

**ANALISIS SENTIMEN HATESPEECH TERHADAP GACOAN
PADA PLATFORM X DENGAN MENGGUNAKAN SUPPORT
VECTOR MACHINE (SVM)**

SKRIPSI

DISUSUN OLEH

ALIFA AZ ZAHRA FAHREZA

NPM. 2109010006



UMSU

Unggul | Cerdas | Terpercaya

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
2025**

**ANALISIS SENTIMEN HATESPEECH TERHADAP GACOAN
PADA PLATFORM X DENGAN MENGGUNAKAN SUPPORT
VECTOR MACHINE (SVM)**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
(S.Kom) dalam Program Studi Sistem Informasi pada Fakultas Ilmu Komputer
dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara**

**ALIFA AZ ZAHRA FAHREZA
NPM. ISI 2109010006**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
2025**

LEMBAR PENGESAHAN

Judul Skripsi : ANALISIS SENTIMEN HATESPEECH TERHADAP
GACUAN PADA PLATFORM X DENGAN
MENGUNAKAN SUPPORT VECTOR MACHINE
(SVM)

Nama Mahasiswa : ALIFA AZ ZAHRA FAHREZA

NPM : 2109010006

Program Studi : SISTEM INFORMASI

Menyetujui
Komisi Pembimbing



(Dr. Firahmi Rizky, S.Kom, M.Kom)
NIDN. 3048770671230343

Ketua Program Studi



(Dr. Firahmi Rizky, S.Kom, M.Kom)
NIDN. 3048770671230343

Dekan



(Dr. Al-Khowarizmi, S.Kom., M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

ANALISIS SENTIMEN HATESPEECH TERHADAP GACOAN PADA PLATFORM X DENGAN MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM)

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, 8 Juli 2025

Yang membuat pernyataan



Alifa Az Zahra Fahreza

NPM. 2109010006

**PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama : Alifa Az Zahra Fahreza
NPM : 2109010006
Program Studi : Sistem Informasi
Karya Ilmiah : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

**ANALISIS SENTIMEN HATESPEECH TERHADAP GACOAN PADA
PLATFORM X DENGAN MENGGUNAKAN SUPPORT VECTOR
MACHINE (SVM)**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, 8 Juli 2025

Yang membuat pernyataan



Alifa Az Zahra Fahreza

NPM. 2109010006

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : Alifa Az Zahra Fahreza
Tempat dan Tanggal Lahir : Medan, 08 Juli 2002
Alamat Rumah : Jl. Sei Asahan no.76 A
Telepon/Faks/HP : 081371795502
E-mail : alifazzhr.tata@gmail.com
Instansi Tempat Kerja : -
Alamat Kantor : -

DATA PENDIDIKAN

SD : SDIT Al – Fityan School Medan TAMAT: 2014
SMP : SMP Swasta Adabiah Padang TAMAT: 2017
SMA : SMA Daarul Quran Cikarang TAMAT: 2021

KATA PENGANTAR



Puji dan Syukur, Alhamdulillah penulis panjatkan kehadiran Allah SWT atas segala rahmat, nikmat, dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan judul **“ANALISIS SENTIMEN HATESPEECH TERHADAP GACOAN PADA PLATFORM X DENGAN MENGGUNAKAN SUPPORT VECTOR MACHINE”** sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi di Universitas Muhammadiyah Sumatera Utara.

Penyusunan skripsi ini tentu bukanlah hal yang mudah dan tidak terlepas dari berbagai tantangan dan hambatan. Namun berkat bimbingan, dukungan, doa serta semangat yang ada, penulis dapat melalui semua proses dengan baik hingga akhirnya skripsi ini dapat terselesaikan. Dengan segala kerendahan hati, Penulis tentunya berterima kasih kepada berbagai pihak dalam dukungan serta doa dalam penyelesaian skripsi. Penulis juga mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Agussani, M.AP., Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Ibu Dr. Firaahmi Rizky, S.Kom, M.Kom Ketua Program Studi Sistem Informasi Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
4. Ibu Yoshida Sari, S.E., S.Kom., M.Kom. Sekretaris Program Studi Sistem Informasi Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
5. Ibu Dr. Firaahmi Rizky, M.Kom, selaku Dosen Pembimbing yang telah dengan sabar membimbing dan memberikan arahan selama penyusunan skripsi ini.

6. Teruntuk yang Tersayang dan Teristimewa kepada papa dan almh. mama, penulis ucapkan terimakasih yang tak terhingga karena telah berjuang untuk memberikan kehidupan dan pendidikan yang baik sedari kecil hingga penulis mampu menyelesaikan studi di titik akhir ini. Yang tidak kenal lelah untuk mendoakan, memberikan semangat serta kasih sayang dan cintanya yang tulus demi kesuksesan saya.
7. Kepada keluarga besar saya baik dipihak mama maupun pihak papa saya, terimakasih telah ada dan memberikan semangat serta mengkhawatirkan kondisi saya selama saya menulis penelitian saya hingga akhir. Yang telah menemani dan membantu saya untuk mendukung dan mengerjakan skripsi ini hingga akhirnya saya sudah berada di tahap ini.
8. Thanks to adik – adik dan kakak – kakak ku tercinta yang telah berkorban banyak demi memberikan dukungan dan semangat untuk dapat membantu menyelesaikan skripsi saya, terkhusus adik Habibie dan Jibril terimakasih kepada kalian sebanyak – banyaknya.
9. Thanks to Muhammad Fityatul Haqqi telah meberikan banyak waktu dan supportnya, yang telah menjadi penyemangat disaat penulis melakukan proses penelitian serta mendengarkan semua keluh kesah hingga dipenghujung semester akhir ini.
10. Thanks to teman – teman seperjuangan skripsi ku Agita dan Putri yang telah banyak membantu selama kita berdiskusi dan memberikan banyak masukan untuk berjuang bersama demi menyelesaikan studi ini dengan sempurna.
11. Thanks to teman – teman KKN saya yang telah memberikan banyak dukungan dan kenangannya, terimakasih atas kerja samanya yang pada akhirnya kita bisa dapat sampai di tahap ini.
12. Semua pihak yang terlibat langsung ataupun tidak langsung yang tidak dapat penulis ucapkan satu-persatu yang telah membantu penyelesaian skripsi ini.

ANALISIS SENTIMEN HATE SPEECH TERHADAP GACOAN PADA PLATFORM X DENGAN MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM)

ABSTRAK

Perkembangan media sosial saat ini memunculkan berbagai bentuk interaksi digital salah satunya adalah *hatespeech* atau biasa disebut ujaran kebencian yang semakin marak terjadi. Platform X yang sebelumnya dikenal sebagai Twitter menjadi salah satu media yang paling banyak digunakan masyarakat dalam menyampaikan opini secara terbuka, khususnya terhadap industri bisnis seperti restoran siap saji Gacoan. Penelitian ini bertujuan untuk melakukan analisis sentimen yang mengandung ujaran kebencian terhadap Gacoan di Platform X dengan menerapkan algoritma Support Vector Machine (SVM) sebagai metode klasifikasi. Data dikumpulkan melalui proses scrapping menggunakan tools tweet-harvest dalam rentang waktu mulai dari Januari hingga Juni 2024 dengan total 2500 data Tweet. Selanjutnya dilakukan tahap preprocessing seperti case folding, cleansing, tokenizing, normalisasi, stopword removal dan stemming. Pelabelan data menggunakan pendekatan INSET Lexicon-Based untuk mengklasifikasikan tweet menjadi *hatespeech* dan *non-hatespeech*. Hasil pengujian model SVM dengan kernel linear menunjukkan tingkat akurasi sebesar 83,3% dengan nilai precision 0,899, recall 0,907 dan f1-score 0,903 pada kategori *hatespeech*. Dari hasil ini membuktikan bahwa algoritma SVM mampu secara efektif mengklasifikasikan ujaran kebencian pada data tidak terstruktur di media sosial.

Kata Kunci: Analisis Sentimen, Hate Speech, Platform X, Support Vector Machine (SVM), Gacoan

HATESPEECH SENTIMENT ANALYSIS ON GACOAN THROUGH PLATFORM X USING SUPPORT VECTOR MACHINE (SVM)

ABSTRACT

The rapid development of social media has led to various forms of digital interaction, one of which is hate speech that continues to increase. Platform X, formerly known as Twitter, has become one of the most widely used platforms for the public to express their opinions openly, particularly regarding businesses such as the fast-food restaurant Gacoan. This research aims to conduct sentiment analysis on hate speech directed at Gacoan on Platform X by applying the Support Vector Machine (SVM) algorithm as the classification method. The data were collected through a scraping process using the Tweet-Harvest tool, covering the period from January to June 2024, with a total of 2,500 tweets. The preprocessing steps included case folding, cleansing, tokenizing, normalization, stopword removal, and stemming. Data labeling was carried out using the INSET Lexicon-Based approach to classify the tweets into hatespeech and non-hatespeech categories. The SVM model with a linear kernel achieved an accuracy of 83.3%, with a precision of 0.899, recall of 0.907, and an F1-score of 0.903 for the hatespeech category. These results demonstrate that the SVM algorithm is effective in classifying hate speech in unstructured social media data.

Keywords: Sentiment Analysis, Hate Speech, Platform X, Support Vector Machine (SVM), Gacoan

DAFTAR ISI

LEMBAR PENGESAHAN	ii
PERNYATAAN ORISINALITAS	iii
KATA PENGANTAR	vi
BAB I.....	1
PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah	3
1.4. Tujuan Penelitian	4
1.5. Manfaat Penelitian	4
BAB II.....	6
LANDASAN TEORI.....	6
2.1 Analisis dan Hate Speech	6
2.1.1 Kategori / Kelas Sentimen	7
2.1.2 Hate Speech	8
2.1.3 Dampak Hatespeech dan Regulasinya	9
2.2 Rebranding Twitter to Platform X	10
2.3 Support Vector Machine (SVM)	11
2.3.1 Cara Kerja SVM	12
2.4 Python	16
2.5 Gacoan	17
2.6 Penelitian Terdahulu	18
BAB III	21
METODOLOGI PENELITIAN.....	21
3.1. Pendekatan Penelitian	21
3.2. Teknik Pengumpulan Data	22
3.2.1 Studi Pustaka	22
3.2.2 Scrapping Data	23
3.3. Metodologi Pengembangan Sistem	24
3.3.1 Preprocessing Data	24
3.3.2 Labelling Data	26
3.3.3 Split Data dan Vektorisasi	27
3.3.4 Implementasi Support Vector Machine	28
3.3.5 Evaluasi Model (Akurasi, Presisi, dan Recall)	29
3.4 Perangkat Penelitian	30

3.5	Waktu dan Tempat Penelitian	31
3.5.1	Tempat Penelitian	31
BAB IV	32
HASIL DAN PEMBAHASAN	32
4.1.	Hasil dan Proses Pengumpulan Data Tweet	32
4.2.	PreProcessing Data	33
4.2.1.	Case Folding	34
4.2.2.	Cleaning	34
4.2.3.	Tokenizing	36
4.2.4.	Normalisasi	36
4.2.5.	Stopword Removal	37
4.2.6.	Stemming	38
4.3.	Labelling Metode Inset Lexicon Based	39
4.4.	Split Data	41
4.5.	Implementasi Model Support Vector Machine	44
4.5.1.	Pemilihan Parameter Optimal dengan GridSearchCV	44
4.5.2.	Latih Model SVM	45
4.6.	Evaluasi Model SVM	46
4.6.1.	Confusion Matrix	46
4.6.2.	Perhitungan Accuracy, Precision, Recall dan F1-Score	47
4.6.3.	Classification Report	49
4.6.4.	ROC Curve SVM	50
4.6.5.	WordCloud	52
BAB V	55
PENUTUP	55
5.1.	Kesimpulan	55
5.2.	Saran	55
DAFTAR PUSTAKA	57
LAMPIRAN	62

DAFTAR TABEL

Tabel 2. 1 Tabel Penelitian Terdahulu	18
Tabel 3. 1 Perangkat Keras	30
Tabel 3. 2 Perangkat Lunak	31
Tabel 4. 1 Hasil case_folding.....	34
Tabel 4. 2 Hasil Cleaning.....	35
Tabel 4. 3 Hasil Tokenisasi.....	36
Tabel 4. 4 Hasil Normalized	37
Tabel 4. 5 Hasil Stopwords.....	38
Tabel 4. 6 Hasil Stemming.....	39
Tabel 4. 7 Tabel Perhitungan Manual Confusion Matrix	47

DAFTAR GAMBAR

Gambar 2. 1 Konsep hyperplane pada SVM Linier.....	14
Gambar 2. 2 Konsep Hyperplane pada SVM Non-Linier.....	14
Gambar 3. 1 Kerangka Penelitian	22
Gambar 3. 2 Scrapping Tweet-harvest.....	24
Gambar 3. 3 Kode train_test_split	27
Gambar 3. 4 Vectorizer Code	28
Gambar 4. 1 Install Package Scrapping.....	32
Gambar 4. 2 Scrapping.....	33
Gambar 4. 3 Hasil Scrapping.....	33
Gambar 4. 4 Case_folding.....	34
Gambar 4. 5 Cleansing	35
Gambar 4. 6 Tokenisasi.....	36
Gambar 4. 7 Tokenisasi.....	37
Gambar 4. 8 Stopword Removal	38
Gambar 4. 9 Stemming.....	39
Gambar 4. 10 Kamus Lexicon	40
Gambar 4. 11 Fungsi Penentuan Skor	40
Gambar 4. 12 Bar Plot Distribusi Sentimen	41
Gambar 4. 13 Split Data.....	42
Gambar 4. 14 Bar Plot Split Data.....	42
Gambar 4. 15 Vektorisasi Teks.....	43
Gambar 4. 16 GridSearchCV	44
Gambar 4. 17 Latih Model SVM.....	45
Gambar 4. 18 Evaluasi Model.....	46
Gambar 4. 19 Confusion Matrix Linear	47
Gambar 4. 20 Hasil Classification Report	49
Gambar 4. 21 ROC Curve SVM.....	51
Gambar 4. 22 Prediksi Hatespeech	52
Gambar 4. 23 Prediksi Non-Hatespeech	52
Gambar 4. 24 WOrd Cloud Hatespeech	53
Gambar 4. 25 Word Cloud Non-Hatespeech.....	54

BAB I PENDAHULUAN

1.1. Latar Belakang Masalah

Membahas tentang Hate Speech atau yang biasa disebut dengan Ujaran kebencian masih menjadi suatu hal yang sangat diremehkan masyarakat terhadap pengaruh dan dampaknya. Istilah Hate Speech sendiri bermakna “ekspresi yang mengandung hasutan untuk merugikan berdasarkan target yang diidentifikasi dengan kelompok sosial ataupun demografis tertentu”. Adanya Ujaran Kebencian seperti itu, telah menggeser pola berkomunikasi masyarakat, berawal dari komunikasi di dunia nyata menjadi komunikasi di dunia maya. Menurut William teknologi pada saat ini telah banyak merubah cara masyarakat berkomunikasi baik itu perorangan, suatu perkumpulan grup, perkumpulan organisasi, publik maupun komunikasi internasional.

Hate Speech atau ujaran Kebencian pada saat ini telah beralih ke dalam dunia maya, yang tidak lain adalah jejaring sosial yang merupakan sebagai metode komunikasi massa khususnya adalah Platform Twitter atau pada saat ini Platform tersebut telah melakukan *rebranding* atau berganti nama menjadi Platform X. Platform X merupakan salah satu aplikasi yang menyediakan kolom ulasan dimana para penggunanya dapat memberikan suara baik itu berupa pengetahuan, pendapat, maupun ujaran – ujaran yang diberikan yang dapat dilihat oleh siapa pun. Banyak dari pengguna dalam menyampaikan pendapatnya dengan menggunakan perasaan yang kemudian dituangkan kedalam bentuk komentar akan menimbulkan dampak berupa komentar baik atau buruk, tergantung bagaimana kebijakan pengguna dalam bersosial media, jika tidak dapat dikendalikan maka akan berpengaruh buruk

(Farhan & Setiaji, 2023). Salah satunya, ujaran kebencian yang diberikan melalui media sosial seperti ini merupakan bentuk adanya penyalahgunaan jejaring sosial dalam bentuk memprovokasikan suatu hal yang buruk, mengabaikan keberagaman, pencemaran nama baik, maupun intoleransi(Sadat et al., 2022).

Oleh karena itu, peneliti melakukan analisis sentimen terhadap kasus Hate Speech terkait dengan Gacoan yang terdapat pada tweet platform X dengan menggunakan salah satu pendekatan analisis sentimen dengan teknik machine learning, salah satunya yaitu Support Vector Machine (SVM). Teknik machine learning satu ini merupakan salah satu algoritma yang efektif dalam mengklasifikasikan data, baik itu mengidentifikasi ataupun mengklasifikasikan konten yang mengandung Hate Speech. Peneliti menggunakan teknik machine learning satu ini dikarenakan Support Vector Machine (SVM) ini merupakan sebuah algoritma yang sudah terkenal dengan kuatnya kinerja dalam tugas – tugas klasifikasi. Beberapa alasannya adalah model machine learning ini berbasis kernel dan telah menunjukkan adanya potensi dalam mengekstraksi informasi dengan set data yang relatif lebih kecil, tidak hanya itu SVM juga bersifat serbaguna dalam memproses data linier dan non-linier (Al-Khowarizmi et al., 2023).

Dalam penelitian ini fokus pada analisis hate speech terhadap Restoran Siap Saji “Gacoan” melalui Platform X dengan menggunakan model Support Vector Machine dengan menggunakan kernel Linear, yang dimana tentunya telah melalui berbagai macam preprocessing dan juga proses untuk mencari kernel mana yang terbaik untuk digunakan dengan menggunakan GridSearchCv. Dalam kasus data yang menggunakan linear, model SVM mampu memberikan dimensi yang efisien. Umumnya setiap data postingan Platform X memiliki sifat yang yang konteksnya

sangat kompleks seperti banyaknya singkatan, emoji dan tanda baca, hal ini membuat Platform X menjadi data tidak terstruktur (Al-Khowarizmi et al., 2023). Gacoan merupakan salah satu resto siap saji yang saat ini populer dikalangan masyarakat terutama mahasiswa dan tidak menutup kemungkinan mereka yang pernah merasakan makan ditempat tersebut memberikan ulasan ataupun pendapat mereka terhadap restoran tersebut dan seringkali menjadi topik pembicaraan di media sosial.

1.2. Rumusan Masalah

Berlandaskan dalam latar belakang penelitian di atas, penulis dapat merumuskan beberapa masalah sebagai berikut:

1. Bagaimana menganalisa sentimen yang mengandung hatespeech terhadap Gacoan melalui Platform X menggunakan Support Vector Machine (SVM)?
2. Bagaimana pengujian metode Support Vector Machine (SVM) dalam mengklasifikasikan Hate Speech yang telah dikumpulkan dari Platform X?
3. Bagaimana hasil dari analisis dalam mengklasifikasikan Hate Speech terhadap Gacoan pada Platform X?

1.3. Batasan Masalah

Dalam penelitian ini, terdapat beberapa Batasan masalah yang ingin dibahas oleh penulis, diantaranya:

1. Pengumpulan data diambil melalui Platform X menggunakan scraping tool bernama Tweet-Harvest dengan menggunakan kata kunci “Gacoan” mulai dari 01 – 01 – 2024 sampai dengan 05 – 06 - 2024.

2. Analisis sentimen terhadap data tweet gacoan yang dibahas terfokus pada kelas sentimen negatif (hatespeech) dan positif (non-hatespeech) dengan menggunakan text mining yang memberikan hasil *Classification Report* terhadap model SVM.

1.4. Tujuan Penelitian

Terdapat beberapa tujuan yang akan dicapai dalam proses penelitian ini adalah:

1. Memahami serta mengevaluasi bagaimana sentimen *assessment* dalam mengidentifikasi dan mengklasifikasi bentuk hatespeech terhadap data tweet yang berkaitan dengan Gacoan di Platform X.
2. Mengetahui keberhasilan akurasi metode Support Vector Machine (SVM) dalam mengklasifikasikan data tweet yang mengandung Hate Speech di Platform X (Akurasi, Presisi, Recall dan F1Score).
3. Menampilkan hasil actual dan prediksi oleh model kedalam bentuk confusion matrix untuk dihitung kembali secara manual dan laporan hasil klasifikasi kedalam bentuk wordcloud.

1.5. Manfaat Penelitian

Adapun beberapa manfaat dari penelitian ini sebagai berikut:

1. Menjadi referensi serta wawasan dan perspektif baru untuk pengembangan model analisis sentimen khususnya metode Support Vector Machine yang digunakan dalam konteks ujaran kebencian terhadap restoran siap saji di media sosial.

2. Memberikan pengalaman praktis dan dapat dijadikan sumber informasi serta bahan penelitian selanjutnya yang dapat dikembangkan lebih lanjut pada pembahasan lain guna meningkatkan kualitas pembelajaran.
3. Dengan hasil penelitian yang ada dapat membantu masyarakat lebih memahami dinamika diskusi dan beropini di ranah publik, khususnya mengenai Gacoan dimedia sosial sehingga dapat lebih bijak lagi dalam berkomunikasi baik itu menjadi lebih positif dan konstruktif.

BAB II

LANDASAN TEORI

2.1 Analisis dan Hate Speech

Sentimen Analisis adalah penggunaan konten analitic untuk mendapatkan berbagai sumber informasi yang beragam baik itu melalui web ataupun berasal dari platform media sosial. Tidak lain, tujuannya tentu untuk memperoleh opini dari pengguna yang berada di stage atau situs tersebut agar dapat membantu dalam pengambilan keputusan yang strategis, selain itu juga membantu memantau dalam setiap perubahan opini yang kemungkinan akan terjadi (Annisa, 2022).

Menurut (Purnamasari et al., 2023) Analisis Sentimen umumnya mengambil berbagai macam karakteristik dari sebuah informasi berbentuk tekstual terstruktur maupun tidak terstruktur kemudian menganalisisnya agar dapat mendapatkan pikiran, pendapat, serta perasaan yang memungkinkan keluar dari sebuah informasi atau data tersebut. Sedangkan menurut (Faruk Hussain Abdullah, 2023) Analisis sentimen merupakan sebuah proses dalam menganalisis teks yang berbasis digital agar dapat mengetahui apakah ada mengandung unsur emosional terhadap pesan tersebut baik itu positif, negatif ataupun netral. Analisis ini melibatkan penggunaan algoritma dalam prosesnya dan juga model machine learning agar dapat mengolah teks serta mengidentifikasi kata ataupun frasa yang mengarahkan kepada sentimen tertentu (Septian, 2023).

2.1.1 Kategori / Kelas Sentimen

Dalam analisis sentimen tentunya terdapat alat yang dapat digunakan sebagai pemindai teks secara otomatis guna untuk mengetahui sikap penulisnya terhadap suatu topik yang sedang dalam perbincangan.

Oleh karena itu, dalam penelitian (Wahyu Sejati et al., 2023) sistem yang dirancang agar bekerja secara otomatis akan mengelompokkan opini dan ulasan yang didapatkan sesuai dengan kelasnya, yang dimana kelas sentimen itu secara umum dibagi menjadi kelas positif, netral dan negatif. Opini dan ulasan diambil melalui pengumpulan data yang berhubungan dengan sistem seperti referensi tentang kritik dan saran. Semakin akurat kualitas data dan informasi yang diperoleh maka sistem akan mendapatkan hasil *vocabulary* yang semakin besar dan berkualitas sehingga keakuratan sistem dalam proses memprediksi kelas sentimennya akan lebih tepat.

Dalam penelitian pada buku (Purnamasari et al., 2023) analisis sentimen dilakukan dengan menggunakan tiga kategori sentimen, diantaranya adalah sentimen positif, sentimen negatif dan terakhir sentimen netral. Dalam penentuan klasifikasinya berlandaskan kategori yang ada yaitu dengan cara menghitung jumlah kata atau frasa yang terkandung didalam sebuah ulasan yang diambil, apabila kata positif lebih mewakili dibandingkan dengan jumlah kata negatif, maka ulasan tersebut merupakan kelas yang berada di dalam klasifikasi sentimen positif dan begitu seterusnya apabila jumlah kata negatif lebih mewakili maka termasuk kedalam kelas sentimen negatif, hal serupa dengan sentimen netral.

2.1.2 Hate Speech

Hate Speech sendiri merupakan bentuk komunikasi yang dilakukan oleh seseorang atau pelaku, baik itu sekelompok orang dalam bentuk provokasi, hasutan maupun penghinaan terhadap individu atau kelompok lain berdasarkan berbagai faktor contohnya seperti ras, warna kulit, etnis, gender, cacat, orientasi seksual, kewarganegaraan, agama maupun aspek lainnya (Sri, 2018). Definisi didalam *Council of Europe Hatespeech* dapat dipahami bahwa “segala bentuk ekspresi apapun itu yang bersifat menyebarkan, menghasut, mendukung ataupun membenarkan sifat kebencian yang rasial, xenophobia, anti-semitisme maupun bentuk kebencian lainnya yang berakar pada sifat tidak toleransi itu semua merupakan nasionalisme agresif yang intoleran serta tindakan diskriminatif dan permusuhan terhadap kelompok minoritas, migran dan individu yang berlatar belakang imigran”.

Teks Hate Speech atau biasa disebut dengan ujaran kebencian ini tersebar luas bukan karena kebetulan, kesenangan atau hiburan semata – mata, akan tetapi mungkin ada tujuan tertentu dibaliknya. Jika dikaitkan dengan teknologi digital, maka akibatnya adalah orang – orang yang menggunakan ruang publik di media sosial lebih cenderung mempercayai informasi yang terdiri dari teks yang mengandung hate speech tersebut dengan mengunggah dan menyebarkan pesan tersebut melalui media sosial yang terkait, hal ini menyebabkan penyebaran hate speech bergerak dengan cepat dan menarik perhatian banyak orang (Kusumasari & Arifianto, 2020). Pada media sosial, hate speech didefinisikan dalam bentuk tulisan, ucapan, pidato atau orasi yang digunakan di media sosial untuk mendorong penghinaan dengan menggunakan akun anonim untuk menyembunyikan

identitasnya. Penyebaran hate speech dapat dijelaskan dengan konsep ruang publik sebagai tempat orang berbicara dan bernegosiasi di media sosial (Ahnaf & Suhadi, 2014).

2.1.3 Dampak Hatespeech dan Regulasinya

Menurut (Pelajar, 2024) semua hal yang berkaitan dengan hatespeech atau ujaran kebencian tentunya dapat memicu Tindakan kekerasan dan konflik antar individu atau kelompok, bahkan dapat merugikan pihak lain. Disamping itu, hatespeech memiliki efek negatif diantaranya:

- a. Meningkatkan konflik dan kekerasan yang dapat menyebabkan ketegangan dan konflik antar masyarakat yang merupakan reaksi timbal balik yang mengarah pada eskalasi kekerasan.
- b. Diskriminasi dan pelanggaran hak asasi manusia seperti hak untuk diperlakukan secara adil dan setara.
- c. Dampak psikologis yang merugikan dengan adanya kekerasan terhadap orang yang dapat menyebabkan trauma, depresi, rendah diri dan masalah kesehatan mental lainnya.
- d. Menghambat kohesi sosial dengan kekerasan yang memecah belah masyarakat, menghambat integrasi sosial dan menciptakan jurang antar kelompok yang saling menjatuhkan.
- e. Menghambat kemajuan sosial seperti menghentikan upaya untuk menghilangkan diskriminasi, mendorong kesetaraan, dan membangun masyarakat yang sejahtera.

Meskipun demikian, pemerintah telah mengeluarkan upayanya untuk mengontrol penyebaran hatespeech didunia digital dengan mengeluarkan surat edaran dan beberapa undang – undang. Aturan -Aturan yang dapat digunakan untuk menangani secara hukum ujaran kebencian ada didalam Pasal 156 KUHP dan Pasal 355 KUHP mengenai “ancaman kekerasan”. Sebagai negara yang telah memberikan antisipasi mengenai hatespeech, Indonesia harus bertanggung jawab untuk menghormati, melindungi dan memenuhi hak warga negaranya untuk terbebas dari ujaran – ujaran kebencian seperti hasutan itu (Ahnaf & Suhadi, 2014).

2.2 Rebranding Twitter to Platform X

Dipandang sebagai aplikasi paling populer ditahun 2023 karena menjadi sumber informasi bermanfaat, selain itu karena aplikasi ini merupakan platform terbuka tentunya memiliki banyak tantangan yang harus dihadapi terutama kontennya dan juga termasuk kesalahan ejaan dan tata bahasa. Twitter bukan hanya digunakan untuk berbagi informasi faktual dan pengetahuan, tetapi juga digunakan untuk mengungkapkan berbagai ide dan bermacam emosi melalui tweet. Karena itu, akan sangat membantu apabila setiap pengguna twitter berkontribusi dengan membagikan tweet yang memiliki nilai bagi seluruh pengguna. Jenis tweet yang tersedia di twitter mencakup berita terbaru, pesan inspirasi, komentar tentang berbagai topik, konten komedi, penafsiran ayat suci, dan kontribusi lainnya yang dibuat oleh penggunanya (Diani et al., 2024).

Menurut laporan pada (Tami, 2023) sebelum itu, Elon Musk si pemilik Twitter telah meminta pengikutnya untuk mengusulkan logo baru untuk X dan meminta istilah baru untuk menggantikan kata “tweeting” jika Twitter tidak lagi

digunakan. Salah satu penggunanya akhirnya mengusulkan logo baru yang dimana akhirnya Elon Musk dan Linda Yaccarino sebagai CEO Twitter, telah resmi mengumumkan bahwa Twitter akan mengganti mereknya menjadi “X”. Linda menyatakan bahwa “X” adalah representasi dimasa yang akan datang melalui interaktivitas tanpa batas dengan terfokus pada konten audio, video, pesan, serta pembayaran dan perbankan sambil membangun pasar global atau biasa disebut marketplace sebagai ide, barang layanan serta peluang. Dimana hal ini dia mengharapkan X akan menghubungkan dunia dengan cara yang belum pernah terlihat sebelumnya disertai adanya dukungan kecerdasan buatan (AI).

Selain itu, Platform X memungkinkan penggunanya berinteraksi dan berteman dengan pengguna lain melalui fitur like, retweet dan balasan. Maka dari itu, Platform X memiliki fitur Top Trending atau Trending For You yang dimana membantu para pengguna melihat tweet apa yang sedang ramai diperbincangkan atau biasa disebut dikicaukan oleh pengguna Platform X diseluruh dunia (Amelia et al., 2024).

2.3 Support Vector Machine (SVM)

Support Vector Machine dikembangkan oleh Boser, Guyon, dan Vapnik pertama kali diperkenalkan di *Annual Workshop on Computational Learning Theory* ditahun 1992, dimana SVM merupakan model atau metode yang baru – baru ini menjadi banyak sorotan sebagai metode yang mutakhir dalam pengenalan pola. Metode SVM ini membagikan kumpulan data kedalam beberapa kelompok dengan membuat garis agar memisahkan satu kelompok dengan meletakkan pemisah yang tegas. Konsep dasar SVM sebenarnya terdiri dari gabungan teori komputasi dari

beberapa dekade yang lalu, contohnya seperti margin hyperplane dan kernel oleh Aronszajn tahun 1950 (Fachri Muhammad, Fauziah, 2019) (Nufairi et al., 2024).

SVM biasanya menghindari overfitting data pelatihan dikarenakan memiliki kemampuan yang dapat menggeneralisasi data pelatihan ke data yang tidak terlihat sebelumnya (Fachri Muhammad, Fauziah, 2019). Dalam SVM terdapat proses klasifikasi yang dilakukan menggunakan hyperplane atau biasa disebut garis pembatas guna untuk membedakan yang mana kelas positif dan yang mana kelas negatif. Margin adalah jarak antara titik vector dalam kelas dan hyperplane, dimana Margin yang lebih besar biasanya tingkat error generalisasi dari pemilahnya lebih rendah dikarenakan terdapat hyperplane yang baik yaitu yang memiliki jarak terdekat ke titik data pelatihan kelas (Siti Aisah et al., 2024).

Menurut (Setiadi, 2023) teknik ini digunakan untuk mengidentifikasi pola dalam data pelatihan berlabel oleh SVM, yang kemudian dilanjutkan dengan membangun model pembelajaran mesin pada saat proses klasifikasi berlangsung. Model ini kemudian akan dipakai pada data tweet tambahan yang tidak berlabel, guna untuk menentukan estimasi kelas sentimen, baik negatif maupun positif. Sedangkan menurut (Laurensz & Eko Sedyono, 2021) SVM merupakan teknik dengan kualitas dan akurasi tingginya sangat diminati dibandingkan dengan algoritma lainnya. Namun, untuk menggunakannya, harus melalui tahap pelatihan sequential dan proses pengujian.

2.3.1 Cara Kerja SVM

Hyperplane merupakan model matematis yang digunakan untuk membedakan antara dua kelompok data berdasarkan kelas pada algoritma SVM,

cara kerja pada SVM sendiri dengan menemukan hyperline terbaik dan mengoptimalkan jarak antar kelas (Farhan & Setiaji, 2023). Didalam penelitian (Darussalam & Arief, 2018) SVM melakukan pengkategorian teks dengan cara sebagai berikut:

a. Diketahui titik data $\{(x_i, y_i)\}$

Dimana, $x_i = \{x_1, x_2, \dots, x_n\} \in R^n$, $y_i \in \{+1, -1\}$

Permasalahan diatas dapat diselesaikan dengan menggunakan persamaan *hard margin linear SVM* , yaitu:

$$\min_{w, b} \frac{1}{2} \|w\|^2 \dots\dots\dots(2.1)$$

diikuti dengan persamaan

$$s. t. y_i(w^T x_i + b) \geq 1, \forall i, \dots\dots\dots(2.2)$$

Dimana, $(w) = wight$ adalah parameter.

$(b) = bias$ adalah parameter.

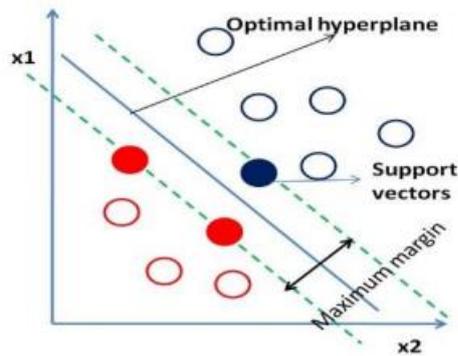
b. Fungsi weight didapatkan melalui persamaan:

$$w = \sum_{i=1}^N a_i y_i x_i \dots\dots\dots(2.3)$$

c. Fungsi bias didapatkan melalui persamaan:

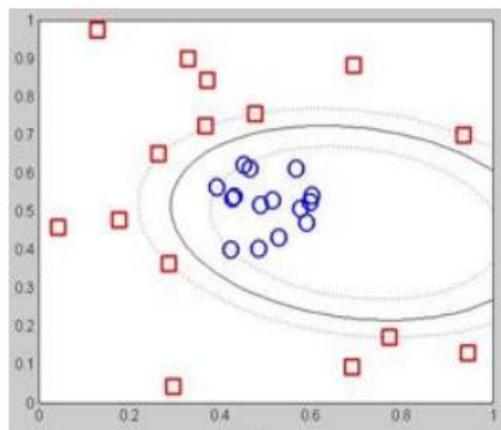
$$b = -\frac{1}{2}(w \cdot x^+ + w \cdot x^-) \dots\dots\dots(2.4)$$

Berdasarkan cara kerja SVM yang diketahui oleh (Oryza Habibie Rahman et al., 2021) pada dasarnya, SVM melakukan klasifikasi biner dan linier. Namun, setelah pengembangan yang dilakukan itu bertujuan agar dapat melakukan pengklasifikasian multiclass dengan teknik yang diterapkan yaitu one vs one, one vs all dan directed acyclic graph yang diarahkan.



Gambar 2. 1 Konsep hyperplane pada SVM Linier

Kemudian agar berjalan pada proses non-linier, yaitu dengan cara memasukkan konsep kernel pada ruang kerja berdimensi tinggi dilanjutkan dengan mencari hyperplane yang mampu mengoptimalkan jarak antara kelas yang ada.



Gambar 2. 2 Konsep Hyperplane pada SVM Non-Linier

Support vector Machine menggambar hyperplane dengan menggunakan fungsi matematika “kernel” gunanya untuk mentransformasikan data (Armadianti et al., 2024). Karakteristik dan parameter kernel sangat bergantung pada keakuratan model yang dibuat, dengan dua jenis SVM (linier dan non-linier) berdasarkan sifatnya. SVM linier memisahkan dua kelas sedangkan SVM non-linier memiliki kemampuan untuk melakukan trik kernel ke ruang dimensi tinggi (Alviani et al., 2023).

Didalam penelitian (Oryza Habibie Rahman et al., 2021) terdapat beberapa fungsi kernel yang populer digunakan oleh banyak orang diantaranya adalah kernel Linear, non – linear, Sigmoid, Polynomial, Radial Basic Function (RBF), biasanya fungsi ini diterapkan guna memetakan dimensi awal pada himpunan data (dimensi lebih rendah) ke dimensi baru yang relatif tinggi (Sujadi, 2022).

Terdapat juga beberapa fungsi pada kernel *trick* yang dibahas didalam penelitian (Ade Dwi Dayani et al., 2024) dan fungsi ini sangat populer digunakan dalam metode SVM, sebagai berikut:

a. Kernel Linear

digunakan dalam data yang memiliki dua kelas ataupun data yang diklasifikasikan secara linear dengan bentuk seperti dibawah ini:

$$K(x_i, x_j) = x_i \cdot x_j \dots\dots\dots(2.5)$$

b. Polynomial

digunakan dalam menyelesaikan suatu permasalahan data yang tidak terpisah berdasarkan linear dengan bentuk seperti dibawah ini:

$$K(x_i, x_j) = (x_i \cdot x_j + c)^d \dots\dots\dots(2.6)$$

c. Radial Bias Function (RBF)

Atau biasa disebut dengan gaussian untuk menyelesaikan suatu permasalahan yang sama dengan polynomial atau data lebih dari dua kelas (banyak kelas) dengan bentuk persamaan seperti dibawah ini:

$$K(x_i, x_j) = \exp(-\gamma ||x_i - x_j||^2) \dots\dots\dots(2.7)$$

K = fungsi kernel

x_i, x_j = dokumen dataset

γ = parameter kecepatan proses *training*

Exp = basis logaritma alami

d = derajat *polynomial*

2.4 Python

Berdasarkan didalam penelitian oleh (Wati & Ernawati, 2021) python itu merupakan bahasa yang mudah dipelajari serta difokuskan pada keterbacaan kode yang bersifat multi platform. Sedangkan menurut (Nurjanah & Insanudin, 2018) python merupakan bahasa pemrograman yang dikemukakan oleh Guido van Rossum dan banyak dikenal sebagai bahasa *scripting* dan pemrograman website. Python biasanya digunakan sebagai bahasa skrip, tetapi ia dapat digunakan dan diterapkan untuk tujuan yang biasanya tidak dapat dilakukan dengan bahasa skrip.

Bahasa pemrograman interpretatif yang bisa digunakan dibanyak platform tentu dengan adanya filosofi perancangan yang dikhususkan. Bahasa ini merupakan bahasa yang paling banyak digunakan dalam bidang Data Science, Machine Learning serta Internet of Things (IoT). Terdapat beberapa manfaat python seperti membangun model, membuat visualisasi serta membersihkan data. Hal ini lah yang menjadikan python banyak digunakan oleh banyak orang, selain manfaatnya itu python juga memiliki library yang lengkap, mulai dari open source hingga ke fleksibelannya yang mampu meningkatkan produktivitas terhadap developer (Susanti & Walid, 2022).

Penelitian oleh (Qisthiano et al., 2023) menjelaskan bahwa beberapa dekade terakhir, terlihat adanya peningkatan terhadap library pada python sehingga menjadikannya pilihan yang kuat sebagai pengerjaan dalam menganalisis data. Salah satu contoh bahasa tingkat tinggi itu adalah python selain itu terdapat pascal,

C++, java, perl yang juga merupakan bahasa pemrograman tingkat tinggi. Selain itu terdapat juga bahasa tingkat rendah yang merupakan bahasa mesin atau biasa disebut assembly dimana sebuah komputer hanya mampu mengeksekusi (Nurjanah & Insanudin, 2018).

Alasan mengapa python banyak digunakan dan populer dikalangan programmer yakni mudah diterapkan kedalam domain Artificial Intelligence diantaranya data mining, pemrograman bahasa alami dan juga image processing. Dalam python, library yang digunakan untuk memproses sebuah kualitas ataupun kuantitas data gambar agar dapat lebih mudah diidentifikasi pada saat proses klasifikasi itu menggunakan *library Open Computer Vision (OpenCv)* (Reza et al., 2022).

2.5 Gacoan

Pt. Pesta Pora Abadi adalah perusahaan lokal Indonesia yang bergerak di industri makanan dan minuman dengan produknya yang terkenal yaitu Mie Gacoan, dimana merupakan waralaba mie pedas yang sedang populer di Indonesia. Perusahaan ini memiliki restoran yang luas dengan makanan, minuman, dan makanan ringan yang dimulai dengan harga yang sangat masuk di kantong pelajar dan mahasiswa (Nugroho et al., 2024). Gacoan merupakan restoran cepat saji nomor satu yang paling banyak diminati oleh kalangan anak muda dengan menyajikan menu utama berupa olahan mie siap santap dengan memiliki cita rasa pedas yang unik (Irsyada, 2024).

Menurut pendapat (Syahara & Suryowati, n.d.) mie yang dibuat, memiliki ciri khas yang kenyal dan gurih dengan harga yang sangat *affordable*. Pada

penelitiannya terdapat hasil tabulasi frekuensi responden yang menunjukkan bahwa pelanggan “setuju” restoran siap saji ini memiliki bukti fisik, empati, keandalan, daya tanggap serta jaminan di mata pelanggannya. Kemudian, terdapat juga responden yang menyatakan “sangat setuju” terkait porsi, rasa, tekstur, warna dan juga cara penyajian pada mie gacoan. Selain itu, mereka juga mengatakan bahwa harga Mie Gacoan yang ditawarkan sangat sesuai dan berkualitas.

Mie Gacoan tentunya memiliki banyak varian rasa termasuk tingkat atau level kepedasan yang beragam terhadap mie gacoannya, hal ini tentunya akan menarik perhatian pelanggan akan rasa puasny terhadap kualitas produk Mie Gacoan. Konsumen dapat memilih level kepedasan pada mie sesuai dengan selera masing - masing. Mereka yang tidak menyukai pedas dapat memesan mie level rendah yang tentunya komposisi cabainya tidak digunakan sama sekali, sementara mereka yang ingin menantang atau menguji adrenalin selera makan mereka dapat mencoba mie yang memiliki level lebih tinggi (Purbowo et al., 2024).

2.6 Penelitian Terdahulu

Tabel 2. 1 Tabel Penelitian Terdahulu

No.	Referensi	Peneliti	Tahun	Hasil Penelitian
1.	Analisis Sentimen Terhadap Opini Publik pada Sosial Media Twitter Menggunakan Metode Support Vector Machine.	(Ade Dwi Dayani, Yuhandri, Gunadi Widi Nurcahyo)	2024	Penelitian yang diperoleh dari kinerja analisis SVM menunjukkan performa yang cukup dalam melakukan

				analisis klasifikasi terhadap opini publik pada sosial media twitter.
2.	Analisis Sentimen Pada Ulasan Aplikasi Indodax di Google Play Store Menggunakan Metode Support Vector Machine.	(Alfio Kusuma, Ermatita, Helena Nurramdhani Irmada)	2022	Penelitian ini melakukan analisis sentimen yang dimulai dari pengambilan data terhadap ulasan yang ada pada aplikasi Indodax kemudian data diberi label negatif atau positif, yang dimana selanjutnya dilakukan praproses data dan pembersihan data serta klasifikasi menggunakan SVM.
3.	Klasifikasi Ujaran Kebencian pada Media Sosial Twitter	(Oryza Habibie Rahman,	2021	Penelitian ini untuk mengetahui bagaimana pola

	<p>Menggunakan Support Vector Machine.</p>	<p>Gunawan Abdillah, Agus Komarudin)</p>	<p>perilaku dalam ujaran kebencian yang terbagi menjadi lima kelas yaitu suku, agama, ras, antar golongan dan netral menggunakan SVM dengan pengujian membandingkan tiga buah kernel yang umum dipakai.</p>
--	--	--	---

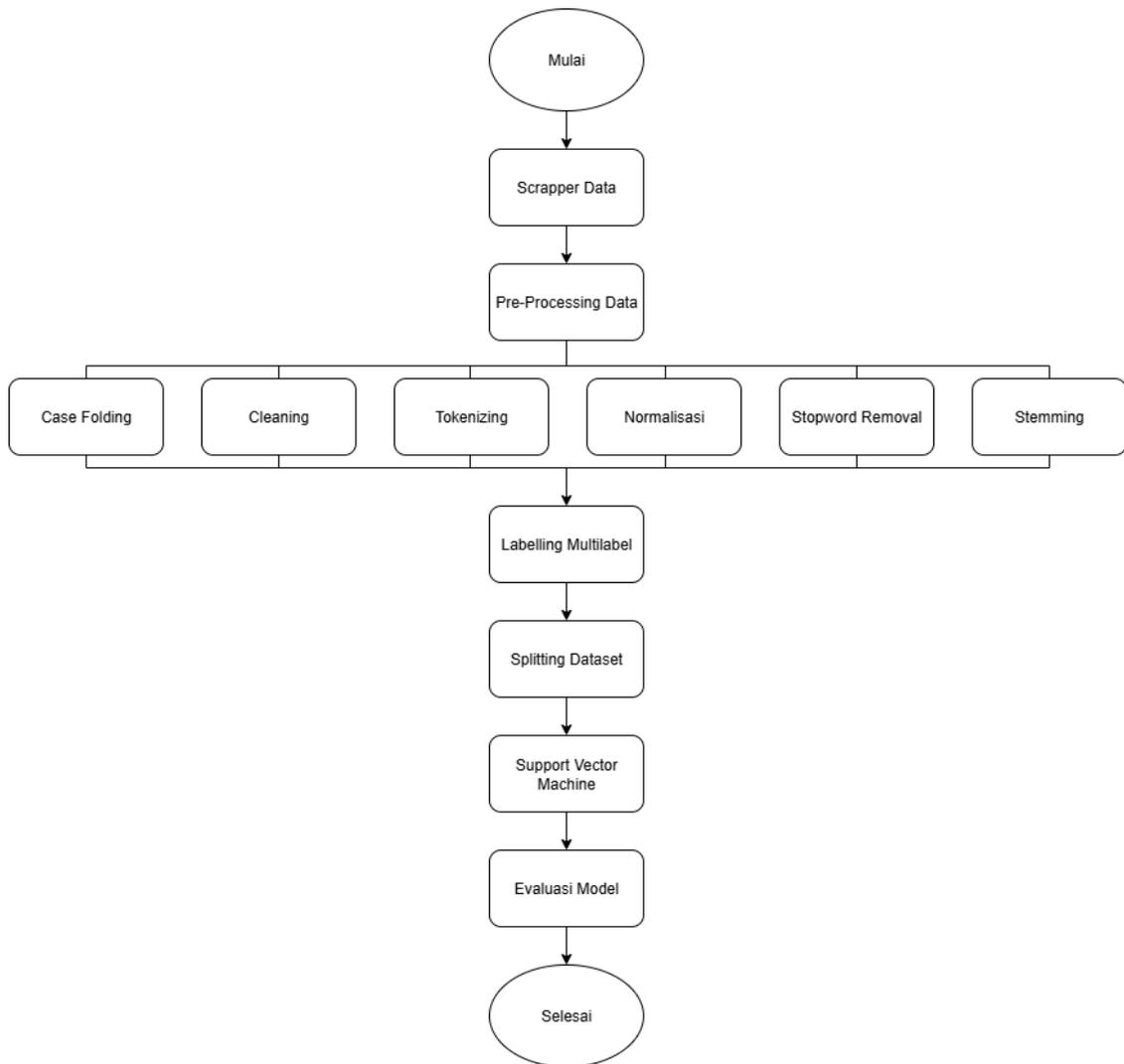
BAB III

METODOLOGI PENELITIAN

3.1. Pendekatan Penelitian

Pendekatan penelitian yang digunakan peneliti merupakan metode penelitian kuantitatif dengan metode Support Vector Machine (SVM) sebagai model yang digunakan untuk melakukan klasifikasi terhadap data tweet atau postingan yang dikumpulkan mengenai Gacoan terutama yang mengandung konten hatespeech pada Platform X. Sehingga dari hasil tersebut dapat dijadikan sebagai bahan evaluasi maupun pengoptimalan strategi pemasaran terhadap pihak restoran siap saji Gacoan serta sebagai bentuk pengembangan dari metode Support Vector Machine dalam hal klasifikasi. Pendekatan kuantitatif merupakan metode penelitian yang mengarah pada prinsip positivisme yang terfokus pada pemanfaatan data numerik, dimana data tersebut di analisis menggunakan teknik machine learning guna melakukan pengukuran serta evaluasi secara objektif (Irsyada, 2024).

Adapun tahapan penelitian ini digambarkan secara sistematis melalui kerangka kerja yang tersusun atas beberapa langkah utama. Setiap langkah dalam kerangka ini saling terhubung dan membentuk alur proses yang runtut, dimulai dari proses pengambilan data hingga evaluasi akhir model. Tujuan dari penyusunan kerangka ini adalah untuk memberikan gambaran menyeluruh terkait proses penelitian yang dilakukan yang dimana kerangka ini mencerminkan pendekatan logis dalam mengolah data teks secara bertahap. Berikut tahapan tersebut dapat dilihat dalam bentuk bagan kerangka kerja seperti yang ditampilkan pada gambar dibawah ini.



Gambar 3. 1 Kerangka Penelitian

3.2. Teknik Pengumpulan Data

Untuk teknik pengumpulan data yang digunakan penulis dalam penelitian ini dengan cara mengumpulkan berbagai sumber informasi dari penelitian terdahulu dan pengetahuan yang mendukung proses penelitian yang disusun dengan terdiri dari beberapa proses rangkaian pengumpulan datanya sebagai berikut:

3.2.1 Studi Pustaka

Studi kepustakaan melibatkan penelitian berbagai buku referensi dan hasil penelitian sebelumnya yang relevan dan memungkinkan untuk memperoleh dasar

teori terkait topik yang diteliti. Selain itu, studi kepustakaan juga mencakup metode pengumpulan data melalui telaah terhadap literatur, catatan serta laporan yang berhubungan dengan permasalahan yang ingin di selesaikan(Purbowo et al., 2024). Berdasarkan penelitian ini, peneliti menyesuaikan serta mempelajari teori yang berasal dari berbagai referensi yang berkaitan dengan model Support Vector Machine sebagai bahan untuk melengkapi penelitian ini.

Menurut (Nufairi et al., 2024) studi literatur berperan sebagai sumber pendukung dalam menyelesaikan sebuah permasalahan dalam penelitian yang tentunya digunakan sebagai referensi pustaka untuk pengembangan penelitian. Proses ini mencakup pengumpulan informasi dari berbagai buku, penelitian atau jurnal sebelumnya serta dari situs web yang membahas analisis sentimen dan algoritma Support Vector Machine. Dari studi literatur berikut, penulis memiliki tujuan yang ingin dicapai guna memperoleh pengetahuan yang lebih luas mengenai teknik dan model yang digunakan dalam penelitian dengan tujuan dapat diketahui perkembangan dari model tersebut dan dapat diterapkan sebagai bahan evaluasi maupun yang lainnya. Terutama kepada permasalahan analisis sentimen yang mengandung hatespeech yang menyebar diranah sosial media.

3.2.2 Scrapping Data

Metode Scrapping yang digunakan peneliti disini bertujuan untuk mengumpulkan data dari Platform X yang sebelumnya dikenal sebagai aplikasi Twitter. Proses ini dilakukan dengan menggunakan scrapping tools bernama tweet-harvest agar dapat mengakses serta mengekstrak data terkait postingan yang berhubungan dengan restoran Gacoan. Dengan memanfaatkan tools tweet-harvest

ini memungkinkan peneliti untuk memperoleh data secara efisien termasuk tweet, komentar dan juga interaksi pengguna yang berhubungan dengan permasalahan terhadap restoran Gacoan ini, terkhususnya jika ada permasalahan yang mengandung hatespeech didalam tweet.

```
#Scrapping Data
filename = 'FinalScrapping.csv'
search_keyword = 'gacoan since:2024-01-01 until:2024-02-01 lang:id'
limit = 500

!npx -y tweet-harvest@2.6.1 -o "{filename}" -s "{search_keyword}" --tab "LATEST" -l {limit} --token {twitter_auth_token}
```

Gambar 3. 2 Scrapping Tweet-harvest

Data yang diambil sesuai dengan kata kunci yang digunakan yaitu “Gacoan” sehingga dapat memberikan gambaran yang jelas. Data yang diambil sekitar 2500 data pada tahun 2024 selama lima bulan yaitu 01-01-2024 hingga 05-06-2024. Data diambil menggunakan twitter_auth_token yang diambil dari website Platform X sendiri, akan tetapi pengambilan data dilakukan secara bertahap seperti mengambilnya setiap per satu bulan dengan limit pengambilan yang dibuat adalah sebanyak 500 data, hal ini dikarenakan adanya keterbatasan pengambilan data yang dapat dilakukan dengan menggunakan tools ini. Maka dari itu, proses scrapping dilakukan sebanyak 5 kali untuk mengambil sebanyak 2500 data selama lima bulan yang kemudian data yang telah diambil tersebut disimpan ke dalam file.

3.3. Metodologi Pengembangan Sistem

3.3.1 Preprocessing Data

Dataset yang diperoleh melalui proses scrapping masih berupa kata atau kalimat yang tidak terstruktur dan tidak teratur. Selain itu dalam klasifikasi teks, proses ini baik itu saat pemilihan fitur memiliki peran krusial karena dapat

memengaruhi tingkat akurasi setiap metode yang digunakan (Oryza Habibie Rahman et al., 2021). Oleh karena itu, perlu melakukan tahap preprocessing terlebih dahulu untuk membersihkan data agar pada saat melakukan analisis sentimen dapat menghasilkan output yang lebih akurat (Ade Dwi Dayani et al., 2024). Berikut terdapat beberapa tahapan yang dilakukan dalam proses preprocessing data, yaitu:

a. Case Folding

Tahap dimana proses mengubah seluruh teks dalam dokumen menjadi format yang seragam dimana seluruh huruf dalam sebuah tweet diubah menjadi huruf kecil. Hal ini bertujuan agar menghindari kesalahan saat melakukan proses tokenisasi.

b. Cleaning

Dilakukan untuk mengurangi *noise* dalam data dengan membersihkan serta menghapus kata – kata yang tidak relevan seperti tanda baca, link url, maupun emoji. Melalui proses cleansing, data teks lebih siap untuk melanjutkan pada tahap analisis berikutnya seperti pemodelan sentimen, klasifikasi ataupun ekstraksi informasi.

c. Tokenizing

Proses pemecah rangkaian karakter menjadi unit – unit yang lebih kecil seperti kata atau frasa yang disebut sebagai token. Dengan fungsinya yang dapat memotong kalimat menjadi kata, karakter, symbol ataupun tanda baca akan mempermudah tahap selanjutnya dalam analisis.

d. Normalisasi

Salah satu proses yang bertujuan untuk menyamakan variasi bentuk kata agar teks menjadi seragam contohnya seperti menangani kata – kata yang tidak baku

menjadi kedalam bentuk kata baku. Proses normalisasi ini dilakukan karena sangat banyak kata yang informal yang digunakan oleh pengguna di media sosial.

e. Stopword Removal

Tahap ini melakukan penghapusan kata – kata yang tidak penting atau bisa dikatakan tidak memiliki makna yang signifikan. Contohnya seperti kata keterangan, kata sambung dan lainnya. Dimana pada akhirnya kata yang tidak berguna akan dibuang dan tidak akan digunakan.

f. Stemming

Setiap kata yang mengandung imbuhan diubah menjadi bentuk dasar, selain itu kata dengan ejaan yang kurang tepat juga diperbaiki atau jika memungkinkan dihapus pada tahap ini. Proses ini tentunya menggunakan *library* untuk mengubah kata imbuhan.

3.3.2 Labelling Data

Pelabelan dilakukan dengan cara memberikan label terhadap data berdasarkan kamus positif dan negatif yang digunakan. Dalam proses labelling ini menggunakan pendekatan INSET Lexicon Based untuk 2 klasifikasi yaitu positif dan negatif untuk menentukan sentiment dari teks yang sudah melalui tahap preprocessing. Kamus lexicon positif dan negatif diunduh pada github agar dapat digunakan pada proses ini, kamu ini berisi daftar kata beserta skor sentimennya. Setiap kata akan dikategorikan berdasarkan skor yang terkandung didalam sebuah teksnya. Berikut penjabaran dari skema pelabelan yang digunakan dalam penelitian ini:

a. Non-Hatespeech

Jika skor sentimen lebih dari 0, sentimen dikategorikan sebagai “Non-Hatespeech”.

b. Hatespeech

Jika skor sentimen kurang dari atau sama dengan 0, sentiment dikategorikan sebagai “Hatespeech”.

3.3.3 Split Data dan Vektorisasi

Setelah data teks selesai melalui proses pembersihan dan praproses seperti tokenisasi, stemming atau stopword removal tahap selanjutnya adalah memisahkan data menjadi dua bagian yaitu data latih (training data) dan data uji (testing data). Pemisahan ini bertujuan untuk melatih model pada sebagian data kemudian menguji performa model tersebut pada data yang belum pernah dilihat sebelumnya untuk mengetahui kemampuannya dalam melakukan generalisasi.

```
# Memisahkan data menjadi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42, stratify=y)
```

Gambar 3. 3 Kode train_test_split

Dalam penelitian ini, pemisahan data dilakukan menggunakan fungsi `train_test_split` dari Pustaka `scikit-learn`. Parameter `test_size=0.2` menunjukkan bahwa sebanyak 20% data digunakan sebagai data uji, sementara 80% sisanya digunakan untuk data latih. Pemisahan dilakukan secara acak namun dikendalikan oleh `random_state=42` agar hasil pemisahan dapat direproduksi (konsisten). Selain itu, digunakan parameter `stratify=y` agar proporsi kelas (kategori) pada data latih dan data uji tetap seimbang terutama Ketika data memiliki distribusi kelas yang tidak merata.

Kemudian model pembelajar mesin tidak dapat memproses data dalam bentuk teks mentah. Oleh karena itu, diperlukan proses transformasi teks menjadi representasi numerik. Salah satu metode yang digunakan dalam penelitian ini adalah `CountVectorizer` yang disediakan oleh Pustaka `Scikit-learn`. `CountVectorizer` bekerja dengan cara mengubah dokumen teks menjadi representasi vektor berdasarkan frekuensi kemunculan setiap kata dalam dokumen. Proses ini sering disebut juga sebagai `Bag-of-Word`. Setiap kata unik dalam seluruh korpus teks dianggap sebagai fitur dan nilai dari fitur tersebut adalah jumlah kemunculannya dalam dokumen tertentu.

```
vectorizer = CountVectorizer()  
X_train_vec = vectorizer.fit_transform(X_train)  
X_test_vec = vectorizer.transform(X_test)  
  
X_train_vec.shape, X_test_vec.shape
```

Gambar 3. 4 Vectorizer Code

`Vectorizer = CountVectorizer()` disini untuk membuat objek `CountVectorizer`. Sedangkan `fit_transform(X_train)` untuk membangun kosakata dari data latih dan sekaligus mengubah data latih menjadi vektor numerik. Kemudian `transform(X_test)` akan mengubah data uji menjadi bentuk vektor berdasarkan kosakata yang telah dibentuk dari data latih (tanpa melakukan fit ulang agar tidak terjadi kebocoran data).

3.3.4 Implementasi Support Vector Machine

Data yang telah melalui proses pembagian data menjadi data uji dan data latih selanjutnya akan dilatih dengan model `Support Vector Machine` menggunakan data latih sebesar 80%. Proses pelatihan ini bertujuan untuk menemukan hasil

klasifikasi yang membedakan antara kelas hatespeech maupun kelas non-hatespeech dengan akurasi tinggi. Implementasi model ini memungkinkan memberikan pemahaman lebih mendalam mengenai pola hatespeech terhadap permasalahan pada restoran siap saji Gacoan dalam upaya pengembangan terhadap model Support Vector Machine dalam analisis sentimen.

Pada tahap melatih model klasifikasi SVM menggunakan data yang sudah di proses dan dibagi sebelumnya untuk mengklasifikasikan tweet ke dalam kategori “Hatespeech” dan “Non-Hatespeech”. Beberapa langkah yang akan dilakukan pada saat implementasi model SVM diantaranya adalah melakukan Tuning Parameter, kemudian melakukan proses pelatihan Model SVM yang kemudian akan divisualisasikan dengan Confusion Matrix dan Classification Report serta Word Cloud. Secara keseluruhan pada tahap ini melibatkan pemilihan Parameter terbaik untuk model SVM, melatih model, mengevaluasi kinerjanya menggunakan berbagai metrik yaitu seperti confusion matrix, laporan klasifikasi dan kurva ROC.

3.3.5 Evaluasi Model (Akurasi, Presisi, dan Recall)

Evaluasi model pada penelitian ini dilakukan pengujian untuk menilai kinerja model klasifikasi yang telah dibangun dimana proses ini mencakup perhitungan *accuracy*, *precision*, *recall* dan *f1-score* guna sejauh mana model mampu melakukan klasifikasi dengan baik. Selain itu juga digunakan Confusion Matrix, Classification Report serta ROC Curve sebagai bentuk perbandingan evaluasi model juga. Akan tetapi, dalam kasus hatespeech, *recall* pada kelas negatif lebih penting agar model tidak melewatkan teks yang mengandung ujaran kebencian.

a. Accuracy

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100\% \dots\dots\dots (3.1)$$

b. Precision

$$Precision = \frac{TP}{(TP+FP)} \dots\dots\dots (3.2)$$

c. Recall

$$Recall = \frac{TP}{(TP+FN)} \dots\dots\dots (3.3)$$

d. F1-Score

$$F1 - Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \dots\dots\dots (3.4)$$

3.4 Perangkat Penelitian

Perangkat atau alat dan bahan baik itu perangkat keras (Hardware) maupun perangkat lunak (Software) yang peneliti gunakan di sesuaikan dengan kebutuhan dalam penelitian ini. Dua hal ini digunakan sesuai dengan kemampuan dan fungsi dari setiap perangkatnya, demi mencapai hasil yang maksimal agar penelitian ini dapat berjalan dengan lancar dan baik. Berikut berupa rincian dari setiap perangkatnya:

Tabel 3. 1 Perangkat Keras

No.	Perangkat	Ket/Deskripsi
1.	Laptop	<i>Asus VivoBook</i>
2.	Processor	<i>Intel Core i3</i>
3.	RAM	4GB
4.	Penyimpanan	1TB HDD + 256GB SSD

Tabel 3. 2 Perangkat Lunak

No.	Perangkat Lunak	Ket/Deskripsi
1.	<i>Windows 10 64-bit</i>	<i>Operating System.</i>
2.	<i>Python version 3.10.0</i>	<i>Tools untuk membangun model.</i>
3.	<i>Google Colab</i>	<i>Tools untuk menjalankan proses Analisis data dan Program.</i>
4.	<i>Scikit-learn</i>	<i>Library Python untuk melakukan analisis data.</i>
5.	Matplotlib dan seaborn	<i>Library Python untuk melakukan visualisasi data yang lebih interaktif.</i>
6.	Microsoft Office Word	<i>Tools untuk membuat sebuah hasil laporan.</i>

3.5 Waktu dan Tempat Penelitian

3.5.1 Tempat Penelitian

Penelitian ini dilakukan pada media sosial Platform X terhadap sentimen mengenai restoran siap saji Gacoan, khususnya yang mengandung ujaran kebencian dengan pengambilan serta pengolahan data menggunakan tools Google Colab dan alat bantu Library Python dan juga Scrapping Tools bernama Tweet-Harvest untuk melakukan proses scrapping data.

BAB IV HASIL DAN PEMBAHASAN

4.1. Hasil dan Proses Pengumpulan Data Tweet

Untuk tahap scrapping beberapa sel kode digunakan untuk mengambil data dari Twitter berdasarkan kata kunci dan rentang waktu yang telah ditentukan yaitu “Gacoan” mulai dari bulan Januari tahun 2024 hingga Mei 2024, seperti pada proses setup awalnya yaitu menginstal beberapa package agar dapat melakukan scrapping diantaranya package pandas yang digunakan untuk manipulasi data atau bisa dikatakan untuk membaca file kedalam dataframe dan kemudian menampilkannya serta node.js yang diperlukan untuk alat scrapping karena tweet-harvest dibangun dengan node.js. Node.js adalah tools untuk menjalankan JavaScript diluar browser yaitu dengan mengambil data secara otomatis dari suatu website menggunakan kode JavaScript. Salah satu bentuk potongan python code nya seperti pada gambar dibawah ini.

```
# import required Python package
!pip install pandas

# Install Node.js (because tweet-harvest built using Node.js)
!sudo apt-get update
!sudo apt-get install -y ca-certificates curl gnupg
!sudo mkdir -p /etc/apt/keyrings
!curl -fsSL https://deb.nodesource.com/gpgkey/nodesource-repo.gpg.key | sudo gpg --dearmor -o /etc/apt/keyrings/nodesource.gpg
```

Gambar 4. 1 Install Package Scrapping

Pada sel – sel berikutnya merupakan proses scrapping itu sendiri yang diawali dengan menentukan filename terlebih dahulu dengan format csv untuk menyimpan hasil scrapping yang dimana telah diinput dengan “search_keyword” yang telah ditentukan sebelumnya serta memberikan limit untuk jumlah tweet maksimum yang ingin diambil yaitu sebanyak 500 data dalam satu kali proses scrapping, dan pada tahap ini scrapping dilakukan sebanyak lima kali dengan limit data yang sama

sehingga total keseluruhan data yang diambil menjadi sekitar 2500 data tweet. Hal ini dikarenakan adanya keterbatasan pengambilan data dari Twitter atau Platform X maka dari itu mengapa perlu di lakukan proses ini secara berulang sebanyak lima kali untuk meminimalisir terjadinya error pada proses pengambilan.

```
#Scrapping Data

filename = 'FinalScrapping.csv'
search_keyword = 'gacoan since:'
limit = 500

!npx -y tweet-harvest@2.6.1 -o
```

Gambar 4. 2 Scrapping

Baris terakhir pada gambar di atas sendiri merupakan perintah untuk menjalankan tools tweet-harvest. Perintah ini menggunakan twitter_auth_token untuk mengakses dan mengambil tweet berdasarkan keyword dan limit yang ditentukan. Dan berikut adalah sebagian hasil proses scrapping yang telah dilakukan sebelumnya, terdapat waktu, id, username, dan postingan tweet.

	time	id	username	text
0	Wed Jan 31 23:59:34 +0000 2024	1752844099307331827	NaN	@loseangeless sebenarnya gabisa menyalahkan or...
1	Wed Jan 31 23:55:19 +0000 2024	1752843030670024988	NaN	pen makan mie gacoan
2	Wed Jan 31 23:53:31 +0000 2024	1752842576619766230	NaN	maaf ya kl ak segoblok ntu? https://t.co/cJ1oS...
3	Wed Jan 31 23:53:19 +0000 2024	1752842529450713434	NaN	@HyrYssc_2612 Gacoan gacoan gacoan.. Mauuuuuuuuu
4	Wed Jan 31 23:52:48 +0000 2024	1752842398085128200	NaN	@loseangeless @jdrunkbetch @liojuks kak ak joi...

Gambar 4. 3 Hasil Scrapping

4.2. PreProcessing Data

Bagian ini bertujuan untuk membersihkan dan mengubah data teks mentah menjadi format yang siap untuk dianalisis dan digunakan dalam model *machine learning*. Langkah awal yang dilakukan adalah memuat data hasil scrapping yang dimana data tersebut masih tidak terstruktur, maka dari itu perlu melakukan

pembersihan data atau preprocessing terlebih dahulu, langkah – langkah preprocessing yang dilakukan meliputi:

4.2.1. Case Folding

Bagian ini melakukan perubahan semua teks yang berada pada kolom “*text*” menjadi huruf kecil, proses ini dilakukan dengan cara mengiterasi setiap teks dikolom “*text*” serta memeriksanya apakah dalam bentuk *string* dan mengubahnya menjadi huruf kecil dengan menggunakan `.lower()` yang kemudian akan disimpan didalam daftar *case_folding* yang diberikan perintah untuk meletakkannya di kolom baru bernama “*case_folded*” di DataFrame.

```
case_folding = []
for w in df['text']:
    if isinstance(w, str):
        data = w.lower()
```

Gambar 4. 4 Case_folding

Untuk hasil pada proses *case_folding* ini bisa dilihat pada tabel dibawah ini:

Tabel 4. 1 Hasil case_folding

text	case_folded
@HyrYssc_2612 Gacoan gacoan	@hyryssc_2612 gacoan gacoan
gacoan.. Mauuuuuuuuu	gacoan.. mauuuuuuuuu

4.2.2. Cleaning

Proses cleaning dilakukan dengan mendefinisikan fungsi *clean_text* yang menggunakan modul bawaan python yaitu modul *re* atau Regular Expression untuk memodifikasi teks seperti menghapus elemen – elemen yang tidak diinginkan dari teks, contohnya seperti Tag ‘RT’ (Retweet), Mention Pengguna, angka – angka, hashtag, URL atau tautan web, serta pengulangan karakter yang berlebihan

misalnya ‘mauuu’ menjadi ‘mau’, kemudian yang terakhir penghapusan karakter non-alfanumerik dan karakter spesifik lainnya. Berikut beberapa potongan kodenya seperti pada gambar dibawah ini.

```
import re
import pandas as pd

# Function to remove non-alphanumeric characters, numbers, hashtags, and links
def clean_text(text):
    # Replace RT tag
    text = re.sub('RT\s', '', text)
    # Replace @_username
    text = re.sub('\B@\w+', '', text)
    # Remove numbers
    text = re.sub(r'\d+', '', text)
```

Gambar 4. 5 Cleansing

Kemudian untuk menghapus baris yang memiliki nilai kosong pada kolom “cleaned” dan “text” meskipun dalam output yang ditampilkan jumlah data tidak berubah, yang dimana artinya menunjukkan tidak ada baris kosong dikolom tersebut setelah melakukan *cleaning* dan menampilkan kolom “cleaned” yang telah dibuat untuk melihat efek *cleaning* seperti pada tabel dibawah ini.

Tabel 4. 2 Hasil Cleaning

text	case_folded	cleaned
@TukangParaf	@tukangparaf	wah biar ngulang
@ariesadhar Wah biar ngulang semester depan yaaa. Btw udah termasuk ekonomis sih uang kuliah s1 disana.	@ariesadhar wah biar ngulang semester depan yaaa. btw udah termasuk ekonomis sih uang kuliah s1 disana.	semester depan yaa btw udah termasuk ekonomis sih uang kuliah s disana

4.2.3. Tokenizing

Pada tahap ini, proses tokenisasi dilakukan menggunakan `nltk.download('punkt')` untuk mengunduh tokenizer kalimat dan kata dari NLTK yaitu singkatan dari Natural Language Toolkit yang merupakan library python untuk melakukan pengolahan NLP (Natural Language Processing). Berikut dibawah ini merupakan potongan kode python untuk proses tokenisasi.

```
import nltk
from nltk.tokenize import word_tokenize

# Tokenisasi teks pada kolom 'cleaned'
df['tokens'] = df['cleaned'].apply(word_tokenize)
```

Gambar 4. 6 Tokenisasi

Dengan mengimpor `word_tokenize` dari `nltk.tokenize` hasilnya nanti akan diterapkan pada kolom “*cleaned*” untuk memecah setiap teks menjadi daftar kata – kata individual (token) dan disimpan dalam kolom baru bernama “*tokens*”. Tampak adanya perbedaan dari kolom “*cleaned*” dengan kolom “*tokens*” yang membuktikan bahwa proses ini berhasil dilakukan seperti pada contoh tabel dibawah ini.

Tabel 4. 3 Hasil Tokenisasi

cleaned	tokens
pen makan mie gacoan	['pen', 'makan', 'mie', 'gacoan']

4.2.4. Normalisasi

Untuk melakukan normalisasi pada kata yaitu dengan cara mendefinisikan fungsi “`normalisasi_kata`” untuk mengambil daftar token dan kamus normalisasi sebagai input yang dimana fungsi ini memeriksa setiap token yang ada didalam kamus, apabila ditemukan dan sesuai maka token asli diganti dengan bentuk

normalisasinya sesuai dengan kamus, jika tidak ditemukan maka token asli akan dipertahankan.

```
def normalisasi_kata(tokens, kamus):
    normalized_tokens = []
    for token in tokens:
        if token.lower() in kamus:
            normalized_tokens.extend(kamus[token.lower()].split())
        else:
            normalized_tokens.append(token)
    return normalized_tokens

kamus_path = '/content/normalisasi.csv'
```

Gambar 4. 7 Tokenisasi

Kamus normalisasi yang digunakan berupa file dengan format CSV yang berisi kata asli (singkatan, slang, dll) dan bentuk normalnya, yang dimana fungsi `normalisasi_kata` ini diterapkan pada kolom “*tokens*” sebagai teks asli untuk proses normalisasi dan hasilnya disimpan pada kolom baru “*normalized*”. Hasil pada kolom “*normalized*” akan tampak seperti kata yang normal atau kata yang asli baik itu berupa singkatan maupun bahasa gaul contohnya seperti ‘yg’ menjadi ‘yang’.

Tabel 4. 4 Hasil Normalized

tokens	normalized
['maaf', 'ya', 'kl', 'ak', 'segoblok', 'ntu']	['maaf', 'ya', 'kalau', 'aku', 'segoblok', 'itu']

4.2.5. Stopword Removal

Untuk melakukan *Stopword Removal* perlu menggunakan `nlTK.download('stopwords')` untuk mengunduh daftar stopwords dalam bahasa Indonesia yang dimana dapat diketahui stopwords merupakan kata – kata umum yang biasanya tidak memiliki makna yang signifikan dalam melakukan analisis contohnya seperti ‘dan’, ‘atau’, ‘yang’, dan lainnya. Maka dari itu, yang perlu

dilakukan adalah dengan mendefinisikan fungsi `remove_stopwords` yang mengambil daftar token sebagai input. Terlihat pada potongan kode seperti dibawah ini.

```
# Mengambil daftar stopwords dalam bahasa Indonesia
stopwords_indonesia = set(stopwords.words('indonesian'))

# Fungsi untuk menghapus stopwords dari daftar token
def remove_stopwords(tokens):
    filtered_tokens = [token for token in tokens if token.lower() not in stopwords_indonesia]
    return filtered_tokens

# Menghapus stopwords dari kolom 'normalized'
df['stopwords_removed'] = df['normalized'].apply(remove_stopwords)
```

Gambar 4. 8 Stopword Removal

Fungsi ini memfilter token seperti hanya menyimpan token yang tidak ada di dalam daftar *stopwords* bahasa indonesia kedalam kolom baru bernama “*stopwords_remove*”, yang dimana tentunya fungsi “*remove_stopwords*” itu diterapkan pada kolom “*normalized*”. Dalam proses ini menggunakan modul *stopwords* dari dalam NLTK.

Tabel 4. 5 Hasil Stopwords

normalized	stopwords_remove
['maaf', 'ya', 'kalau', 'aku', 'segoblok', 'itu']	['maaf', 'ya', 'segoblok']

4.2.6. Stemming

Stemming merupakan tahap dimana menghapus imbuhan yang ada pada sebuah kata sekaligus mengembalikannya menjadi kata awal. Pada proses *stemming* menggunakan *library* Sastrawi dan Swifter yang digunakan untuk mempercepat operasi `apply` pada `pandas DataFrame`. Untuk membuat hasil stemmer menggunakan “*StemmerFactory*” dari Sastrawi yang dimana stemmer ini akan

mengubah kata – kata menjadi bentuk dasarnya (akar kata). Selain itu untuk menerapkan stemmer pada setiap token dalam daftar token perlu mendefinisikan fungsi “stem_tokens”.

```
# Membuat objek stemmer menggunakan INSET
stemmer_factory = StemmerFactory()
stemmer = stemmer_factory.create_stemmer()

# Fungsi untuk melakukan stemming pada token
def stem_tokens(tokens):
    stemmed_tokens = [stemmer.stem(token) for token in tokens]
    return stemmed_tokens

# Melakukan stemming pada kolom 'stopwords_removed'
df['stemmed'] = df['stopwords_removed'].swifter.apply(stem_tokens)
```

Gambar 4. 9 Stemming

Kemudian fungsi stem_tokens diterapkan pada kolom “stopwords_removed” menggunakan “swifter.apply” untuk efisiensi. Hasilnya akan disimpan dalam kolom baru lagi bernama “stemmed”. Dikarenakan hasil stemming masih berupa bentuk token, maka perlu dilakukan proses penggabungan kembali menjadi satu string per baris seperti pada contoh tabel dibawah ini.

Tabel 4. 6 Hasil Stemming

stopwords_removed	stemmed
['maaf', 'ya', 'segoblok']	maaf ya goblok

4.3. Labelling Metode Inset Lexicon Based

Untuk melakukan pelabelan kategori terhadap dataset ini menggunakan pendekatan lexicon-based berbasis INSET (Indonesia Sentiment Lexicon). Pendekatan ini mengandalkan daftar kata – kata berkonotasi positif dan negatif dalam bahasa Indonesia untuk menentukan kategori sentimen suatu tweet. Dataset yang digunakan merupakan hasil preprocessing dari data tweet yang telah melalui

berbagai tahap pembersihan mulai dari case folding, normalisasi, tokenisasi dan juga stemming. Hasil pada kolom “*stemmed*” akan digunakan untuk proses klasifikasi karena berisi kata – kata yang telah dibersihkan yang merupakan kata dasar agar dapat dilakukan proses analisis selanjutnya.

```
positive_lexicon = set(pd.read_csv(positive_url, sep='\t', header=None)[0])
negative_lexicon = set(pd.read_csv(negative_url, sep='\t', header=None)[0])
```

Gambar 4. 10 Kamus Lexicon

Lexicon dimuat dari dua file eksternal *positive.tsv* dan *negative.tsv* yang masing – masing berisi daftar kata – kata positif dan negatif dari INSET. Setiap tweet dihitung jumlah kemunculan kata – kata dari kamus positif dan negatif yang kemudian skor sentimen akan diperoleh dari selisih antara jumlah kata positif dan kata negatif, skor ini akan positif jika lebih banyak kata positif dan akan menjadi negative jika lebih banyak kata negatif. Dengan adanya ketentuan klasifikasi seperti jika skor lebih dari 0 maka “Non-Hatespeech” dan jika skor kurang dari 0 maka hasilnya “Hatespeech”.

```
if isinstance(text, str):
    positive_count = sum(1 for word in text.split() if word in positive_lexicon)
    negative_count = sum(1 for word in text.split() if word in negative_lexicon)
    sentiment_score = positive_count - negative_count

    if sentiment_score == 0:
        sentiment_score = random.choice([1, -1])

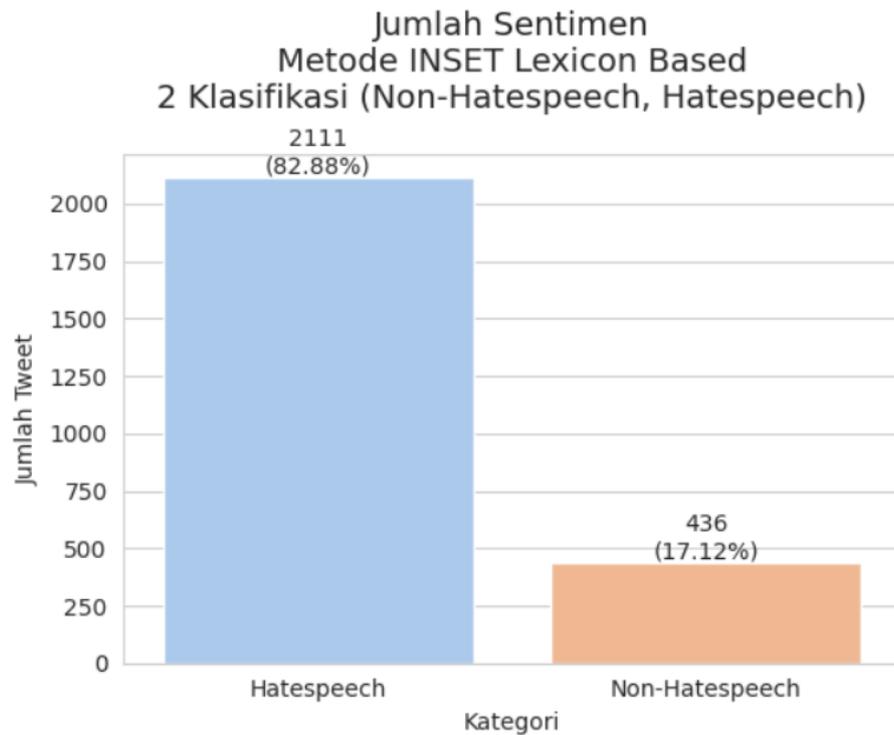
    if sentiment_score > 0:
        sentiment = 'Non-Hatespeech'
    else:
        sentiment = 'Hatespeech'

    return sentiment_score, sentiment
return 0, 'Hatespeech'
```

Gambar 4. 11 Fungsi Penentuan Skor

Dengan proses ini, setiap teks yang telah melalui proses preprocessing dalam dataset diberi label kategori “*Hatespeech*” atau “*Non- Hatespeech*” berdasarkan

kamus kata positif dan negatif yang digunakan. Hasilnya kemudian divisualisasikan kedalam bentuk bar plot untuk melihat distribusi jumlah tweet di setiap kategori sentimen. Hasilnya menunjukkan bahwa mayoritas tweet diklasifikasikan sebagai “*Hatespeech*” yaitu sekitar 82,88% dan sisanya Non-Hatespeech sebesar 17,12%.



Gambar 4. 12 Bar Plot Distribusi Sentimen

4.4. Split Data

Setelah data diberi label selanjutnya mempersiapkan data untuk melakukan proses pelatihan model yaitu dengan memisahkan data menjadi data latih dan data uji serta melakukan ekstraksi fitur menggunakan metode *CountVectorizer*. Pemisahan data dibagi menjadi dua bagian, yaitu Data latih sebesar 80% dan Data Uji sebesar 20%. Pembagian dilakukan menggunakan fungsi *train_test_split* dari Pustaka *sklearn.model_selection* dengan parameter *stratify=y*, untuk memastikan proporsi label tetap seimbang didata latih dan uji.

```

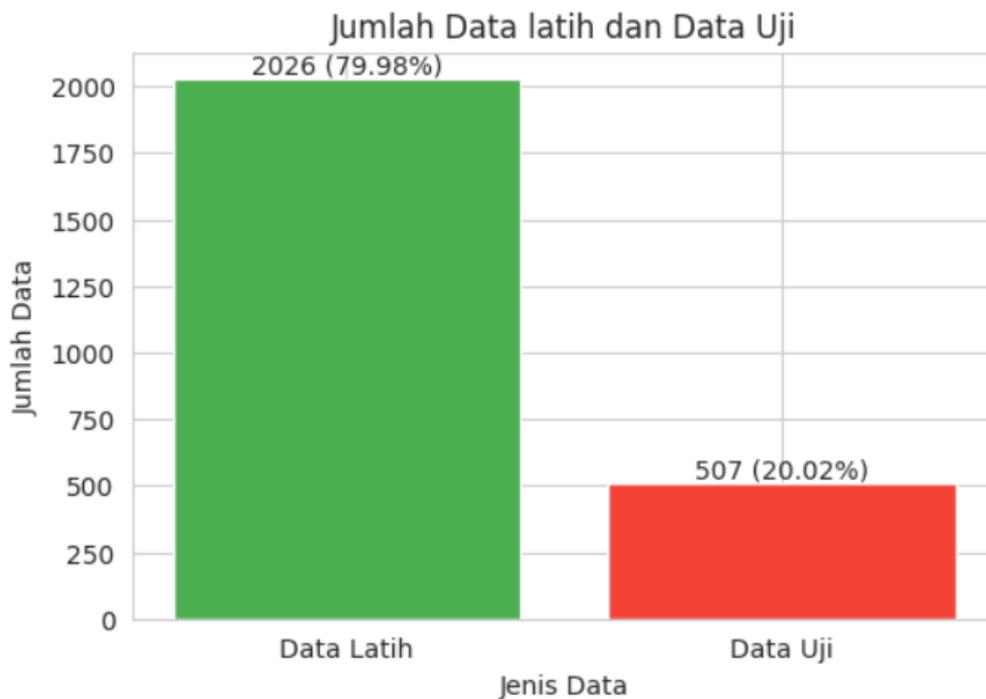
X = data['stemmed']
y = data['kategori']

# Memisahkan data menjadi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42, stratify=y)

```

Gambar 4. 13 Split Data

Untuk menentukan fitur X dan target Y yaitu mengambil kolom “*stemmed*” untuk dijadikan sebagai fitur (input) untuk model sedangkan kolom “Kategori” yang berisi label sentimen seperti “*Hatespeech*” dan “*Non-Hatespeech*” dijadikan sebagai target (output) yang akan diprediksi oleh model. Kemudian untuk “*random_state=42*” digunakan untuk memastikan bahwa pembagian data akan tetap sama setiap kali kode dijalankan sehingga hasilnya dapat direproduksi sedangkan penggunaan “*stratify=y*” untuk memastikan bahwa proporsi setiap kelas dalam data latih dan data uji proporsional dengan proporsi kelas dalam data asli ,argumen ini sangat berguna untuk dataset yang tidak seimbang. Dan untuk hasil proses Train-test_split adalah Data Latih sebesar 79,98% dan Data Uji 20,02%



Gambar 4. 14 Bar Plot Split Data

Setelah data dipisahkan, selanjutnya dilakukan proses ekstraksi fitur menggunakan metode Bag of Words melalui *CountVectorizer*. Metode ini mengubah teks menjadi representasi numerik berupa matriks sparse yang menunjukkan jumlah kemunculan kata (*term frequency*) dalam dokumen. Hasil dari ekstraksi fitur tersebut adalah `x_train_vec` (2026, 2898) dan `x_test_vec` (507, 2898) yang dimana artinya terdapat 2.898 kata unik (fitur) yang dihasilkan dari data latihan dan masing -masing tweet direpresentasikan sebagai vector dengan panjang 2.898.

```
vectorizer = CountVectorizer()
X_train_vec = vectorizer.fit_transform(X_train)
X_test_vec = vectorizer.transform(X_test)

X_train_vec.shape, X_test_vec.shape

Jumlah data latihan: 2026
Jumlah data uji: 507
=====
((2026, 2898), (507, 2898))
```

Gambar 4. 15 Vektorisasi Teks

Untuk “`X_train_vec`” adalah langkah ini mempelajari kosa kata dari data latihan dan mengubah teks menjadi vektor (representasi numerik) berdasarkan hitungan kemunculan setiap kata. Sedangkan “`X_test_vec`” digunakan untuk mengubah teks data uji menjadi vektor menggunakan kosakata yang sudah dipelajari dari data latihan. Hal ini penting untuk menghindari kebocoran data (*data leakage*) dari data uji ke data latihan. Setelah langkah ini, data teks sudah diubah menjadi representasi numerik yang siap digunakan untuk melatih model machine learning.

4.5. Implementasi Model Support Vector Machine

4.5.1. Pemilihan Parameter Optimal dengan GridSearchCV

Langkah pertama dalam proses implementasi model SVM adalah dengan melakukan pencarian parameter terbaik menggunakan GridSearchCV, parameter yang diuji antara lain parameter regularisasi, jenis kernel mana yang akan digunakan, kemudian *gamma* hanya jika kernel yang digunakan adalah “rbf”. Model SVM diinisialisasi terlebih dahulu kemudian dilakukan pencarian kombinasi terbaik dengan menggunakan cv (cross validation) menggunakan metrik f1-macro sebagai acuan evaluasi, ini digunakan dikarenakan adanya ketidak seimbangan data maka digunakan agar tidak bias ke kelas mayoritas dan kemampuan model dalam mengklasifikasikan semua kelas secara seimbang atau adil. Dan hasil dari proses ini adalah menggunakan kernel Linear, dengan nilai $C = 1$ dan $\text{gamma} = \text{scale}$.

 Best Parameters: {'C': 1, 'gamma': 'scale', 'kernel': 'linear'}

Gambar 4. 16 GridSearchCV

Artinya model terbaik menggunakan kernel Linear dikarenakan cocok jika data sudah cukup bisa dipisahkan tanpa perlu pemetaan dimensi tinggi, dalam kasus ini (hatespeech vs non-hatespeech) model bisa membedakan kelas hanya dengan garis lurus diruang fitur dan juga lebih sederhana dan cepat serta menghindari overfitting. Kemudian untuk C sebagai parameter regularisasi yaitu 1 karena nilai sedang artinya model seimbang antara tidak terlalu ketat terhadap data training (tidak overfit dan tidak terlalu longgar terhadap kesalahan (tidak underfit). Dan yang terakhir hasil gamma sebenarnya tidak mempengaruhi karena kernel yang digunakan adalah Linear. Akan tetapi, dari semua proses ini yang dibutuhkan hanya

untuk mengetahui kernel mana yang bagus untuk digunakan dalam implementasi model SVM ini.

4.5.2. Latih Model SVM

Untuk proses Latih model dikembangkan dengan menggunakan bahasa pemrograman python dengan bantuan library *Scikit-Learn*, dalam implementasinya terdapat beberapa alur seperti Inisialisasi model kemudian baru dapat di lakukan pelatihan dan pengujian. Model SVM di inisialisasi menggunakan kernel linear, ini dikarenakan telah melakukan hasil pencarian parameter terbaik melalui *GridSearchCV*, kernel ini memberikan performa terbaik pada data yang digunakan. Selain itu, parameter `probability = True` ditambahkan untuk memungkinkan prediksi probabilitas kelas dan `random_state=42` digunakan agar hasil yang diperoleh konsisten.

```
# Inisialisasi Model
models = {
    "SVM": SVC(kernel='linear', probability=True, random_state=42)
}

# Train Model / Latih Model
results = {}
for model_name, model in models.items():
    model.fit(X_train_vec, y_train)
    y_pred = model.predict(X_test_vec)
```

Gambar 4. 17 Latih Model SVM

Data yang telah diproses sebelumnya dibagi menjadi data latih dan data uji, yang kemudian digunakan untuk melatih dan menguji performa model. Proses pelatihan dilakukan menggunakan fungsi `.fit()` dengan menggunakan data latih yaitu `(X_train_vec, y_train)` artinya model belajar dari data ini agar bisa mengenali seperti apa karakteristik dari data “*hatespeech*” dan “*non-hatespeech*”, setelah

model selesai di latih, kemudian digunakan untuk prediksi terhadap data uji yaitu (X_{test_vec}) menggunakan fungsi `.predict()`, hasil dari prediksi ini yang nantinya akan dibandingkan dengan label asli untuk mengetahui apakah prediksinya benar atau tidak.

4.6. Evaluasi Model SVM

Evaluasi dilakukan untuk mengukur performa model berdasarkan metrik akurasi, confusion matrix dan classification report yang terdiri dari precision, recall, dan f1-score. Setelah model melakukan prediksi, hasil evaluasi akan disimpan didalam *dictionary* “results” untuk kemudian ditampilkan dan dianalisis lebih lanjut. Adapun metrik evaluasi yang digunakan adalah sebagai berikut:

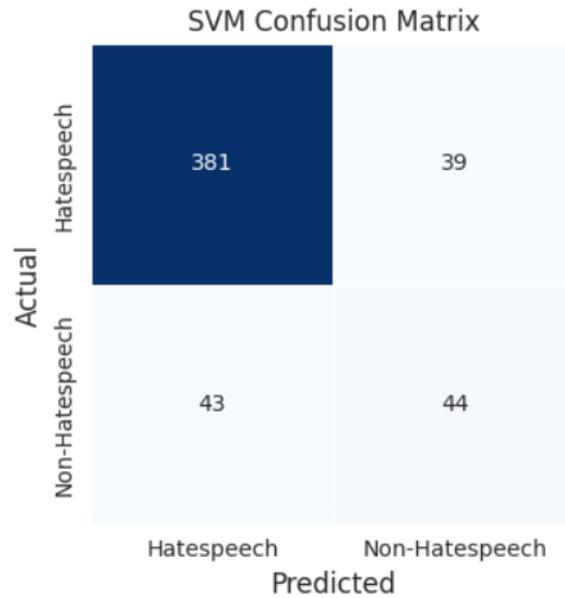
```
results[model_name] = {  
    "accuracy": accuracy_score(y_test, y_pred),  
    "classification_report": classification_report(y_test, y_pred, output_dict=True),  
    "confusion_matrix": confusion_matrix(y_test, y_pred)  
}
```

Gambar 4. 18 Evaluasi Model

4.6.1. Confusion Matrix

Matrix yang menunjukkan jumlah prediksi benar dan salah yang dibuat oleh Model terhadap masing – masing kelas. Matrix ini membantu melihat performa model dalam mengklasifikasikan setiap kelas seperti berapa banyak “*hatespeech*” yang benar diprediksi sebagai “*hatespeech*” dan berapa banyak “*non-hatespeech*” yang salah di prediksi sebagai “*hatespeech*”.

Confusion matrix for SVM:



Gambar 4. 19 Confusion Matrix Linear

Berdasarkan confusion matrix tersebut diperoleh bahwa dari total 507 data yang diklasifikasikan sebagai “*Hatespeech*” sebanyak 381 data berhasil diklasifikasikan dengan benar (*True Negative*) sementara 43 data “*Non-Hatespeech*” salah diklasifikasikan sebagai “*Hatespeech*” (*False Negative*). Sebaliknya dari total 83 data yang merupakan “*Non-Hatespeech*” kemudian sebanyak 44 data berhasil diklasifikasikan dengan benar (*True Positive*) sedangkan 39 data “*Hatespeech*” salah diklasifikasikan sebagai “*Non-Hatespeech*” (*False Positive*).

4.6.2. Perhitungan Accuracy, Precision, Recall dan F1-Score

Tabel 4. 7 Tabel Perhitungan Manual Confusion Matrix

Aktual	Prediksi Positif	Prediksi Negatif
Positif	44 <i>TP</i>	43 <i>FN</i>
Negatif	39 <i>FP</i>	381 <i>TN</i>

a. Accuracy

Akurasi mengukur proporsi prediksi yang benar dari seluruh data.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{44+381}{44+381+39+43} = \frac{425}{507} = 0.8385 / 83.85\%$$

b. Precision

Presisi mengukur seberapa akurat prediksi positif.

$$\text{Precision Positif} = \frac{TP}{TP + FP} = \frac{44}{44 + 39} = \frac{44}{83} = 0.5301 / 53.01\%$$

Presisi mengukur seberapa akurat prediksi negatif.

$$\text{Precision Negatif} = \frac{TN}{TN + FN} = \frac{381}{381 + 43} = \frac{381}{424} = 0.8986 / 89.86\%$$

c. Recall

Recall mengukur berapa banyak dari total kelas positif yang berhasil ditemukan.

$$\text{Recall Positif} = \frac{TP}{TP + FN} = \frac{44}{44 + 43} = \frac{44}{87} = 0.5057 / 50.57\%$$

Recall mengukur berapa banyak dari total kelas negatif yang berhasil ditemukan.

$$\text{Recall Negatif} = \frac{TN}{TN + FP} = \frac{381}{381 + 39} = \frac{381}{420} = 0.9071 / 90.71\%$$

d. F1-Score

$$\begin{aligned} \text{F1 Positif} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \times \frac{0.5301 \times 0.5057}{0.5301 + 0.5057} \\ &= 0.5177 / 51.77\% \end{aligned}$$

$$\begin{aligned} \text{F1 Negatif} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \times \frac{0.8986 \times 0.9071}{0.8986 + 0.9071} \\ &= 0.9028 / 90.28\% \end{aligned}$$

4.6.3. Classification Report

Evaluasi lanjutan terhadap performa model SVM dilakukan menggunakan *classification report* yang mencakup metrik precision, recall dan f1-score untuk masing – masing kelas yaitu “*Hatespeech*” dan “*Non-Hatespeech*”. Hasil dari *Classification Report* ini akan dijadikan perbandingan dengan hasil perhitungan manual akurasi, presisi, recall dan f1-score dari confusion matrix linear sebelumnya.

Classification Report for SVM:

	precision	recall	f1-score	support
Hatespeech	0.899	0.907	0.903	420.000
Non-Hatespeech	0.530	0.506	0.518	87.000
accuracy	0.838	0.838	0.838	0.838
macro avg	0.714	0.706	0.710	507.000
weighted avg	0.835	0.838	0.837	507.000

Gambar 4. 20 Hasil Classification Report

Pada kelas “*Hatespeech*” model menunjukkan performa yang sangat baik dengan nilai precision sebesar 0.899 artinya sekitar 89,9% dari seluruh prediksi “*Hatespeech*” adalah benar. Nilai Recall sebesar 0.907 menunjukkan bahwa 90,7% dari seluruh data “*Hatespeech*” berhasil terdeteksi dengan tepat oleh model. Kombinasi antara precision dan recall menghasilkan nilai f1-score sebesar 0.903 yang mencerminkan keseimbangan yang baik antara keduanya.

Sementara itu, pada kelas “*Non-Hatespeech*” performa model masih perlu ditingkatkan, kemudian precision sebesar 0.530, recall sebesar 0.506 dan f1-score sebesar 0.518. nilai – nilai yang lebih rendah ini menunjukkan bahwa model

memiliki kesulitan dalam mengenali kelas “*Non-Hatespeech*” yang kemungkinan disebabkan oleh ketidakseimbangan jumlah data antara kedua kelas.

Secara keseluruhan, model mencapai akurasi sebesar 83,3% yang berarti sekitar 84% dari seluruh prediksi model sesuai dengan label yang sebenarnya. Untuk memberikan gambaran lebih seimbang terhadap performa antar kelas digunakan dua jenis rata – rata, diantaranya:

a. Macro Average

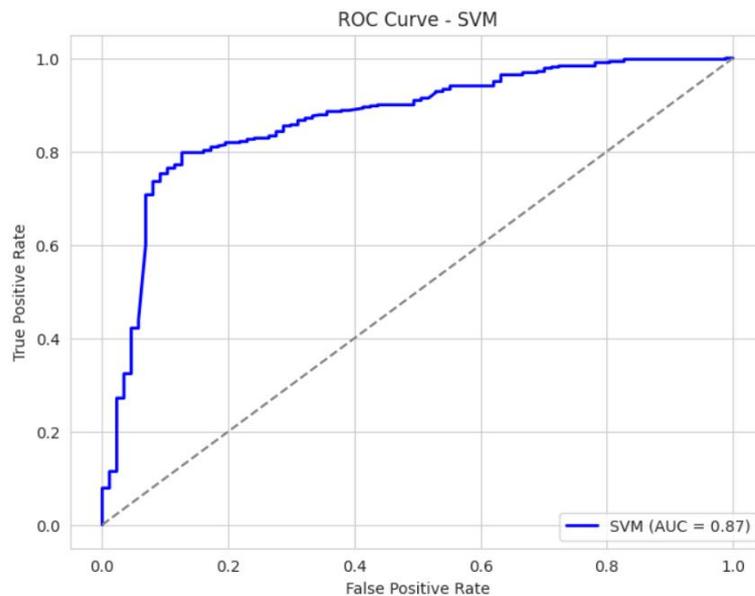
Yaitu rata – rata dari precision, recall dan f1-score tanpa memperhatikan jumlah sampel pada setiap kelas dengan masing – masing sebesar 0.714, 0.706, dan 0.710.

b. Weighted Average

Yaitu rata – rata tertimbang berdasarkan jumlah data per kelas dengan hasil precision 0.835, recall 0.8 dan f1-score 0.837.

4.6.4. ROC Curve SVM

Selain menggunakan *Confusion Matrix* dan *Classification Report*, evaluasi performa model SVM juga dilakukan dengan menggunakan Receiver operating characteristic (ROC) Curve dan Area Under the Curve (AUC). ROC Curve merupakan grafik yang menggambarkan hubungan antara *True Positive Rate* (Recall) dan *False Positive Rate* pada berbagai nilai ambang (*threshold*). Semakin mendekati sudut kiri atas grafik, semakin baik kinerja model dalam membedakan anatara kelas positif dan negatif.



Gambar 4. 21 ROC Curve SVM

Gambar diatas menunjukkan hasil ROC Curve untuk model SVM yang digunakan dalam klasifikasi teks “*Hatespeech*” dan “*Non-Hatespeech*”. Dari kurva tersebut, diperoleh nilai AUC sebesar 0.87 yang menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam membedakan kedua kelas tersebut. Secara umum, nilai AUC berada pada rentang 0 hingga 1. Proses ini penting untuk mengevaluasi kinerja model secara menyeluruh terutama pada kasus klasifikasi yang tidak seimbang seperti pada dataset ini.

Dalam proses ini ROC Curve harus dibangun menggunakan data uji bukan data latih. Hal ini dilakukan untuk menghindari overfitting dan memastikan bahwa penilaian performa model bersifat objektif serta mencerminkan kemampuan generalisasi terhadap data yang belum pernah dilihat sebelumnya. Hasil AUC 0.87 yang diperoleh bukanlah hasil dari proses pelatihan model, melainkan hasil evaluasi terhadap data baru yang tidak digunakan selama pelatihan.

4.6.5. WordCloud

Wordcloud adalah visualisasi yang menunjukkan frekuensi kata – kata dalam sebuah teks. Kata – kata yang lebih sering muncul akan ditampilkan dengan ukuran font yang lebih besar. Disini wordcloud dibuat untuk dua kategori sentiment yang diprediksi oleh model SVM yaitu “*Hatespeech*” dan “*Non-Hatespeech*”. Prediksi pada data latih dilakukan disini karena word cloud akan dibuat berdasarkan teks dari data latih yang dikelompokkan berdasarkan prediksi model. Kemudian membuat DataFrame baru (`X_train_df`) yang berisi dua kolom yaitu “*text*” (teks asli dari data latih sebelum di vectorize) dan kolom “*predicted_sentiment*” (label sentimen yang diprediksi oleh model untuk teks tersebut). Proses ini mempermudah pemfilteran teks berdasarkan hasil prediksi model.

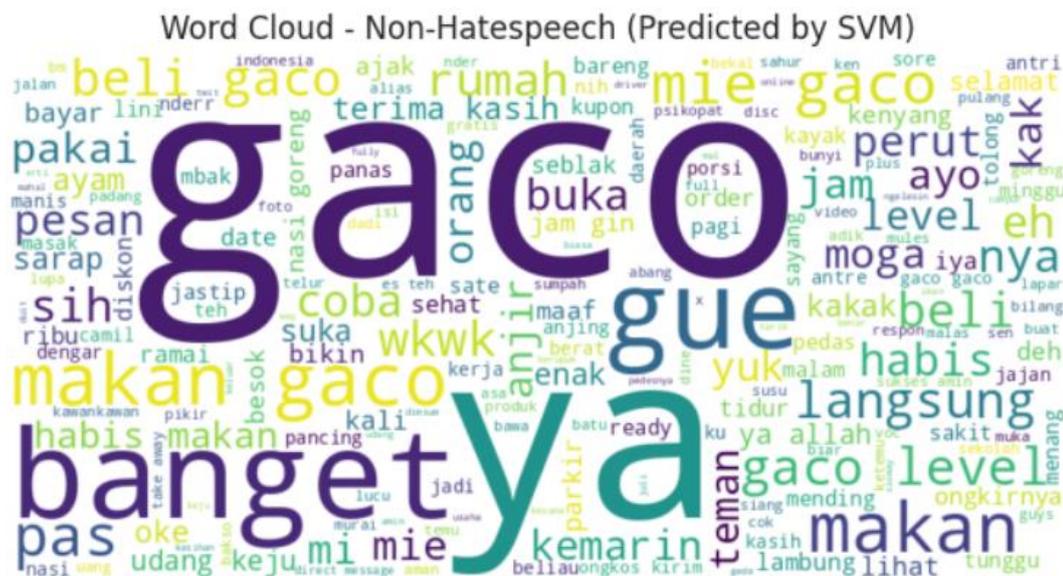
```
hatespeech_text = ' '.join(X_train_df[X_train_df['predicted_sentiment'] == 'Hatespeech']['text'])
```

Gambar 4. 22 Prediksi Hatespeech

Kode diatas untuk mengambil kolom yang bernilai “*hatespeech*” semua teksnya digabungkan menjadi satu string tinggal yang dipisahkan oleh spasi. String ini berisi semua teks dari data latih yang diprediksi oleh model sebagai “*Hatespeech*”. Sama halnya untuk teks yang diprediksi oleh model SVM sebagai kelas “*Non-Hatespeech*”.

```
non_hatespeech_text = ' '.join(X_train_df[X_train_df['predicted_sentiment'] == 'Non-Hatespeech']['text'])
```

Gambar 4. 23 Prediksi Non-Hatespeech



Gambar 4. 25 Word Cloud Non-Hatespeech

Kata yang paling dominan yang muncul pada wordcloud diantaranya adalah “gaco”, “ya”, “banget”, “makan”, “gue”, “mie”, “beli” tampil lebih besar menandakan kata – kata ini paling sering muncul dalam tweet yang dikategorikan sebagai tidak mengandung ujaran kebencian. Meskipun beberapa kata seperti “gaco”, “mie”, dan “makan” muncul di kedua wordcloud konteks penggunaannya berbeda. Artinya kata seperti “gaco” tidak selalu berarti *hatespeech*, karena model SVM mempelajari pola kata dan konteks sekitar, jadi kalau kata – kata negatif, umpatan atau keluhan muncul bersamaan maka tweet cenderung diklasifikasikan sebagai *hatespeech*. Sebaliknya jika kata yang sama muncul dengan kata – kata positif, ajakan, atau pengalaman biasa maka dikategorikan sebagai non-*hatespeech*.

Banyak istilah makanan dan aktivitas seperti “mie”, “keju”, “ayam”, “nasi goreng”, “minum”, “sarapan”, “habis” yang menggambarkan percakapan seputar pengalaman kuliner. Dalam wordcloud ini hampir tidak ditemukan kata kasar atau berkonotasi agresif seperti “anjing”, “sampah”, “jele”, dan lainnya yang sering menjadi indikator *hatespeech*.

BAB V

PENUTUP

5.1. Kesimpulan

Penelitian yang telah dilakukan mengenai analisis sentimen negative (Hatespeech) terhadap restoran siap saji Gacoan pada Platform X menggunakan algoritma Support Vector Machine (SVM) maka dapat disimpulkan beberapa hal seperti proses analisis sentimen berhasil dilakukan dengan tahapan yang sistematis dengan model SVM menggunakan kernel Linear menunjukkan performa yang cukup baik dalam mengklasifikasikan tweet kedalam kategori Hatespeech dan non-Hatespeech. Hasil evaluasi menunjukkan akurasi model sebesar 83,3% dengan nilai precision, recall dan f1-score pada kelas hatespeech yang tinggi yaitu masing – masing 0,899, 0,907 dan 0,903.

Sedangkan sebagian besar tweet tentang Gacoan pada Platform X diklasifikasikan sebagai Hatespeech yaitu sebesar 82,88% dan sisanya adalah Non-Hatespeech sebesar 17,12% berdasarkan hasil klasifikasi menggunakan metode INSET. Model SVM sendiri terbukti efektif digunakan dalam konteks klasifikasi biner terutama untuk mendeteksi ujaran kebencian secara otomatis dari media sosial dengan keunggulan mampu menangani data tidak terstruktur serta efisien dalam pemrosesan.

5.2. Saran

Berdasarkan hasil penelitian dan temuan yang diperoleh maka beberapa saran yang dapat disampaikan adalah untuk melakukan klasifikasi dengan menambahkan fitur agar hasil analisis sentimen menjadi lebih lengkap dan representatif terhadap

opini publik secara umum. Baik itu penambahan algoritma lain untuk dijadikan perbandingan performa model dalam mendeteksi ujaran kebencian ataupun penggabungan metode lexicon-based dan machine learning agar dapat memperoleh hasil klasifikasi yang lebih akurat dan relevan terutama dalam konteks bahasa informal dan gaya bahasa media sosial.

Untuk mengatasi ketidakseimbangan kelas dalam data penelitian selanjutnya dapat menggunakan teknik oversampling seperti SMOTE untuk meningkatkan performa model khususnya kelas minoritas. Penelitian ini diharapkan dapat menjadi referensi bagi pelaku usaha seperti restoran Gacoan dalam memahami opini public di media sosial sebagai bahan evaluasi layanan maupun strategi komunikasi digital.

DAFTAR PUSTAKA

- Ade Dwi Dayani, Yuhandri, & Widi Nurcahyo, G. (2024). Analisis Sentimen Terhadap Opini Publik pada Sosial Media Twitter Menggunakan Metode Support Vector Machine. *Jurnal KomtekInfo*, 11, 1–10.
<https://doi.org/10.35134/komtekinfo.v11i1.439>
- Ahnaf, M. I., & Suhadi. (2014). Isu-isu Kunci Ujaran Kebencian (Hate Speech): Implikasinya terhadap Gerakan Sosial Membangun Toleransi. *Jurnal Multikultural & Multireligius*, 13(3), 153–164. <http://www.youtube.com/>
- Al-Khowarizmi, Sari, I. P., & Maulana, H. (2023). Detecting Cyberbullying on Social Media Using Support Vector Machine: A Case Study on Twitter. *International Journal of Safety and Security Engineering*, 13(4), 709–714.
<https://doi.org/10.18280/ijssse.130413>
- Alviani, V., Alam, S., & Kurniawan, I. (2023). Analisis Sentimen Review Aplikasi Wetv Pada Platform Twitter Menggunakan Support Vector Machine. *STORAGE: Jurnal Ilmiah Teknik Dan Ilmu Komputer*, 2(3), 143–149.
<https://doi.org/10.55123/storage.v2i3.2351>
- Amelia, R. A., Sari, E. A., Zulfiqor, Almaas, A., Cahyo, G. N., & Nisa, K. P. (2024). Pola Interaksi Online Pengguna X terhadap Komunitas "Marah-Marah" di Platform X. *Jurnal Akademik Pengabdian Masyarakat*, 2(3), 54–67.
- Annisa, T. (2022). *Mengenal peran sentiment analysis beserta cara kerjanya*. Ekrut.Com.
<https://www.ekrut.com/media/sentiment-analysis-adalah>
- Armadianti, W., Syeh, A., Lastono, B., Putra, F. R., Kamil, I., Ghazi, A., & Rakhmawati, N. A. (2024). Analisis Sentimen Netizen Terhadap Personal Branding Elon Musk Pada Platform X Dengan Pendekatan Analisis Support Vector Machine. 9(1).
- Darussalam, & Arief, G. (2018). *Jurnal Resti*. *Resti*, 1(1), 19–25.
- Diani, N. R., Cantika, G. U., Zahra, A., Nurlaila, M., & Julianto, A. (2024). Analisis Twitter dalam Pembelajaran Bahasa Asing: Peran Konten Twitter Terhadap

- Pembelajaran Online. *Artikulasi: Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, 4(1), 5–11. <https://doi.org/10.17509/artikulasi.v4i1.68515>
- Fachri Muhammad, Fauziah, R. (2019). Jurnal Mantik Jurnal Mantik. *Mobile-Based National University Online Library Application Design*, 3(2), 10–19. <http://iocscience.org/ejournal/index.php/mantik/article/view/882/595>
- Farhan, N. M., & Setiaji, B. (2023). Indonesian Journal of Computer Science. *Indonesian Journal of Computer Science*, 12(2), 284–301. <http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135>
- Faruk Hussain Abdullah. (2023). *SENTIMENT ANALYSIS: A brief Explanation*. Medium. <https://medium.com/@farukhussainkbt/sentiment-analysis-a-brief-explanation-7be9833a452c>
- Irsyada, M. Z. (2024). *Pengaruh Harga , Promosi Dan Kualitas Produk Terhadap Keputusan Pembelian Mie Gacoan Cabang Tulungagung*. 1(1), 34–41.
- Kusumasari, D., & Arifianto, S. (2020). Makna Teks Ujaran Kebencian Pada Media Sosial. *Jurnal Komunikasi*, 12(1), 1. <https://doi.org/10.24912/jk.v12i1.4045>
- Laurensz, B., & Eko Sedyono. (2021). Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 10(2), 118–123. <https://doi.org/10.22146/jnteti.v10i2.1421>
- Nufairi, F., Pratiwi, N., & Herlando, F. (2024). Analisis Sentimen Pada Ulasan Aplikasi Threads Di Google Play Store Menggunakan Algoritma Support Vector Machine. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 9(1), 339–348. <https://doi.org/10.29100/jupi.v9i1.4929>
- Nugroho, A. S., Saputro, A., Suhardi, F. A., & Heikal, J. (2024). Dinamika Kreatif Manajemen Strategis Digital Marketing Analysis of Mie Gacoan Customer at Jakarta Using RFM and K-Means Clustering Methode. *Journalpedia*, 1(1), 223–240. <https://journalpedia.com/1/index.php/dkms/index>

- Nurjanah, T. S., & Insanudin, E. (2018). *Hack Database Website Menggunakan Python dan Sqlmap Pada Windows Hack Database Website Menggunakan Python dan Sqlmap Pada Windows Abstrak*. May, 0–7.
- Oryza Habibie Rahman, Gunawan Abdillah, & Agus Komarudin. (2021). Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 17–23. <https://doi.org/10.29207/resti.v5i1.2700>
- Pelajar, L. (2024). *SOSIO RELIGI : Jurnal Kajian Pendidikan Umum Pengaruh Social Media terhadap Perubahan Sosial di*. 22(1), 27–32.
- Purbowo, F., Azulfa, K. R., Al Hasan, N., & Masruroh, S. R. (2024). Keputusan Pembelian Konsumen Mie Gacoan di. *Jombang. Agrimics Journal*, 1(1), 33–43. <https://journal.sativapublishing.org/index.php/aj/>
- Purnamasari, D., Bayu, A., Desy, A., Fanka, W. A. P., Reza, A., Safrila, M., Yanda, O. N., & Hidayati, U. (2023). Pengantar Metode Analisis Sentimen. In M. A. S. dr. rer. nat. I Made Wiryana, SKom., S.Si. (Ed.), *Gunadarma Penerbit* (1st ed.). Gunadarma Penerbit. chrome-extension://efaidnbmninnibpcajpcgiclfndmkaj/https://penerbit.gunadarma.ac.id/wp-content/uploads/2023/09/Pengantar-Metode-Analisis-Sentimen_Detty-cs_Watermark.pdf
- Qisthiano, M. R., Ruswita, I., Prayesy, A., & Korespondensi, *. (2023). Implementasi Metode SVM dalam Analisis Sentimen Mengenai Vaksin dengan Menggunakan Python 3. *Online) Teknologi: Jurnal Ilmiah Sistem Informasi*, 13(1), 1–7.
- Reza, M., I. A. Q. Maududi, M. Rifki, A. Mujaddid, F. Ikhsanudin, Y. Adharani, S. N. Ambo, & N. Rosanti. (2022). Artificial Intelligence : Image Processing & Application with Python. *Seminar Nasional Pengabdian Masyarakat LPPM UMJ*, 1(1), 1–8. <http://jurnal.umj.ac.id/index.php/semnaskat>
- Sadat, A., Lawelai, H., & Suherman, A. (2022). Analisis Sentimen Media Sosial: Hate

- Speech Kepada Pemerintah di Twitter. *PRAJA: Jurnal Ilmiah Pemerintahan*, 10(1), 69–76.
- Septian, B. (2023). *Apa Itu Analisis Sentimen : Pengertian, Tipe, dan Cara Kerjanya*. Kazeo. <https://blog.kazeo.id/apa-itu-analisis-sentimen/>
- Setiadi, K. (2023). *Komputa : Jurnal Ilmiah Komputer dan Informatika Shopeefood Pada Media Sosial Twitter Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM)* *Komputa : Jurnal Ilmiah Komputer dan Informatika*. 12(1), 29–38.
- Siti Aisah, I., Irawan, B., & Suprpti, T. (2024). Algoritma Support Vector Machine (Svm) Untuk Analisis Sentimen Ulasan Aplikasi Al Qur'an Digital. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3759–3765. <https://doi.org/10.36040/jati.v7i6.8263>
- Sri, M. (2018). Fenomena ujaran kebencian Dampak Ujaran Kebencian. *Media Ilmiah Komunikasi Umat Beragama*, 10(1), 83–95. <http://artikata.com>
- Sujadi, H. (2022). Analisis Sentimen Pengguna Media Sosial Twitter Terhadap Wabah Covid-19 Dengan Metode Naive Bayes Classifier Dan Support Vector Machine. *INFOTECH Journal*, 8(1), 22–27. <https://doi.org/10.31949/infotech.v8i1.1883>
- Susanti, N. A., & Walid, M. (2022). *Klasifikasi Data Tweet Ujaran Kebencian Di Media Sosial*. 6(2), 538–543.
- Syahara, S. A., & Suryowati, B. (n.d.). *QUALITY OF SERVICE , PRODUCTS , PRICES , WORD OF MOUTH AND CUSTOMER SATISFACTION MIE GACOAN*. 91–100.
- Tami. (2023). Rebranding Twitter Ke X. In *SCRIBD*. <https://id.scribd.com/document/686051721/Rebranding-Twitter-ke-X>
- Wahyu Sejati, Ankur Singh Bist, & Amirsyah Tambunan. (2023). Pengembangan Analisis Sentimen dalam Rekayasa Software Engineering menggunakan tinjauan literatur sistematis. *Jurnal MENTARI: Manajemen, Pendidikan Dan Teknologi Informasi*, 2(1), 95–103. <https://doi.org/10.33050/mentari.v2i1.377>

Wati, R., & Ernawati, S. (2021). Analisis Sentimen Persepsi Publik Mengenai PPKM Pada Twitter Berbasis SVM Menggunakan Python. *Jurnal Teknik Informatika UNIKA Santo Thomas*, 06, 240–247. <https://doi.org/10.54367/jtiust.v6i2.1465>

LAMPIRAN



MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/III/2019
 Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003
<https://fkti.umsu.ac.id> fkti@umsu.ac.id [umsumedan](#) [umsumedan](#) [umsumedan](#) [umsumedan](#)

Berita Acara Pembimbingan Skripsi

Nama Mahasiswa : Alifa Az Zahra Fahresa Program Studi : Sistem Informasi
 NPM : 2102010006 Konsentrasi :
 Nama Dosen Pembimbing : Dr. Firahmi Rizky S.Kom, M.Kom Judul Penelitian : ANALISIS SENTIMEN
 HATE SPEECH TERHADAP ACCOUNT
 PADA PLATFORM X DENGAN MENGGUNAKAN
 SUPPORT VECTOR
 MACHINE (SVM)

Item	Hasil Evaluasi	Tanggal	Paraf Dosen
	Revisi Judul	14/01-2025	<i>[Signature]</i>
	Revisi mulai dari Bab 1 - 2	04/02-2025	<i>[Signature]</i>
	Revisi Bab 3	19/02-2025	<i>[Signature]</i>
	Acc Seminar Proposal	21/02-2025	<i>[Signature]</i>
	revisi hasil Seminar Proposal	11/03-2025	<i>[Signature]</i>
	revisi Bab 4	11/07-2025	<i>[Signature]</i>
	revisi Bab 5	12/07-2025	<i>[Signature]</i>
	Acc Sidang	14/07-2025	<i>[Signature]</i>

Medan, 11 Juli 2025

Disetujui oleh :

Dosen Pembimbing

Diketahui oleh :

Ketua Program Studi
 Sistem Informasi

[Signature]

(Martiano S.Pd, S.Kom.,M.Kom)

(Dr. Firahmi Rizky, S.Kom., M.Kom)



UMSU

Unggul | Cerdas | Terpercaya
Bila menulis surat ini agar direspon, nomor dan tanggalnya

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/III/2019
Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003
<https://fkti.umsu.ac.id> fkti@umsu.ac.id [f/umsumedan](https://www.facebook.com/umsumedan) [ig/umsumedan](https://www.instagram.com/umsumedan) [in/umsumedan](https://www.linkedin.com/company/umsumedan) [yt/umsumedan](https://www.youtube.com/channel/UCumsumedan)

**PENETAPAN DOSEN PEMBIMBING
PROPOSAL/SKRIPSI MAHASISWA
NOMOR : 968/IL3-AU/UMSU-09/F/2024**

Assalamu'alaikum Warahmatullahi Wabarakatuh

Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara, berdasarkan Persetujuan permohonan judul penelitian Proposal / Skripsi dari Ketua / Sekretaris.

Program Studi : Sistem Informasi
Pada tanggal : 16 Desember 2024

Dengan ini menetapkan Dosen Pembimbing Proposal / Skripsi Mahasiswa.

Nama : Alifa Az Zahra Fahreza
NPM : 2109010006
Semester : VII (Tujuh)
Program studi : Sistem Informasi
Judul Proposal / Skripsi : Pendeteksian Hate Speech Gacoan pada Platform Twitter dengan Sentimen Analisis Menggunakan SVM

Dosen Pembimbing : Dr. Firahmi Rizky, S.Kom.,M.Kom.

Dengan demikian di izinkan menulis Proposal / Skripsi dengan ketentuan

1. Penulisan berpedoman pada buku panduan penulisan Proposal / Skripsi Fakultas Ilmu Komputer dan Teknologi Informasi UMSU
2. Pelaksanaan Sidang Skripsi harus berjarak 3 bulan setelah dikeluarkannya Surat Penetapan Dosen Pembimbing Skripsi.
3. **Proyek Proposal / Skripsi** dinyatakan " **BATAL** " bila tidak selesai sebelum Masa Kadaluarsa tanggal : **16 Desember 2025**
4. Revisi judul.....

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Ditetapkan di : Medan
Pada Tanggal : 15 Jumadil Akhir 1446 H
16 Desember 2024 M

Dekan

Dr. A. Khoirizmi, S.Kom., M.Kom.
NIDN : 0127099201



Cc. File

