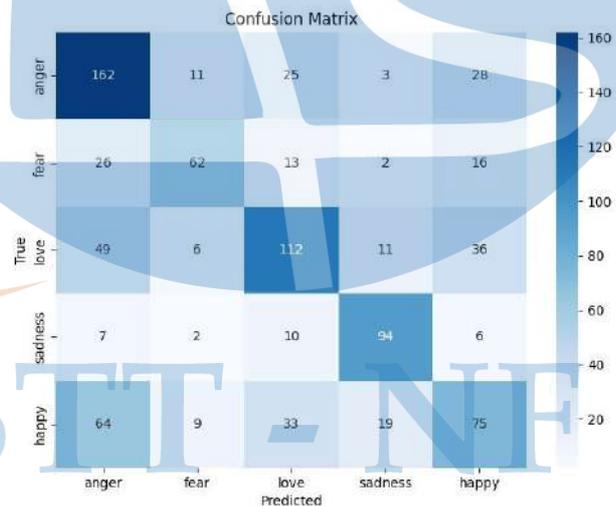
	precision	recall	f1-score	support
anger	0.53	0.71	0.60	229
fear	0.69	0.52	0.59	119
happy	0.58	0.52	0.55	214
love	0.73	0.79	0.76	119
sadness	0.47	0.38	0.42	200
accuracy			0.57	881
macro avg	0.60	0.58	0.58	881
weighted avg	0.57	0.57	0.57	881

Setelah mendapatkan hasil akurasi berdasarkan evaluasi model *machine learning*, hasil evaluasi juga dapat dilihat *confusion matrix*-nya berikut hasil *confusion matrix* dari evaluasi metode *machine learning* menggunakan fitur *TF-IDF*:



**Gambar 4.3** *Confusion Matrix* Analisis Sentimen menggunakan *TF-IDF*

Lalu hasil *confusion matrix* dari evaluasi metode *machine learning* menggunakan fitur *LSA*:



**Gambar 4.4** *Confusion Matrix* Analisis Sentimen menggunakan *LSA*

Dari hasil pengujian analisis sentimen ini didapatkan akurasi sebesar 66% jika menggunakan fitur *TF-IDF* dan akurasi sebesar 57% jika menggunakan fitur *LSA* berdasarkan *classification report* dari pengujian yang dilakukan. Selanjutnya dari hasil pengujian didapatkan pada *confusion matrix*, yaitu 881 kasus pengujian





## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan rumusan masalah dan pengujian yang dilakukan pada penelitian “Analisis Sentimen pada Komentar di Platform X metode *Support Vector Machine*” dapat diperoleh kesimpulan sebagai berikut:

1. Metode *Support Vector Machine* dapat mendeteksi emosi pada cuitan dengan hasil akurasi sebesar 66% dan 57% menggunakan fitur *TF-IDF* dan *LSA* dengan emosi yang dianalisis yaitu marah, takut, cinta, sedih, dan senang.
2. Penelitian ini menunjukkan bahwa langkah penulisan pemrosesan data dalam program yang ada saat ini masih belum mencapai tingkat efisiensi yang optimal. Ada proses normalisasi yang dilakukan lebih lama dibandingkan proses lainnya, seperti implementasi. Untuk meningkatkan efektivitas penelitian, sangat penting untuk mengeksplorasi dan mengadopsi lebih banyak metode pemrosesan dataset yang lebih canggih.
3. Hasil evaluasi mengalami perbedaan akurasi sebesar 9% menggunakan ekstraksi fitur yang berbeda
4. Berdasarkan hasil penelitian yang dilakukan penerapan ekstraksi fitur menggunakan *TF-IDF* lebih unggul dibandingkan menggunakan *LSA* berdasarkan hasil akurasinya.

#### 5.2 Saran

Pada penelitian ini memiliki keterbatasan dan potensi peningkatan. Oleh karena itu, penulis menyarankan pengembangan supaya penelitian selanjutnya menjadi lebih baik, pada penelitian ini penulis hanya meneliti bahasa indonesia dengan menggunakan menggunakan metode *Support Vector Machine* disarankan untuk mengeksplorasi dengan berbagai macam jenis metode seperti *Naive Bayes*, *K-Nearest Neighbors*, *Artificial Neural Networks*, dan metode *machine learning* lainnya.

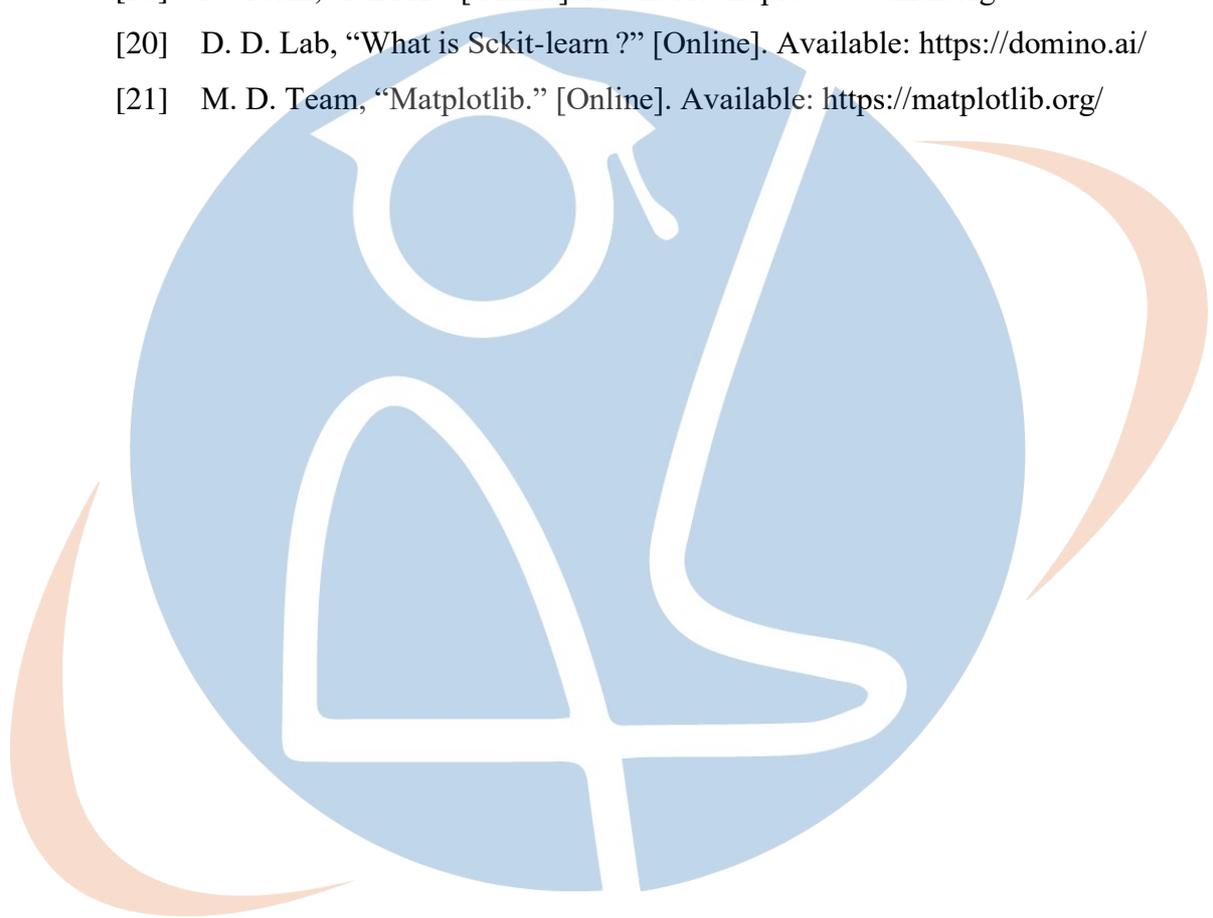
## DAFTAR PUSTAKA

- [1] F. Mulyono, “Dampak Media Sosial bagi Remaja,” *J. Simki Econ.*, vol. 4, no. 1, pp. 57–65, 2021, doi: 10.29407/jse.v4i1.66.
- [2] B. Kurniawan, A. Ari Aldino, and A. Rahman Isnain, “SENTIMEN ANALISIS TERHADAP KEBIJAKAN PENYELENGGARA SISTEM ELEKTRONIK (PSE) MENGGUNAKAN ALGORITMA BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS (BERT),” 2022. [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [3] A. D. Haspari, T. A. Maulana, and C. R. Yuningsih, “Lima emosi dasar manusia melalui seni new media,” vol. 11, no. 2, pp. 3662–3673, 2024.
- [4] A. Rahman Isnain, A. Indra Sakti, D. Alita, and N. Satya Marga, “SENTIMEN ANALISIS PUBLIK TERHADAP KEBIJAKAN LOCKDOWN PEMERINTAH JAKARTA MENGGUNAKAN ALGORITMA SVM,” *JDMSI*, vol. 2, no. 1, pp. 31–37, 2021, [Online]. Available: <https://t.co/NfhmfMjtXw>
- [5] Y. Amalia, N. Jannah, and R. B. Prasetyo, “Analisis Sentimen dan Emosi Publik pada Awal Pandemi COVID-19 Berdasarkan Data Twitter dengan Pendekatan Berbasis Leksikon (Analysis of Public Sentiment and Emotion at the Beginning of COVID-19 Pandemic Based on Twitter Data with Lexicon Based Approach).”
- [6] R. W. Pratiwi, S. F. H. D. Dairoh, D. I. Af'idah, Q. R. A, and A. G. F, “SENTIMEN Pada Review Skincare Female Daily Menggunakan Metode Support Vector Machine (SVM),” *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 4, no. 1, pp. 40–46, 2021, doi: 10.20895/inista.v4i1.387.
- [7] V. Kevin, S. Que, : Analisis, S. Transportasi, A. Iriani, and H. D. Purnomo, “Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization (Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimization),” 2020. [Online]. Available: [www.tripadvisor.com](http://www.tripadvisor.com),
- [8] A. A. Kasim and M. Sudarsono, “Algoritma Support Vector Machine

- (SVM) untuk Klasifikasi Ekonomi Penduduk Penerima Bantuan Pemerintah di Kecamatan Simpang Raya Sulawesi Tengah,” *Semin. Nas. APTIKOM*, pp. 568–573, 2019.
- [9] Ichsan Nur Irmasyah Nurul Chafid, Luqman Mujianto, “Penerapan Filter Kata Menggunakan Metode Stemming Pada Aplikasi Chatting Berbasis Web,” vol. 1, no. 1, pp. 1–9, 2020.
- [10] NLTK Team, “NLTK Tokenize Package.” [Online]. Available: <https://www.nltk.org/>
- [11] R. Arthana, “Mengenal Accuracy, Precision, Recall dan Specificity serta yang diprioritaskan dalam Machine Learning.” [Online]. Available: <https://medium.com/>
- [12] D. P. Wijaya, L. D. Murti, and M. R. Rachman, “Recall dan Precision pada Online Public Access Catalog (OPAC) Dinas Arsip dan Perpustakaan Kota Bandung,” *VISI PUSTAKA Bul. Jar. Inf. Antar Perpust.*, vol. 24, no. 1, pp. 81–91, 2022, doi: 10.37014/visipustaka.v24i1.2915.
- [13] Noermanzah, “Bahasa sebagai Alat Komunikasi, Citra Pikiran, dan Kepribadian,” *Pros. Semin. Nas. Bulan Bhs.*, pp. 306–319, 2019, [Online]. Available: <https://ejournal.unib.ac.id/index.php/semiba>
- [14] T. Curlin, B. Jaković, and I. Miloloža, “Twitter usage in Tourism: Literature Review,” *Bus. Syst. Res.*, vol. 10, no. 1, pp. 102–119, 2019, doi: 10.2478/bsrj-2019-0008.
- [15] K. Tri Putra, M. Amin Hariyadi, and C. Crysdiyan, “Perbandingan Feature Extraction Tf-Idf Dan Bow Untuk Analisis Sentimen Berbasis Svm,” *J. Cahaya MAndalika*, p. 1449, 2023.
- [16] M. H. Mahendra, D. T. Murdiansyah, and K. M. Lhaksmana, “Analisis Sentimen Tweet COVID-19 Menggunakan Metode K-Nearest Neighbors dengan Ekstraksi Fitur TF-IDF dan CountVectorizer,” *Dike J. Ilmu Multidisiplin*, vol. 1, no. 2, pp. 37–43, 2023.
- [17] M. S. Wahyuni, D. Setiawan, and T. Syahputra, “Sistem Temu Kembali Informasi Dengan Latent Semantic Analysis Pada Kesamaan Tugas Akhir Mahasiswa,” *J. Tek.*, vol. 1, no. 2, p. 46, 2021, doi:

10.54314/teknisi.v1i2.655.

- [18] K. Chromiński, E. Benko, Z. J. Hernández-Figueroa, J. D. González-Domínguez, and J. C. Rodríguez-del-Pino, “Python Fundamentals,” *Python Fundam.*, no. c, 2021, doi: 10.17846/fpvai-2021-14.
- [19] N. Team, “NLTK.” [Online]. Available: <https://www.nltk.org/>
- [20] D. D. Lab, “What is Sckit-learn ?” [Online]. Available: <https://domino.ai/>
- [21] M. D. Team, “Matplotlib.” [Online]. Available: <https://matplotlib.org/>



STT - NF