

**ANALISIS SENTIMEN MULTI SOSIAL MEDIA TENTANG PELAYANAN
TENAGA MEDIS MENGGUNAKAN METODE KLASIFIKASI
NAÏVE BAYES DAN SVM**

SKRIPSI

CINDY HARTANTI NAINGGOLAN
181401036



PROGRAM STUDI S1 ILMU KOMPUTER
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA
MEDAN
2022

**ANALISIS SENTIMEN MULTI SOSIAL MEDIA TENTANG PELAYANAN
TENAGA MEDIS MENGGUNAKAN METODE KLASIFIKASI *NAÏVE*
BAYES DAN SVM**

SKRIPSI

Diajukan untuk melengkapi tugas dan memenuhi syarat memperoleh ijazah
Sarjana Ilmu Komputer

CINDY HARTANTI NAINGGOLAN

181401036



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

2022

PERSETUJUAN

Judul : ANALISIS SENTIMEN MULTI SOSIAL
MEDIA TENTANG PELAYANAN TENAGA
MEDIS MENGGUNAKAN METODE
KLASIFIKASI NAÏVE BAYES DAN SVM

Kategori : SKRIPSI

Nama : CINDY HARTANTI NAINGGOLAN

Nomor Induk Mahasiswa : 181401036

Program Studi : SARJANA (S-1) ILMU KOMPUTER

Fakultas : ILMU KOMPUTER DAN TEKNOLOGI
INFORMASI UNIVERSITAS SUMATERA
UTARA

Komisi Pembimbing :

Pembimbing 2

Pembimbing 1

Amer Sharif S.Si, M.Kom

NIP. 19691021 202101 1 001

Dr. Amalia, S.T., M.T.

NIP. 19781221 201404 2 001

Diketahui/disetujui oleh

Program Studi S-1 Ilmu Komputer

Ketua

Dr. Amalia, S.T., M.T.

NIP. 19781221 201404 2 001

PERNYATAAN

ANALISIS SENTIMEN MULTI SOSIAL MEDIA TENTANG PELAYANAN TENAGA MEDIS MENGGUNAKAN METODE KLASIFIKASI *NAÏVE BAYES* DAN SVM

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

Medan, 22 September 2022

Cindy 181401036

PENGHARGAAN

Segala puji syukur kehadiran Tuhan Yang Maha Esa yang telah melimpahkan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Analisis Sentimen dari Multi Sosial Media Tentang Pelayanan Tenaga Medis Menggunakan Metode Klasifikasi *Naïve Bayes* dan SVM”.

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi S-1 Ilmu Komputer Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara.

Pada kesempatan ini, saya ingin mengucapkan terimakasih kepada pihak yang telah banyak membantu baik secara langsung maupun tidak langsung hingga skripsi ini bisa saya selesaikan.

Ucapan terimakasih tersebut saya tujukan kepada :

1. Bapak Dr. Muryanto Amin, S.Sos., M.Si. selaku Rektor Universitas Sumatera Utara.
2. Ibu Dr. Maya Silvi Lydia, B.Sc, M.Sc selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara.
3. Ibu Dr. Amalia ST., M.T. selaku Ketua Program Studi S-1 Ilmu Komputer Universitas Sumatera Utara sekaligus selaku Dosen Pembimbing I yang telah memberikan saran dan masukan hingga skripsi ini terselesaikan.
4. Bapak Amer Sharif S.Si, M.Kom. selaku Dosen Pembimbing II yang telah memberikan bimbingan dan masukan hingga skripsi ini dapat terselesaikan.
5. Seluruh dosen dan staf pengajar di Program Studi S-1 Ilmu Komputer Universitas Sumatera Utara yang telah memberikan ilmu pengetahuan yang bermanfaat selama saya mengikuti pendidikan di S-1 Ilmu Komputer.
6. Orangtua penulis, Bonar Nainggolan dan Nurmalela Marpaung yang telah memberikan banyak motivasi, dukungan, bantuan serta doa

kepada saya selama menjalani masa perkuliahan hingga penyusunan skripsi. Abang dan Adik tersayang, Niko dan Hana yang telah mendoakan dan menyemangati dalam mengerjakan skripsi ini.

7. Kak Bambang, yang turut membantu dalam pengerjaan skripsi.
8. Kak Yanti, yang turut membantu dalam pengerjaan skripsi.
9. Seluruh teman-teman Stambuk 2018 khususnya KOM C yang telah menemani dan menyemangati selama kuliah.

Semoga Tuhan Yang Maha Esa memberikan rahmat dan karunia-Nya kepada semua pihak yang telah memberikan bantuan serta dukungan kepada saya dalam menyelesaikan skripsi ini.

Medan, 22 September 2022

Penulis,

Cindy
Hartanti
Nainggolan

ABSTRAK

Perkembangan teknologi pada bidang informasi telah membuat berbagai aplikasi media sosial bermunculan seperti halnya Facebook, Twitter, Tiktok dan lain-lain. Banyaknya informasi yang tersebar termasuk komentar-komentar pengguna sosial media sulit untuk mengetahui kecenderungan apakah hal tersebut termasuk negatif atau positif, untuk dapat mengetahui kecenderungan dari banyaknya komentar dapat dilakukan dengan analisis sentimen. Analisis sentimen memiliki beberapa tahapan penting yakni tahapan preprocessing dan klasifikasi. Beberapa metode machine learning yang mempunyai tingkat keberhasilan klasifikasi terbaik diantaranya *Naïve Bayes* dan SVM. Hasil dari penelitian ini ialah mengklasifikasikan komentar pada sosial media Tiktok, Facebook dan Twitter berupa sentimen positif, negatif, netral. Untuk mengetahui kualitas dari proses analisis sentiment menggunakan algoritma *Naïve Bayes* dan Support Vector Machine. Hasil dari akurasi yang didapat pada media sosial Tiktok adalah 46% pada percobaan dengan menggunakan algoritma *Naïve Bayes*, pada percobaan dengan menggunakan algoritma *Support Vector Machine* akurasi yang didapat sebesar 91%. Hasil dari akurasi yang didapat pada media sosial Facebook adalah 37% pada percobaan algoritma *Naïve Bayes*, pada percobaan dengan menggunakan algoritma Support Vector Machine akurasi yang didapat sebesar 71%. Hasil dari akurasi yang didapat pada media sosial Twitter adalah 63% pada percobaan algoritma *Naïve Bayes*, pada percobaan dengan menggunakan algoritma Support Vector Machine akurasi yang didapat sebesar 67%.

Kata Kunci: Analisis Sentimen, *Naive Bayes*, *Support Vector Machine*, Tiktok, Facebook, Twitter.

**SENTIMEN ANALYSIS OF MULTI SOCIAL MEDIA ABOUT
MEDICAL SERVICES USING CLASSIFICATION
METHOD NAÏVE BAYES AND SVM**

ABSTRACT

The development of technology in the field of information has made various social media applications appear such as Facebook, Twitter, Tiktok and others. The amount of information that is spread including comments from social media users is difficult to know whether this trend is negative or positive, to be able to know the tendency of the number of comments can be done by sentiment analysis. Sentiment analysis has several important stages, namely the preprocessing and classification stages. Some machine learning methods that have the best classification success rate include Naïve Bayes and SVM. The results of this study are to classify comments on social media Tiktok, Facebook and Twitter in the form of positive, negative, neutral sentiments. To determine the quality of the sentiment analysis process using the Naïve Bayes algorithm and Support Vector Machine. The results of the accuracy obtained on Tiktok social media are 46% in experiments using the Naïve Bayes algorithm, in experiments using the Support Vector Machine algorithm the accuracy obtained is 91%. The results of the accuracy obtained on Facebook social media are 37% in the Naïve Bayes algorithm experiment, in experiments using the Support Vector Machine algorithm the accuracy is 71%. The results of the accuracy obtained on Twitter social media are 63% in the Naïve Bayes algorithm experiment, in experiments using the Support Vector Machine algorithm the accuracy obtained is 67%.

Key Word: *Sentiment Analysis, Naive Bayes, Support Vector Machine, Tiktok, Facebook, Twitter.*

DAFTAR ISI

	Halaman
PERNYATAAN.....	4
PENGHARGAAN	5
ABSTRAK	7
<i>ABSTRACT</i>	8
DAFTAR ISI.....	9
BAB 1 PENDAHULUAN	12
1.1 Latar Belakang	12
1.2 Rumusan Masalah	14
1.3 Batasan Masalah.....	15
1.4 Tujuan Penelitian.....	15
1.5 Manfaat Penelitian.....	15
1.6 Metodologi Penelitian	16
1.7 Sistematika Penulisan.....	16
BAB 2 LANDASAN TEORI.....	18
2.1 Analisis Sentimen.....	18
2.2 <i>Scraping</i>	19
2.3 <i>Data Scraping</i>	19
2.4 <i>Scraping</i> Sosial Media.....	21
2.4.1 Twitter	21
2.4.2 Facebook	21
2.4.3 Tiktok	21
2.5 Tenaga Medis	22
2.6 <i>Naïve Bayes</i>	23

2.7	<i>Support Vector Machine</i>	24
2.8	Penelitian yang Relevan	25
BAB 3	ANALISIS DAN PERANCANGAN SISTEM	28
3.1	Analisis Sistem	28
3.1.1	Analisis Masalah	28
3.1.2	Analisis Kebutuhan	29
3.2	General Arsitektur Sistem	30
3.2.1	<i>Scraping Data</i>	31
3.2.2	<i>Filtering</i>	33
3.2.3	<i>Label In Dataset</i>	33
3.2.4	<i>Preprocessing</i>	33
3.2.5	Klasifikasi Naïve Bayes dan SVM	33
3.2.6	Training Data	33
3.2.7	Evaluate Data	34
3.2.8	Komparasi Akurasi Naïve Bayes dan SVM	34
3.3	Perancangan Sistem Analisis Sentimen	34
3.3.1	Algoritma Naïve Bayes	34
3.3.2	Algoritma Support Vector Machine	49
BAB 4	IMPLEMENTASI DAN PENGUJIAN SISTEM	71
4.1	Implementasi Sistem	71
4.1.1	Spesifikasi Kebutuhan Perangkat Keras	71
4.1.2	Spesifikasi Kebutuhan Perangkat Lunak	71
4.2	Implementasi Sistem Analisis Sentimen	72
4.2.1	<i>Scraping</i>	72
4.2.2	<i>Preprocessing</i>	73

4.3	Pengujian Sistem	74
4.3.1	Confussion Matriks Facebook	74
4.3.2	<i>Confussion Matriks</i> Tiktok.....	85
4.3.3	Confussion Matriks Twitter	95
BAB 5	KESIMPULAN DAN SARAN	108
5.1	Kesimpulan.....	108
5.2	Saran	108
DAFTAR PUSTAKA		109

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi pada bidang informasi telah membuat berbagai aplikasi media sosial bermunculan seperti halnya Facebook, Twitter, Tiktok dan lain-lain. Menurut Andreas Kaplan dan Michael Heinlein, mendefinisikan “media sosial adalah suatu pengelompokan *software* atau perangkat lunak berbasis Internet yang berada di atas dasar ideologi dan teknologi Web 2.0, serta yang dapat memungkinkan *user* atau pengguna untuk penciptaan dan pertukaran “user-generated content” (Kaplan & Haenlein, 2019).

Dengan adanya media sosial setiap orang dapat saling berbagi informasi terhadap orang lain tanpa harus bertemu satu dengan yang lainnya dan juga memiliki kebebasan untuk mengemukakan pendapat. Pengguna media sosial pengguna dapat mempengaruhi hal buruk pada pengguna lain dengan membuat dan menyebarkan informasi yang bersifat tuduhan, fitnah, berita hoax, maupun SARA. “Dalam media sosial dikenal istilah ucapan kebencian atau dikenal dengan *Hate Speech*, yang makin populer saat ini, hal ini disebabkan gesekan atau perbedaan yang mewakili kelompok-kelompok tertentu baik Suku, Agama, Ras, Etnis, Golongan, (Rahman, 2016)”. Atas dasar berbagai permasalahan di media sosial tersebut akhirnya pemerintah Indonesia membuat aturan terkait berbagai kejahatan yang terjadi di sosial media (Pemerintah Indonesia, 2008). Banyaknya informasi yang tersebar termasuk komentar-komentar pengguna sosial media sulit untuk mengetahui kecenderungan apakah hal tersebut termasuk negatif atau positif, untuk dapat mengetahui kecenderungan dari banyaknya komentar dapat dilakukan dengan analisis sentimen.

Analisis sentimen merupakan suatu metode untuk melihat pandangan masyarakat terhadap produk dan jasa. Saat ini opini atau sentimen masyarakat dapat dilihat berbagai *platform* media sosial seperti Twitter, TikTok, Instagram, Facebook, Youtube dan lain sebagainya. Pengumpulan data dari berbagai media sosial ini mempunyai tantangan tersendiri karena masing-masing *platform*

memiliki kebijakan yang berbeda. Sebagai contoh media sosial Twitter menyediakan *Application Programming Interface* (API) dalam mengakses API maupun fungsi twitter itu sendiri. Dalam hal *crawling tweet* ada beberapa ketentuan yang perlu diperhatikan seperti batasan 2.400 *tweet* per 15 menit (Twitter, 2020). Media sosial Facebook memiliki aturan yang berbeda dimana dibutuhkan *written permission* sebelum meng-*crawling* data seseorang sama halnya dengan saat menginstall apps baru di *smartphone*. Dengan kata lain, selama tidak melakukan *crawling* data personal seseorang atau semua postingan yang bersifat pribadi, hal tersebut diizinkan sesuai dengan peraturan GDPR (*General Data Protection Regulation*), sedangkan untuk TikTok, *developer* TikTok menyediakan API untuk setiap akun sehingga para scrapers bisa mengambil data dari beberapa akun yang berbeda.

Mengumpulkan data opini dari berbagai media sosial tidak mudah dan membutuhkan observasi lebih lanjut. Maka dapat dipahami jika kecenderungan peneliti analisis sentimen menggunakan satu sumber data dari satu media sosial saja. Seperti Penelitian oleh (Osaldi, Gregorius Bryan. 2021) dan penelitian yang dilakukan oleh (Noviyanti, Wahdatun. 2018) hanya menggunakan Twitter sebagai sumber data. Penelitian lainnya hanya menggunakan sumber data dari facebook seperti penelitian oleh (Maisarah, Meissy Ayu. 2020) yang memanfaatkan komentar dari fanpage kandidat presiden dan penelitian yang dilakukan oleh (Sidiq dkk. 2020) yang memanfaatkan Grup Facebook Game Online sebagai sumber datanya. Penelitian ini bertujuan untuk membangun analisis sentiment dengan sumber data dari berbagai *platform* sosial media seperti Twitter, Facebook, dan Tiktok. Studi kasus dari penelitian ini mengenai pelayanan tenaga medis di sosial media dengan algoritma klasifikasi *Naïve Bayes* dan *Support Vector Machine* (SVM). Alasan pemilihan topik kesehatan karena kesehatan merupakan hal yang paling mendasar dan sangat penting dalam kehidupan manusia. Kepuasan masyarakat terhadap layanan kesehatan publik mengacu kepada aspek persepsi atau sikap masyarakat yang mana untuk memenuhi harapan hidup. Pada penelitian ini, klasifikasi review tenaga medis dari berbagai sosial media dibagi menjadi polaritas positif, negatif dan netral.

Penelitian terkait analisis sentimen sebelumnya pernah dilakukan yaitu oleh (Saepulrohman, Saepudin, & Gustian, 2021), terkait analisis sentimen pengguna Whatsaspp menggunakan Algoritma *Naïve Bayes* dan *Support Vector Machine*, dataset yang diambil adalah berupa komentar pengguna aplikasi Whatsapp. Berdasarkan hasil analisis kepuasan pengguna aplikasi WhatsApp pada *play store* dengan algoritma NB dan SVM dapat diketahui bahwa hasil akhir nilai akurasi NB adalah sebesar 70,40% dan SVM sebesar 77,00% dan nilai AUC pada algoritma NB sebesar 0,585 dan SVM adalah 0,876. SVM memiliki tingkat akurasi yang unggul 6,6%, daripada NB. Maka metode SVM dapat digunakan pada karakteristik data yang sama.

Peneliti memilih metode *Naïve Bayes* dan *Support Vector Machine* (SVM). Dua metode ini dipilih karena merupakan salah satu algoritma klasifikasi yang sederhana namun memiliki kemampuan dan akurasi tinggi. Metode SVM sebagai pembanding dengan *Naïve Bayes* karena merupakan suatu teknik untuk melakukan prediksi, baik prediksi dalam kasus regresi maupun klasifikasi. Multimedia sosial digunakan dalam penelitian ini guna mengetahui perbedaan kecenderungan opini dari masing-masing media sosial terkait topik yang sama.

Manfaat dari penelitian ini yaitu variasi data dari berbagai sosial media ini, tentunya diharapkan dapat meningkatkan akurasi klasifikasi sentimen karena jumlah data yang digunakan lebih banyak dan bervariasi dibandingkan jika hanya mengambil data dari satu sumber saja.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang ada diketahui bahwa sulit untuk mengetahui kecenderungan opini dari banyaknya komentar yang ada di media sosial terkait pelayanan tenaga medis, perbedaan media sosial tentu akan mendapatkan opini yang berbeda pula. Penelitian ini bertujuan mengumpulkan data opini dari berbagai media sosial tentang pelayanan kesehatan yang kemudian dilakukan analisis sentimen dengan metode klasifikasi *Naïve Bayes* dan *Support Vector Machine* untuk dapat mengetahui kecenderungan opini dari masing-masing media sosial.

1.3 Batasan Masalah

Dalam melakukan penelitian ini, penulis membatasi ruang masalah yang akan diteliti. Batasan masalahnya antara lain :

1. Penelitian ini memakai variasi sosial media seperti Twitter, Tiktok, Facebook dengan bantuan *library* API Developer.
2. Dataset menggunakan bahasa Indonesia
3. Bahasa pemrograman yang digunakan ialah Python 3.7.6
4. Hanya membagikan pendapat menjadi 3 klasifikasi yakni positif, negatif dan netral.
5. Menggunakan analisis sentimen dengan algoritma *Naïve Bayes* dan *Support Vector Machine*.
6. Tidak membedakan dataset baik yang anonim, robot, dst.
7. Objek yang diteliti adalah pekerja tenaga medis.
8. Aspek dari komentar yang diteliti yakni Pelayanan, Empati, Figure, Price dan Umum.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan dari penelitian ini adalah melakukan analisis sentimen terhadap pelayanan kesehatan dalam bidang medis berdasarkan kutipan ataupun pendapat masyarakat di berbagai sosial media, seperti Tiktok, Facebook, dan Twitter dengan menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine*.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah :

1. Memperoleh informasi tentang kecenderungan sentimen masyarakat terhadap bagaimana pelayanan pemerintah dalam bidang medis yang mencakup aspek yakni figure, pelayanan, empati dan hal umum.
2. Hasil penelitian dapat dijadikan salah satu referensi dalam memberikan tindak pencegahan kalau adanya berita *hoax* yang dapat merugikan image

pemerintah.

3. Mengetahui perbedaan sentimen analisis pelayanan pemerintah bidang medis divariasi sosial media dengan menggunakan metode *Naïve Bayes* dan *Support Vector Machine*.

1.6 Metodologi Penelitian

Penelitian yang dilakukan dalam penelitian ini adalah :

1. Studi Pustaka Pada tahap ini dilakukan pengumpulan referensi yang berkaitan dengan penelitian. Referensi yang digunakan berasal dari buku, jurnal, artikel, e-book, makalah, skripsi maupun situs internet. Pencarian referensi harus berkaitan dengan bagaimana pelayanan pemerintah bidang medis, analisis sentimen, algoritma *Naïve Bayes* maupun *SVM*.
2. Analisa dan Perancangan Pada tahap ini penulis melakukan analisa terhadap apa saja yang nantinya akan dibutuhkan dalam penelitian untuk dirancang dalam sebuah model diagram.
3. Implementasi Pada tahap ini, penelitian akan diimplementasikan menggunakan bahasa pemrograman Python berdasarkan hasil yang telah didapat.
4. Pengujian Pada tahap ini, sistem yang telah dirancang akan dilakukan analisis sentimen terhadap perilaku masyarakat kepada pelayanan publik tenaga medis di *platform* sosial Media Twitter, Tiktok dan Facebook apakah bersifat positif, negatif maupun netral sehingga pemerintah dapat menjadikan hasil yang didapat sebagai acuan untuk meningkatkan kualitas pelayanannya dalam bidang medis.
5. Dokumentasi Pada tahap ini, dilakukan dokumentasi tahap demi tahap sampai dengan pengujian dalam bentuk skripsi.

1.7 Sistematika Penulisan

Sistematika dari penyusunan skripsi ini terdiri dari beberapa bagian yang paling utama untuk diketahui yaitu sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini berisi tahapan awal penelitian yaitu dimulai dari latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, dan manfaat penelitian, metodologi penelitian dan sistematika penulisan

BAB 2 LANDASAN TEORI

Bab ini berisi tentang teori-teori singkat yang berkaitan dengan penelitian, antara lain pengertian dari analisis sentimen, crawling, data scraping, scraping social media, twitter, facebook dan tiktok, tenaga medis, *Naïve Bayes* dan *support vector machine*

BAB 3 ANALISIS DAN PERANCANGAN SISTEM

Bab ini berisi tentang analisis sentiment multi social media tentang pelayanan tenaga medis dengan menggunakan algiotitma *Naïve Bayes* dan SVM sebagai algoritma untuk mengklasifikasikan sentiment positif, negative dan netral dari setiap social media.

BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM

Bab ini berisi tentang implementasi atau penerapan dari rancangan sistem analisis sentiment multi social media tentang pelayanan tenaga medis dengan menggunakan algoritma *Naïve Bayes* dan SVM dan menguji sistem yang telah dibuat untuk menemukan kelemahan serta kelebihan dari sistem.

BAB 5 KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan berdasarkan hasil pengujian terhadap sistem analisis sentiment multi social media tentang pelayanan tenaga medis dengan menggunakan algoritma *Naïve Bayes* dan SVM dan saran yang dapat dijadikan rujukan untuk pengembangan sistem dan penelitian selanjutnya.

BAB 2

LANDASAN TEORI

2.1 Analisis Sentimen

Analisis sentimen adalah proses yang bertujuan untuk menentukan apakah isi dari dataset bernilai positif, negatif, atau netral. Analisis sentimen atau *opinion mining* adalah metode analisis yang berbasis komputasi (Liu,2010). Analisis sentimen ini diperlukan untuk melihat kecenderungan dari suatu opini, apakah bernilai positif, negatif atau netral. Salah satu contoh pemakaian analisis sentimen dalam kehidupan nyata adalah mengukur tingkat kepuasan konsumen terhadap suatu barang.

Dalam bidang *Natural Language Processing (NLP)*, *Mining Text*, maupun *DataMining*, penggunaan analisis sentimen berguna untuk bidang produk dan jasa yang manaberguna untuk hal evaluasi serta meningkatkan kualitas produk dan jasa melalui pendapatdan opini dari pengguna di berbagai *platform* (King, 2011).

Empati adalah kemampuan untuk bisa memahami perasaan dan permasalahan orang lain, berpikir dengan sudut pandang mereka, serta menghargai perbedaan perasaanorang lain tentang berbagai hal. (Goleman,2000) Kemampuan bagaimana kita menempatkan diri kita sendiri kedalam posisi orang lain serta bagaimana menghayati pengalaman agar dapat melihat situasi dari sudut pandang orang lain merupakan kemampuan yang sangat diperlukan dalam pelayanan publik. Salah satunya adalah pelayanan publik tenaga medis.

Sentimen analisis dikelompokkan menjadi 3 jenis yakni positif, negatif dan netral. Polaritas positif (1), mengartikan sentimen positif yang artinya mengarah ke persetujuan terhadap suatu hal. Polaritas negatif (-1), mengartikan sentimen negatif yang artinya mengarah ke penolakan terhadap suatu hal. Polaritas netral (0), mengartikan sentimen netral yang artinya tidak menunjukkan sifat dari sentimen positif maupun negatif.

2.2 *Scraping*

Teknik *scraping* merupakan proses pengambilan sebuah dokumen semiterstruktur dari internet, kemudian diekstrak untuk diambil data tertentu dari halaman tersebut agar bisa digunakan bagi kepentingan lain. Manfaatnya ialah agar informasi yang diambil atau digunakan lebih terfokus sehingga memudahkan dalam melakukan pencarian sesuatu (Arisandi, Indra, & Kartini, 2021).

Web scraping adalah teknik untuk mendapatkan informasi dari website secara otomatis tanpa harus menyalinnya secara manual. Tujuan dari web scraper adalah untuk mencari informasi tertentu dan kemudian mengumpulkannya dalam web yang baru. Web scraping berfokus dalam mendapatkan data dengan cara pengambilan dan ekstraksi. Manfaat dari web scraping ialah agar informasi yang dikeruk lebih terfokus sehingga memudahkan dalam melakukan pencarian sesuatu (Ayani, Pratiwi, & Muhandi, 2019).

2.3 *Data Scraping*

Data yang sudah dikumpulkan tadi akan diekstrak yang kemudian akan menjadi suatu target yang spesifik untuk dianalisa. Sebagai contoh pada penelitian ini akan dilakukannya scraping pada objek para tenaga kerja medis di halaman berbagai *platform* sosial media. Misalnya kita akan memilih atribut *figure*, pelayanan, *empathy*, umum yang akan diekstraksi dan data lain seperti nama dokter, spesialisasinya, kode dokter diabaikan. Atribut yang spesifik tersebut kemudian diekstraksi menjadi format CSV atau *spreadsheet* atau bisa juga kedalam format JSON.

Berikut adalah beberapa contoh hasil dari data *scraping* yang sudah dikutip di twitter.

Tabel 2. 1 contoh data scraping dari twitter

Teks	Polaritas
<p>MasyaAllah.. dokternya mana masih muda, gagah, ganteng, kooperatif lagi. Aku rela dah diperiksa lama.. 🤑</p> <p>@lautbieru (Feb, 14 2022)</p>	<p>Positif +</p>
<p>Bu dokterpunyabakatterpendam, lebih baik dipendam aja ya budokter 🤑</p> <p>@puji_hndk (Jun, 10 2020)</p>	<p>Netral</p>
<p>“gatau kenapa dari dulu gue anti banget sama dokter cewek kaya gini, jujurly pas kemarin usg takut banget di judge gue milih dokternya yg chinese ternyata bener enak banget di ajak ngobrol dan dokter gaul banget jadi santai banget org nya🤑”</p> <p>@maruk0chann (Feb, 4 2022)</p>	<p>Netral</p>
<p>Mau lu di rumah sakit yg elit sekalipun pasti deh susternya pada galak2 bgt. Galak yg jutek gitu tauga</p> <p>@anggitarahaha (Oct, 26 2021)</p>	<p>Negatif -</p>
<p>Pengalaman mamaku kali ya? Di RS dibentak2 suster, susternya lebih galak drpd dokternya pdhl dokternya aja chill abezz</p> <p>@alvinamkc (Oct, 17 2021)</p>	<p>Negatif -</p>

2.4 Scraping Sosial Media

2.4.1 Twitter

Untuk melakukan proses *scraping* pada media sosial Twitter, kita memerlukan bantuan *Twitter Developer* yang disebut *Tweepy*. *Tweepy* diperlukan untuk mengambil *Key* dan *Access Token* yang akan digunakan untuk mengambil suatu data dalam Twitter, jadi harus diperlukan izin terlebih dahulu oleh Twitternya sebelum mengambil data. *Twitter API* mempunyai sistem keamanan dimana menggunakan identitas yang disebut dengan *Customer Key* dan *Access Token*. Tujuan dari *auth key* tersebut adalah untuk menjaga *privacy* akun twitter *user*. Nilai dari *key* tersebut berupa pin atau nilai unik yang harus terjaga kerahasiaannya dan hanya pemilik akun twitter tersebut yang tahu.

2.4.2 Facebook

Saat ini facebook menjadi salah satu *platform* sosial media yang banyak digemari pengguna dan memiliki banyak data tentang berbagai individu. Melakukan *scraping* pada facebook artinya mengekstrak data tentang minat, perilaku, serta demografi individu lainnya yang kemudian akan disimpan dalam suatu dokumen.

Sebelum melakukan *scraping* pada *platform* sosial media Facebook, ada beberapa ketentuan yang perlu diperhatikan. Diantaranya harus mengecek [robot.txt](#) terlebih dahulu. *Robot.txt* ini digunakan oleh suatu situs *web* untuk memberi tahu “bot” bagaimana atau apakah suatu situs bisa dicrawled. Facebook memperingatkan di awal pada *file bot* mereka yang isinya “*Crawling Facebook is prohibited unless you have express written permission*”, yang artinya kita harus memerlukan izin terlebih dahulu dalam mengakses data di akun facebook seseorang. Tetapi tentunya anda masih bisa meng-*crawl* data facebook selama tidak memakai postingan ataupun data yang bersifat personal sesuai dengan ketentuan [GDPR](#). Ada beberapa tools yang dapat digunakan untuk mempermudah proses *scraping* facebook diantaranya *Zapier*, *Graph API*, *Facepager*, *Octoparse*, dll.

2.4.3 Tiktok

Saat ini tiktok menjadi *platform* sosial media yang sangat populer dan

menjadi salah satu aplikasi yang paling banyak diunduh. Aplikasi tiktok mengumpulkan dan menggunakan data dari akun *user*. Sehingga diperlukan VPN yang dapat dipercaya untuk melindungi akun pengguna saat pengguna tersebut menggunakan tiktok. *Developer* memerlukan API dalam *menscraping* data akun pengguna tiktok yang lainnya. Sama seperti *platform* media Facebook, ada beberapa *tools* yang dapat membantu developer dalam *menscraping* data tiktok diantaranya apify. Tiktok *scraper* akan mengekstraksi data menjadi file CSV atau JSON.

2.5 Tenaga Medis

Setiap orang berhak memperoleh kesehatan sebagai wujud dari HAM. Untuk itu, perlu dilakukan pembagunan kesehatan secara keseluruhan dan berkesinambungan, guna untuk meningkatkan kemauan, kesadaran dan kemampuan bagaimana hidup sehat dari setiap orang agar terwujudnya tingkat kesehatan masyarakat setinggi-tingginya. (Depkes RI, 2004)

Semakin berkembangnya zaman, terlihat bahwa tenaga medis terus berkembang, baik dalam kapasitas, jumlah maupun sarana prasarana diikuti dengan perkembangan teknologi. Walaupun rumah sakit terus berkembang, namun fungsi dasar dari rumah sakit tidak berubah. Fungsi dasar dari rumah sakit ialah pemulihan maupun perawatan kesehatan masyarakat, baik itu secara pelayanan rawat jalan, rawat inap serta konsultasi pemulihan atau pemeliharaan kesehatan anggota masyarakat (Kuncoro, 2000).

Fenomena atau peristiwa yang sering terjadi di rumah sakit, selalu berkaitan dengan pelayanan tenaga medis itu sendiri yang menyebabkan kesenjangan kualitas pelayanan yang ideal dengan tenaga medis. Hal ini dikarenakan adanya tuntutan pasien yang cukup tinggi, adanya ketidakmampuan tenaga medis, atau bisa jadi lemahnya pengetahuan dan ketrampilan yang dimiliki tenaga medis dalam menangani pasien. Mengingat dari tugas seorang tenaga medis sangat penting diantaranya melakukan diagnosa, perawatan, pengobatan maupun melaksanakan rujukan, maka perlu dilakukannya peningkatan kualitas dari tenaga medis agar tingkat kepuasan seorang pasien tinggi.

Agar terciptanya tingkat kepuasan pasien yang tinggi, maka perlu menciptakan HDP atau hubungan antara profesional (dokter) maupun tenaga medis lainnya dengan klien (pasien). Hubungan tersebut dilandasi dengan semua aspek praktik kedokteran yakni saat menentukan diagnosis maupun mengelola pasien. Bila seorang pasien telah menentukan dokter pilihannya, artinya pasien telah menyerahkan sepenuhnya dalam pengelolaan penyakitnya dan yakin bahwa dokter tersebut tidak akan melakukan tindakan diluar persetujuan pasien. Kepercayaan dari seorang pasien merupakan suatu amanah sehingga para pekerja tenaga medis harus memiliki ilmu dan kemampuan yang terbaik serta memiliki kode etik kedokteran, moral dan hukum yang berlaku. Dalam melakukan komunikasi dengan pasien, seorang tenaga medis harus mampu menunjukkan respect kepada pasien, rasa percaya diri dan empati. Untuk bisa berempati dengan pasien, diharapkan para pekerja tenaga medis harus bisa menjadi pendengar yang baik dan dapat memberi respons yang baik terhadap apa yang diceritakan pasien.

2.6 *Naïve Bayes*

Naïve Bayes Classifier adalah suatu cara pengklasifikasian yang bersumber dari *teorema Bayes*, yakni menggunakan *probability* dan *statistic* yang dikemukakan oleh ilmuan Thomas Bayes. *Teorema Bayes* yakni memprediksi peluang yang ada di masa depan berdasarkan pengalaman yang sudah ada di masa lalu. Algoritma *Naïve Bayes* diidentik dengan pendekatan peluang ataupun probabilitas.

Menurut Kusriani (Kusriani, 2009), *Naïve Bayes Classification* memiliki tingkat akurasi dan kecepatan yang tinggi ketika diaplikasikan kedalam dataset yang besar. Keuntungan dari penggunaan *Naïve Bayes* adalah hanya memerlukan jumlah data training yang sedikit untuk pengklasifikasiannya.

Algoritma dari *Naïve Bayes* sendiri memiliki persamaan yang dibuat sebagai acuan dalam menghitung nilai probabilitas ketika mengambil suatu keputusan.

Persamaannya sebagai berikut :

$$P(X|H) = \frac{P(H|X), P(X)}{P(H)}$$

X : Data dengan *class* yang belum diketahui

H : Hipotesis data X yang merupakan suatu *class* spesifik

P(H|X) : Probabilitas hipotesis H berdasarkan kondisi si X

P(H) : Probabilitas hipotesis H

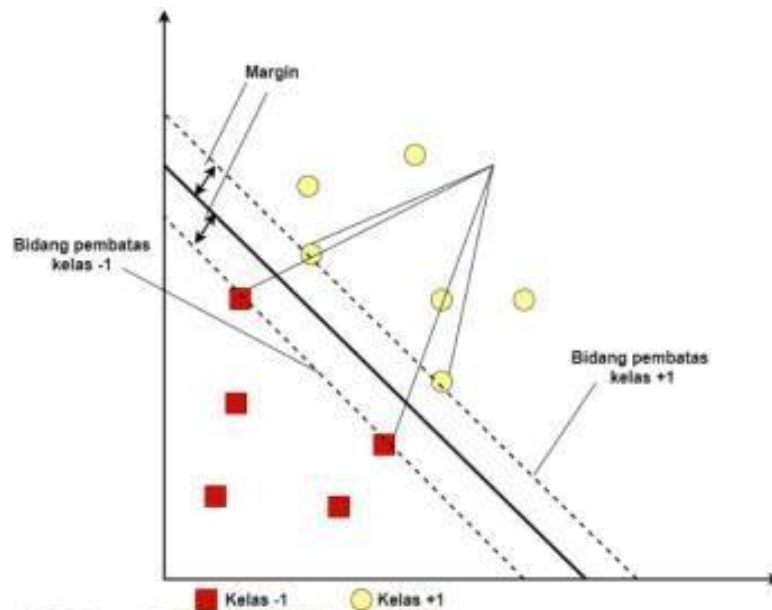
P(X|H) : Probabilitas X berdasarkan kondisi padahipotesis H

P(X) : Probabilitas X

2.7 *Support Vector Machine*

Support Vector Machine atau yang disebut SVM adalah salah satu metode supervised learning yang artinya ada label di tiap datanya. Sehingga dalam proses penentuan keputusan, mesin akan mengklasifikasikan data testing kedalam label yang sesuai dengan atribut yang dimilikinya.

Mulanya SVM hanya dapat menyelesaikan klasifikasi linear, kemudian berkembang sehingga dapat menyelesaikan klasifikasi non-linear dengan memasukkan konsep kernel kedalam ruang berdimensi tinggi. Tujuannya adalah mencari *hyperplane* atau pemisah yang bisa memaksimalkan jarak atau margin antar kelas data. Yang dimaksud dengan margin disini adalah jarak terdekat antara *pattern* dari setiap kelas dengan *hpyerplane*. Proses pencarian *hyperplane* yang terbaik adalah dengan metode SVM.



Gambar 2. 1 Support Vector Machine

2.8 Penelitian yang Relevan

Berikut ini beberapa penelitian yang berkaitan dengan penelitian peneliti antara lain :

1. Berdasarkan penelitian yang berjudul *Analisis Sentimen Pengguna Youtube di Indonesia pada Review Smartphone menggunakan Naïve Bayes*, datasetnya adalah komentar di aplikasi Youtube mengenai *review smartphone*. Penilaiannya dibagi menjadi 3 klasifikasi diantaranya positif, netral, dan negatif dengan menggunakan tahap *preprocessing*. (Balya, 2019)
2. Berdasarkan penelitian yang berjudul *Analisis Sentimen Kepuasan Pengguna Aplikasi Whatsapp Menggunakan Algoritma Naïve Bayes dan Support Vector Machine*, *dataset* yang diambil adalah berupa komentar pengguna aplikasi Whatsapp. Berdasarkan hasil analisis kepuasan pengguna aplikasi WhatsApp pada *play store* dengan algoritma NB dan SVM melalui proses *preprocessing* dan *modeling* serta proses evaluasi pengkategorian ulasan berlabel positif dan ulasan berlabel negatif dapat diketahui bahwa hasil akhir nilai akurasi NB adalah sebesar 70,40% dan SVM sebesar 77,00% dan nilai AUC pada algoritma NB sebesar 0,585 dan SVM adalah 0,876. SVM memiliki tingkat akurasi yang unggul 6,6%, daripada NB. Maka metode SVM dapat digunakan

pada karakteristik data yang sama. (Saepulrohman, Saepudin, & Gustian, 2021)

3. Berdasarkan penelitian yang berjudul *Analisis Sentimen Pembelajaran Daring Pada Media Sosial Twitter Menggunakan Multinomial Naïve Bayes dan Support Vector Machine*, dataset yang diambil adalah tweet berbahasa Indonesia dengan domain “belajar daring”, “kuliah daring”, dan “sekolah daring” sebanyak 9908 tweet. Data dilabeli dengan label positif dan negatif. Pada pengujian menggunakan *Multinomial Naïve Bayes* didapat tingkat akurasi 69,616% data training, dan 71,790% data testing. Sedangkan pada *Support Vector Machine*, didapat tingkat akurasi 73,372% data training serta 74,426% data testing. (Gregorius, 2021)
4. Berdasarkan penelitian yang berjudul *Perbandingan Metode Naïve Bayes dan Support Vector Machine Pada Analisis Sentimen Twitter* menggunakan dataset dari twitter, dengan hasil kedua metode dapat melakukan performa yang baik dan metode *Naïve Bayes* memiliki hasil akurasi, presisi, *recall*, dan *F1-Score* yang lebih baik dibandingkan dengan metode SVM. Dalam penelitian ini menggunakan metode *Naïve Bayes* dan SVM untuk proses klasifikasi. Selain kedua metode tersebut, dapat juga dicoba berbagai metode klasifikasi yang lain seperti *K-Nearest Neighbor*, *Decision Tree*, *rule-based*, dll untuk analisis sentimen. (Fikri, Sabrilla, & Azhar, 2020)
5. Berdasarkan penelitian yang berjudul *Komparasi Algoritma SVM dan Naïve Bayes dengan Algoritma Genetika pada Analisis Sentimen Calon Gubernur Jabar 2018- 2023* dengan menggunakan data *tweet* pemilihan gubernur jawa barat 2018. Hasil dari penelitian ini Hasil pengujian data *tweet* mengenai calon gubernur jawa barat periode 2018-2023 dengan Algoritma *Support Vector Machine* menghasilkan rata-rata akurasi 92,61% dengan AUC 0,950, Algoritma *Naïve Bayes* menghasilkan rata-rata akurasi 93,29% dengan AUC 0,525, Algoritma *Support Vector Machine* berbasis *Genetic Algorithm* menghasilkan rata-rata akurasi 93,03% dengan AUC 0,869 dan Algoritma *Naïve Bayes* berbasis *Genetic Algorithm* menghasilkan rata-rata akurasi 92,85% dengan AUC 0,543. Dengan demikian model algoritma *Support*

Vector Machine berbasis *Genetic Algorithm* adalah model algoritma terbaik dalam penelitian ini dan dapat memberikan hasil terbaik dalam pengujian dan pengklasifikasian analisis sentiment *tweet* calon gubernur jawa barat periode 2018-2023 dibandingkan dengan model algoritma *Naïve Bayes* berbasis *Genetic Algorithm* (NB-GA). (Gunawan & Riana, 2020)

6. Berdasarkan penelitian yang berjudul *Analisis Sentimen Zoom Cloud Meetings di Playstore Menggunakan Naïve Bayes dan SVM* dengan dataset komentar tentang *zoom* dari *playstore*. Hasil dari penelitian ini dari hasil evaluasi diketahui bahwa nilai akurasi untuk mengklasifikasikan label positif dan label negatif pada ulasan para pengguna aplikasi *zoom cloud meetings* di *Google Play Store*, dapat dibuktikan dengan nilai akurasi dan nilai AUC dari masing-masing algoritma yaitu untuk NB nilai akurasi = 74,37% dan nilai AUC = 0,659. Sedangkan untuk algoritma SVM nilai akurasi = 81,22% dan nilai AUC = 0,886. Dalam penelitian ini dapat diketahui bahwa tingkat akurasi yang didapatkan algoritma *Support Vector Machine* (SVM) lebih unggul 6,85% dibandingkan algoritma *Naïve Bayes* (NB) dengan 1.007 *record dataset*. (Herlinawati & Yuliani, 2020)
7. Berdasarkan penelitian yang berjudul *Scalable sentiment classification for Big Data analysis using Naïve Bayes Classifier* pengklasifikasian NBC pada *Big Data* memiliki tingkat akurasi 82% pada saat menganalisis sentimen jutaan ulasan film. (Bingwei Liu,dkk 2013).
8. Berdasarkan penelitian yang berjudul *A Feature Based Approach for Sentiment Analysis by Using Support Vector Machine* peneliti mengembangkan proses keseluruhan 'Analisis Sentimen berbasis Aspek atau Fitur' dengan memanfaatkan pengklasifikasi yang disebut *Support Vector Machine* (SVM) dalam pendekatan barudan terbukti menjadi salah satu cara paling efektif untuk menganalisis dan mengekstrak pandangan pengguna secara keseluruhan tentang fitur tertentu dan jugakeseluruhan produk *e-commerce*. (D.V. Nagarjuna Devi,dkk 2016).

BAB 3

ANALISIS DAN PERANCANGAN SISTEM

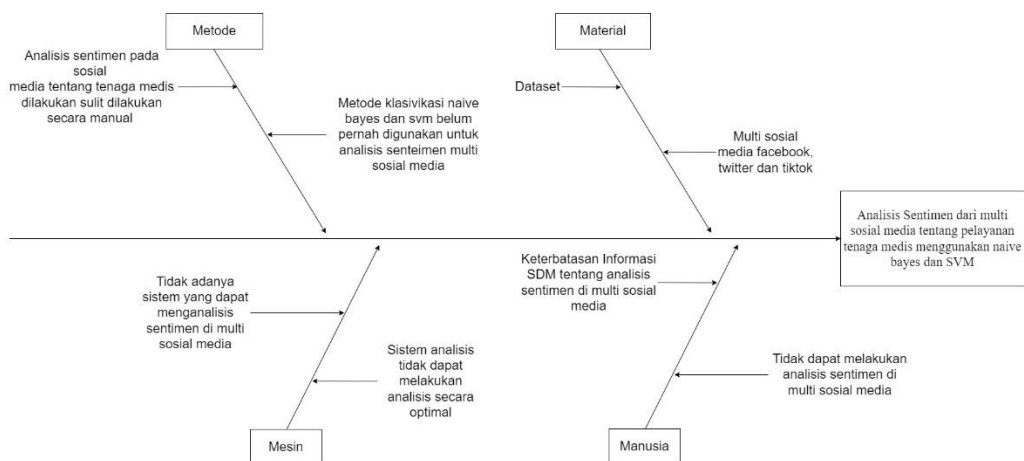
3.1 Analisis Sistem

Pada bab ini akan membahas tentang analisis dan perancangan sistem analisis sentimen dengan menggunakan algoritma *Naïve Bayes* dan *support vector machine*. Analisis sistem adalah metode untuk menemukan kelemahan sistem guna memperoleh gambaran terhadap sistem yang akan dibuat agar dapat sesuai dengan yang dibutuhkan pengguna.

3.1.1 Analisis Masalah

Dalam penelitian ini, permasalahan yang akan dibahas adalah bagaimana membangun sebuah sistem yang dapat melakukan analisis sentimen dari multi social media tentang pelayanan tenaga medis menggunakan metode klasifikasi *Naïve Bayes* dan SVM secara otomatis, cepat dan tepat.

Setelah melakukan analisis masalah, selanjutnya penulis mengidentifikasi dan menjelaskan penyebab masalah yang terjadi untuk kemudian dicari solusinya. Kemudian proses dari analisis dan identifikasi masalah yang telah dilakukan dapat dijelaskan melalui *Ishikawa Diagram (fishbone)* yang dapat dilihat pada gambar 3.1 dibawah.



Gambar 3. 1 Ishikawa Diagram

Pada gambar diatas merupakan *Ishikawa Diagram* yang dapat menjelaskan penyebab atau landasan masalah dalam penelitian ini sehingga dibangun sebuah

sistem yang akan menjadi solusi dari masalah tersebut. Dari permasalahan yang ada pada sistem penelitian ini dibagi menjadi 4 komponen yaitu metode, material, mesin dan manusia. Pada metode dijelaskan mengenai bagaimana sistem dirancang dan dikembangkan berdasarkan metode yang belum pernah digunakan pada penelitian yang sudah pernah digunakan. Sedangkan pada material dijelaskan mengenai bagian utama dan penting yang diperlukan untuk menjalankan sistem pada penelitian ini. Mesin menjelaskan mengenai hal apa saja kemungkinan yang akan mempengaruhi kinerja dari perangkat keras (*hardware*) dalam menjalankan sistem pada penelitian ini. Manusia menjelaskan mengenai berbagai hal yang kemungkinan akan diterima oleh *user* dalam menjalankan sistem pada penelitian ini.

3.1.2 Analisis Kebutuhan

Analisis kebutuhan dalam penelitian ini dibagi dua, yaitu analisis kebutuhan fungsional dan analisis kebutuhan nonfungsional. Dalam analisis kebutuhan fungsional dijelaskan mengenai hal yang harus ada pada sistem tersebut dapat mencapai tujuan utamanya, sedangkan dalam analisis kebutuhan nonfungsional dijelaskan mengenai hal yang menjadi kebutuhan tambahan sehingga mendukung kinerja sistem agar lebih optimal.

a. Analisis kebutuhan fungsional

Kebutuhan fungsional pada sistem dalam penelitian ini terdiri dari

1. Sistem dapat menghasilkan *output* berupa hasil analisis sentiment multi social media tentang pelayanan tenaga medis menggunakan metode *Naïve Bayes* dan SVM
2. Analisis sentimen kalimat yang terdapat pada dataset akan diklasifikasikan menggunakan algoritma *Naïve Bayes* dan SVM

b. Analisis kebutuhan non fungsional

Kebutuhan nonfungsional pada sistem dalam penelitian ini terdiri dari:

1. Performa Sistem

Sistem dapat melakukan proses klasifikasi kalimat yang terdapat pada komentar menggunakan algoritma *Naïve Bayes* dan SVM sehingga diperoleh hasil analisis sentiment yang bernilai

2. Pengendalian / *control*

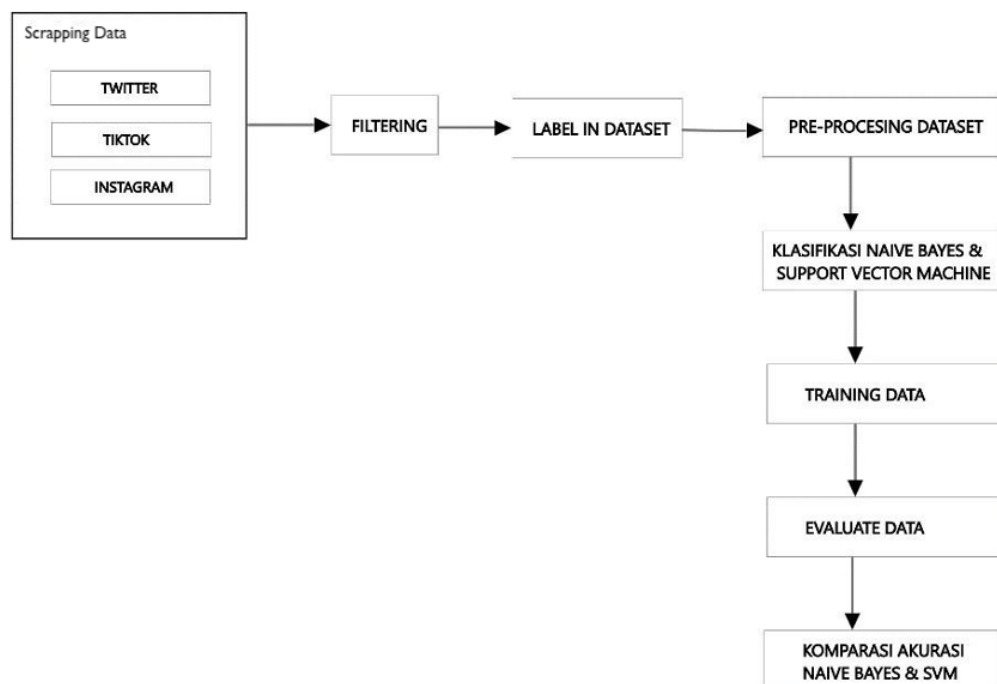
Sistem akan menampilkan pesan *error* apabila pengguna melakukan kesalahan pada saat menggunakan sistem

3. Manajemen Kualitas

Sistem memiliki kualitas yang baik yaitu dengan menghasilkan analisis sentiment secara cepat dan tepat.

3.2 General Arsitektur Sistem

Gambar 2.2 berikut merupakan arsitektur sentimen analisis yang dilakukan pada penelitian ini.



Gambar 2.2 Arsitektur umum sentimen analisis

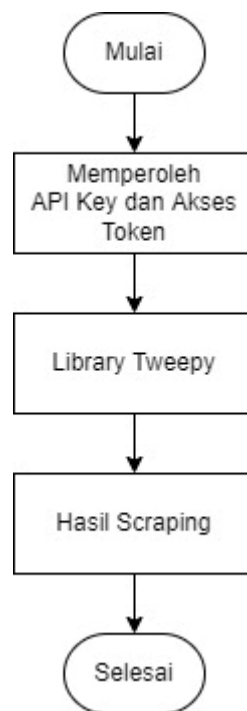
Berdasarkan Gambar 2.2 maka dapat dijelaskan sebagai berikut:

3.2.1 Scraping Data

Langkah pertama dalam analisis sentimen dimulai dengan melakukan *scraping* data pada *platform* sosial media (Twitter,Tiktok,Facebook) dengan menggunakan hashtag #tenagamedis yang merupakan domain atau topik penelitian, dan hal-hal yang berkaitan dengan objek yang diteliti yang kemudian akan dijadikan sebagai dataset.

1. Scraping Twitter

Adapun tahapan dalam melakukan *scraping* terhadap *media sosial twitter* dapat dilihat pada gambar berikut:



Gambar 3.2 Flowchart Melakukan Scraping Twitter

Berdasarkan Gambar 3.2 maka dapat dijelaskan sebagai berikut:

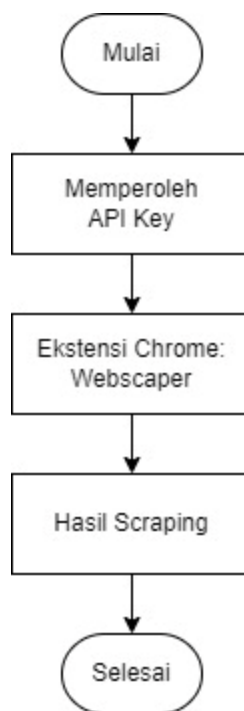
- a. Tahapan pertama yang harus dilakukan adalah memperoleh API key serta akses token twitter.
- b. Tahapan berikutnya adalah melakukan *scraping* menggunakan bahasa pemrograman python dengan melakukan import *library tweepy*,

setelah itu melakukan pengaturan *tweepy* dengan cara mengisi data dengan API key serta Akses token yang diperoleh.

- c. Setelah hasil scraping tampil kemudian melakukan penyimpanan atau ekstraksi dari *tweets* yang sudah didapatkan.

2. *Scraping Facebook dan Tiktok*

Adapun tahapan dalam melakukan *scraping* terhadap *media sosial Facebook dan Tiktok* dapat dilihat pada gambar berikut:



Gambar 3.3 Flowchart Melakukan Scraping Facebook dan Tiktok

Berdasarkan Gambar 3. 3 maka dapat dijelaskan alur dalam melakukan scraping facebook dan TikTok sebagai berikut:

- a. Tahapan pertama yang harus dilakukan adalah memperoleh API key pada sosmed yang akan dilakukan *scraping*.
- b. Tahapan berikutnya adalah melakukan *scraping* menggunakan ekstensi *browser google chrome* dengan melakukan penambahan ekstensi bernama *webscraper*, setelah itu melakukan pengaturan untuk melakukan proses *scraping*.

- c. Setelah itu hasil *scraping* akan tampil kemudian melakukan penyimpanan data kedalam perangkat.

3.2.2 Filtering

Dataset yang telah terkumpul akan dipisah menjadi per kalimat dan dilakukannya filterisasi. *Filtering* adalah proses untuk mengambil hanya data opini saja, sedangkan data/twit/status yang bukan merupakan opini atau *review* yang akan dikeluarkan dari dataset.

3.2.3 Label In Dataset

Dataset yang telah melalui tahapan *filtering* akan dilakukan pelabelan apakah *tweet* tersebut tergolong netral, positif, ataupun negatif. Dalam proses ini, pelabelan dilakukan secara manual.

3.2.4 Preprocessing

Dataset yang sudah terpisah akan dimasukkan ke dalam *text preprocessing* atau sering disebut dengan praproses teks. Praproses teks terdiri atas *case folding*, *tokenizing*, *stopwords*, dan *stemming*. *Case folding* merupakan praproses untuk mengubah seluruh teks menjadi huruf kecil. *Tokenizing* adalah proses membagi teks yang berupa kalimat, paragraf, ataupun dokumen menjadi bagian yang terkecil. *Stopwords* adalah *common words* yang sering muncul dalam suatu teks yang dianggap tidak memiliki makna. *Stemming* adalah proses pembentukan kata dasar dengan cara menghilangkan imbuhan pada kata.

3.2.5 Klasifikasi Naïve Bayes dan SVM

Pada tahapan ini, hasil *preprocessing* akan dilakukan klasifikasi menggunakan bahasa pemrograman python, dimana pada tahap ini akan dilakukan perhitungan menggunakan Naïve Bayes dan SVM agar dataset dapat diklasifikasikan menjadi positif, netral, dan negatif.

3.2.6 Training Data

Dalam proses klasifikasi terdapat data training yang bertujuan sebagai data train dalam mengklasifikasi menggunakan Naïve Bayes dan SVM. Dalam melakukan penelitian ini, peneliti menggunakan data training sebesar 75% dan data test sebesar 25%.

3.2.7 Evaluate Data

Tahapan berikutnya adalah mengevaluasi data menggunakan *confusion matrix*. Sehingga diketahui hasil akurasi dari algoritma Naïve Bayes dan SVM.

3.2.8 Komparasi Akurasi Naïve Bayes dan SVM

Tahap terakhir adalah melakukan perbandingan tingkat akurasi terhadap algoritma *Naïve Bayes* dan SVM. Sehingga diketahui algoritma mana yang memiliki nilai akurasi terbaik.

3.3 Perancangan Sistem Analisis Sentimen

3.3.1 Algoritma Naïve Bayes

Dasar dari *Naïve Bayes* yang dipakai dalam pemrograman adalah rumus Bayes seperti pada persamaan (Zaenal Abidin, 2021).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Keterangan:

$P(A | B)$: Probabilitas A terjadi dengan bukti bahwa B telah terjadi (probabilitas superior)

$P(B | A)$: Probabilitas B terjadi dengan bukti bahwa A telah terjadi

$P(A)$: Peluang terjadinya A

$P(B)$: Peluang terjadinya B

Rumus Probabilitas Class:

$$P(c) = \frac{Nc}{N}$$

Dimana:

$P(c)$ = Prior dari kelas c

Nc = Banyaknya dokumen pada data latih yang menjadi kategori c

N = Banyaknya dokumen pada data latih

Rumus Multinomial Model:

$$P(w|c) = \frac{\text{count}(w, c) + 1}{\text{count}(c) + |V|}$$

Dimana:

$\text{count}(w,c)$ = Jumlah kemunculan kata w pada kelas c

$\text{count}(c)$ = Jumlah total kemunculan kata pada kelas c

$|V|$ = Jumlah fitur

Rumus Posterior Naïve Bayes:

$$\text{Posterior} = \frac{\text{Likelihood} * \text{Prior}}{\text{Evidence}}$$

Untuk menghitung posterior dibutuhkan nilai prior dari perhitungan conditional probability:

$$\text{Posterior} = \text{Likelihood} \times \text{Prior}$$

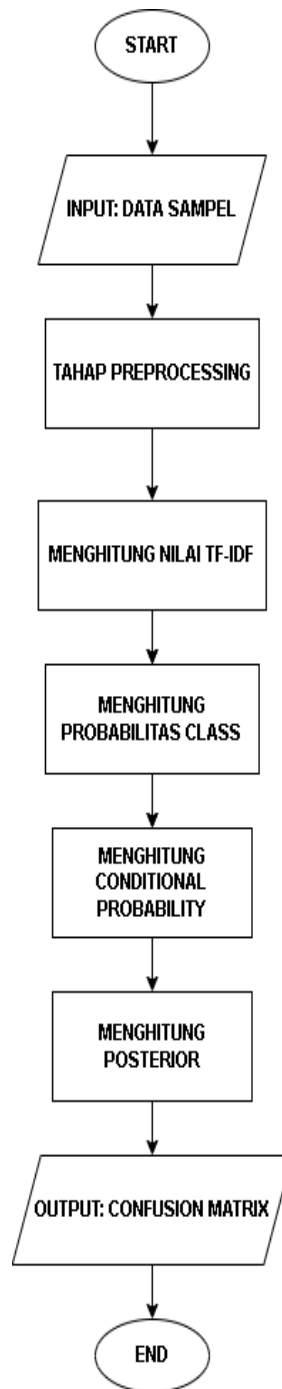
$$P(c|d) = \prod_{w \in d} P(w|c) * P(c)$$

Dimana:

$P(w|c)$ = Nilai conditional probability

$P(c)$ = Nilai probailitas label/class

Dibawah ini merupakan flowchart dari algoritma Naïve Bayes



Gambar 3. 4 Flowchart Algoritma Naïve Bayes

Dibawah ini merupakan perhitungan dari algoritma Naïve Bayes

1) Input Data Sampel

Pada penelitian ini pelabelan disimbolkan menggunakan angka yaitu 0 untuk netral, -1 untuk negatif dan 1 untuk positif.

Tabel 3. 1 Input Data Sampel

No	Stemming	Label
1	mudah pahami mirip tenaga kesehatan dokter	0
2	dokter tenaga medis kasih tahu waktu pagi siang sore masyarakat	-1
3	donasi orang sakit selalu dilakukan secara rutin kepada tenaga medis	-1
4	tenaga medis lebih bijak percaya ilmu sains	1
5	masalah faktor kesehatan pendidikan tenaga medis	1
6	dinas kelautan perikanan peternakan cianjur kurang tenaga medis antisipasi wabah penyakit mulut kuku	1
7	pemerintah siaga dokter anak tenaga medis daerah fasilitas kesehatan	1
8	gejala mirip diagnosa tenaga medis psikiater	-1
9	pemerintah koordinasi dokter anak tenaga medis fasilitas kesehatan	0
10	semangat tinggi tenaga medis	1

2) Tahap Preprocessing

Tabel 3. 2 Tahap Preprocessing

No	Comment	Case Folding	Tokenizing	Normalisasi	Stopwords	Stemming	Label
1	@Muhdzahir @hrdbacot Biar mudah di pahami, mirip seperti tenaga kesehatan, nah PNS itu dokter, PPPK itu perawat, asn itu tenaga medis/kesehatan	biar mudah di pahami mirip seperti tenaga kesehatan nah pns itu dokter pppk itu perawat asn itu tenaga medis kesehatan	biar mudah di pahami mirip seperti tenaga kesehatan nah pns itu dokter pppk itu perawat asn itu tenaga medis kesehatan	biar mudah di pahami mirip seperti tenaga kesehatan nah pns itu dokter pppk itu perawat asn itu tenaga medis kesehatan	mudah pahami mirip tenaga kesehatan dokter	mudah pahami mirip tenaga kesehatan dokter	0
2	"Gimana ya, para dokter atau tenaga medis lain tuh gak ngasih tau	gimana ya para dokter atau tenaga medis lain tuh gak ngasih tau	gimana ya para dokter atau tenaga medis lain tuh gak	gimana ya para dokter atau tenaga medis lain tuh gak kasih	dokter tenaga medis kasih tahu waktu pagi	dokter tenaga medis kasih tahu	-1

	waktu ² yg pas cuma blg kalo 3x1 ya diminum pagi, siang, sore	waktu yg pas cuma blg kalo ya diminum pagi siang sore masyarakat ngikutin apa yg tertulis masalahnya juga masyarakat kita jarang ada yg nanya rentan waktu kalo udh tertera ya udh gitu	ngasih tau waktu yg pas cuma blg kalo ya diminum pagi siang sore masyarakat ngikutin apa yg	tahu waktu yg pas cuma blg kalo ya diminum pagi siang sore masyarakat ikut apa yg tertulis masalahnya juga masyarakat kita jarang ada yg nanya rentan waktu kalo udh tertera ya udh gitu	siang sore masyarakat	waktu pagi siang sore masyarakat at	
	Masyarakat ngikutin apa yg tertulis		tertulis				
	Masalahnya juga masyarakat kita jarang ada yg nanya rentan waktu, kalo udh tertera 3x1 ya udh gitu 😊 https://t.co/4qZyHAJI3B		masalahnya juga masyarakat kita jarang ada yg nanya rentan waktu kalo udh tertera ya udh gitu				
3	@kiaraaaaaa aas bau-bau donasi org sakit yang selalu	bau bau donasi org sakit yang selalu dilakukan secara rutin	bau bau donasi org sakit yang selalu dilakukan	bau bau donasi orang sakit yang selalu dilakukan	donasi orang sakit selalu dilakukan secara rutin	donasi orang sakit selalu dilakuka	-1

	dilakukan secara rutin kepada tenaga medis ...	kepada tenaga medis	secara rutin kepada tenaga medis	secara rutin kepada tenaga medis	kepada tenaga medis	n secara rutin kepada tenaga medis	
4	Sebagai tenaga medis, beliau terasa lebih bijak aja. Kita percaya ilmu sains dan ngehargai sesuatu yang "tidak kelihatan" karena nyatanya emang ada sesuatu yang kadang gak bisa dijelaskan dengan medis juga. Menjalani satu profesi tanpa mendepak fakta di luar kuasanya itu bagus	sebagai tenaga medis beliau terasa lebih bijak aja kita percaya ilmu sains dan ngehargai sesuatu yang tidak kelihatan karena nyatanya emang ada sesuatu yang kadang gak bisa dijelaskan dengan medis juga menjalani satu profesi tanpa mendepak fakta di luar kuasanya itu bagus	sebagai tenaga medis beliau terasa lebih bijak aja kita percaya ilmu sains dan ngehargai sesuatu yang tidak kelihatan karena nyatanya emang ada sesuatu yang kadang gak bisa jelaskan dengan medis juga menjalani satu profesi tanpa mendepak fakta di luar kuasanya itu bagus	sebagai tenaga medis beliau terasa lebih bijak aja kita percaya ilmu sains dan hargai sesuatu yang tidak kelihatan karena nyata emang ada suatu yang kadang gak bisa jelaskan dengan medis juga jalani satu profesi tanpa mendepak fakta di luar kuasanya itu bagus	tenaga medis lebih bijak percaya ilmu sains	tenaga medis lebih bijak percaya ilmu sains	1
			luar kuasanya itu bagus				

5	<u>Selanjutnya,</u> <u>masalah faktor</u> <u>kesehatan dan</u> <u>pendidikan itu</u> <u>belum</u> <u>terjangkau,</u> <u>guru atau</u> <u>tenaga</u> <u>pengajar</u> <u>maupun tim</u> <u>medis atau</u> <u>tenaga medis</u> <u>itu belum bisa</u> <u>ke daerah</u> <u>tertentu karena</u> <u>pembangunan</u> <u>jalannya</u> <u>belum ada.</u> https://t.co/1AVzXZ6pgu	selanjutnya masalah faktor kesehatan dan pendidikan itu belum terjangkau guru atau tenaga pengajar maupun tim medis atau tenaga medis itu belum bisa ke daerah tertentu karena pembangunan jalannya belum ada	selanjutnya masalah faktor kesehatan dan pendidikan itu belum terjangkau guru atau tenaga pengajar maupun tim medis atau tenaga medis itu belum bisa ke daerah tertentu karena pembanguna n jalannya belum ada	selanjutnya masalah faktor kesehatan dan pendidikan itu belum terjangkau guru atau tenaga pengajar maupun tim medis atau tenaga medis itu belum bisa ke daerah tertentu karena pembanguna n jalannya belum ada	masalah faktor kesehatan dan pendidikan tenaga medis	masalah faktor kesehatan pendidikan tenaga medis	1
6	<u>Dinas</u> <u>Kelautan</u> <u>Perikanan dan</u> <u>Peternakan</u> <u>(Dislutkanak)</u> <u>Cianjur</u> <u>kekurangan</u> <u>tenaga medis</u> <u>dalam</u> <u>mengantisipasi</u> <u>penyebaran</u> <u>wabah</u> <u>penyakit mulut</u> <u>dan kuku</u> <u>(PMK).</u>	dinas kelautan perikanan dan peternakan dislutkanak cianjur kekurangan tenaga medis dalam mengantisipasi penyebaran wabah penyakit mulut dan kuku	dinas kelautan perikanan dan peternakan dislutkanak cianjur kekurangan tenaga medis dalam mengantisipasi asi penyebaran wabah	dinas kelautan perikanan dan peternakan dislutkanak cianjur kurang tenaga medis dalam antisipasi penyebaran wabah penyakit mulut kuku	dinas kelautan perikanan dan peternakan cianjur kurang tenaga medis antisipasi wabah penyakit mulut kuku	dinas kelautan perikanan dan peternakan cianjur kurang tenaga medis antisipasi wabah penyakit mulut kuku	1

	<u>#Dislutkanakc</u> <u>ianjur</u> <u>https://t.co/HP</u> <u>hznMc3vQ</u>		penyakit mulut dan kuku	mulut dan kuku			
7	Puan mengingatkan pemerintah agar menyiagakan dokter anak dan tenaga medis lainnya di setiap daerah, serta mempersiapka n fasilitas kesehatan untuk skenario terburuk.	puan mengingatkan pemerintah agar menyiagakan dokter anak dan tenaga medis lainnya di setiap daerah serta mempersiapka n fasilitas kesehatan untuk skenario terburuk	puan mengingatka n pemerintah agar menyiagaka n dokter anak dan tenaga medis lainnya di setiap daerah serta mempersiap kan fasilitas kesehatan untuk skenario terburuk	pemerintah siaga dokter anak tenaga medis lainnya di setiap daerah serta mempersiap kan fasilitas kesehatan untuk skenario terburuk	pemerintah siaga dokter anak tenaga medis daerah fasilitas kesehatan	pemerintah siaga dokter anak tenaga medis daerah fasilitas kesehatan	1
8	Hanya karena gejalanya mirip, belum tentu diagnosanya sama. Tenaga medis juga bukan, tapi gampang banget nyuruh orang lain ke psikiater. Ga sopan.	hanya karena gejalanya mirip belum tentu diagnosanya sama tenaga medis juga bukan tapi gampang banget nyuruh orang lain ke psikiater ga sopan	hanya karena gejalanya mirip belum tentu diagnosanya sama tenaga medis juga bukan tapi gampang banget nyuruh	hanya karena gejala mirip belum tentu diagnosa sama tenaga medis juga bukan tapi gampang banget nyuruh orang lain ke psikiater ga sopan	gejala mirip diagnosa tenaga medis psikiater	gejala mirip diagnosa tenaga medis psikiater	-1

			orang lain ke psikiater ga sopan				
9	“Saya mengingatkan pemerintah agar terus berkoordinasi dengan WHO. Pemerintah melalui Kemenko PMK juga perlu mengoordinasi kan dokter anak & tenaga medis lain di setiap daerah, & persiapkan fasilitas kesehatan untuk skenario terburuk.” – Puan Maharani Puan PM	saya mengingatkan pemerintah agar terus berkoordinasi dengan who pemerintah melalui kemenko juga perlu mengoordinasi kan dokter anak amp tenaga medis lain di setiap daerah amp persiapkan fasilitas kesehatan untuk skenario terburuk puan maharani puan	saya mengingatka n pemerintah agar terus berkoordinas i dengan who pemerintah melalui kemenko juga perlu mengoordina sikan dokter anak amp tenaga medis lain di setiap daerah amp persiapan fasilitas kesehatan untuk skenario terburuk puan maharani puan	saya mengingatka n pemerintah agar terus koordinasi dengan who pemerintah melalui kemenko juga perlu koordinasi dokter anak amp tenaga medis lain di setiap daerah amp persiapkan fasilitas kesehatan untuk skenario terburuk puan maharani puan	pemerintah koordinasi dokter anak tenaga medis fasilitas kesehatan	pemerintah koordinasi dokter anak tenaga medis fasilitas kesehatan	0
10	Semangat juanganya tinggi sekali sampai harus dapatkan tenaga medis :(semangat juanganya tinggi sekali sampai harus dapatkan tenaga medis	semangat juanganya tinggi sekali sampai harus dapatkan	semangat juanganya tinggi sekali sampai harus dapatkan tenaga medis	semangat tinggi tenaga medis	semangat tinggi tenaga medis	1

	https://t.co/FTofUmtc8N		tenaga medis					
--	---	--	--------------	--	--	--	--	--

3) Menghitung Nilai TF-IDF

Tabel 3. 3 Data untuk Perhitungan Nilai TF-IDF

No	Stemming	Label
1	mudah pahami mirip tenaga kesehatan dokter	0
2	dokter tenaga medis kasih tahu waktu pagi siang sore masyarakat	-1
3	donasi orang sakit selalu dilakukan secara rutin kepada tenaga medis	-1
4	tenaga medis lebih bijak percaya ilmu sains	1
5	masalah faktor kesehatan pendidikan tenaga medis	1
6	dinas kelautan perikanan peternakan cianjur kurang tenaga medis antisipasi wabah penyakit mulut kuku	1
7	pemerintah siaga dokter anak tenaga medis daerah fasilitas kesehatan	1
8	gejala mirip diagnosa tenaga medis psikiater	-1
9	pemerintah koordinasi dokter anak tenaga medis fasilitas kesehatan	0
10	semangat tinggi tenaga medis	1

Pembagian Data:

80% Data Training

20% Data Testing

Diketahui:

Tabel 3. 4: Data Term

Dokumen	Term
1	tenaga
2	dokter
3	medis
4	bijak
5	kesehatan
6	penyakit
7	pemerintah
8	gejala
9	fasilitas
10	semangat

Nilai Term yang dihasilkan:

Tabel 3. 5 Data Nilai Term

TF		IDF									
No	Term	D1(1)	D2(1)	D3(1)	D4(1)	D5(1)	D6(1)	D7(1)	D8(1)	D9(0)	D10(0)
1	tenaga	1	1	1	1	1	1	1	1	1	1
2	dokter	1	1	0	0	0	0	1	0	1	0
3	medis	0	1	1	1	1	1	1	1	1	1
4	bijak	0	0	0	1	0	0	0	0	0	0
5	kesehatan	1	0	0	0	1	0	1	0	1	0
6	penyakit	0	0	0	0	0	1	0	0	0	0
7	pemerintah	0	0	0	0	0	0	1	0	1	0
8	gejala	0	0	0	0	0	0	0	1	0	0
9	fasilitas	0	0	0	0	0	0	1	0	1	0
10	semangat	0	0	0	0	0	0	0	0	0	1

Data Training:

Tabel 3. 6 TF Data Training

TF		IDF							
No	Term	D1(1)	D2(1)	D3(1)	D4(1)	D5(1)	D6(1)	D7(1)	D8(1)
1	tenaga	1	1	1	1	1	1	1	1
2	dokter	1	1	0	0	0	0	1	0
3	medis	0	1	1	1	1	1	1	1
4	bijak	0	0	0	1	0	0	0	0
5	kesehatan	1	0	0	0	1	0	1	0
6	penyakit	0	0	0	0	0	1	0	0
7	pemerintah	0	0	0	0	0	0	1	0
8	gejala	0	0	0	0	0	0	0	1
9	fasilitas	0	0	0	0	0	0	1	0
10	semangat	0	0	0	0	0	0	0	0

Dalam data nilai term terdapat D1-D10 dimana data nilai term tersebut merupakan keseluruhan data (100%). *Machine learning* membutuhkan sebuah *data learning*. *Data training* atau *data learning* ini adalah *dataset* yang digunakan untuk pembelajaran atau mencari model yang sesuai untuk sistem analisis sentiment dengan menggunakan metode *naïve bayes*. Data nilai term yang digunakan adalah data dari D1-D8 atau 80% data untuk proses training.

Data Testing:

Tabel 3. 7 TF Data Testing

TF	IDF		
No	Term	D9(0)	D10(0)
1	tenaga	1	1
2	dokter	1	0
3	medis	1	1
4	bijak	0	0
5	kesehatan	1	0
6	penyakit	0	0
7	pemerintah	1	0
8	gejala	0	0
9	fasilitas	1	0
10	semangat	0	1

Dalam data nilai term terdapat D1-D10 dimana data nilai term tersebut merupakan keseluruhan data (100%). Untuk mengevaluasi *machine learning* membutuhkan sebuah *data testing*. *Data testing* ini adalah dataset yang digunakan untuk mengetahui performa algoritma yang sudah dilatih pada sistem analisis sentiment dengan menggunakan metode *naïve bayes*. Data nilai term yang digunakan adalah data dari D9-D10 atau 20% data untuk proses training.

4) Menghitung Probabilitas Class

Rumus Probabilitas Class:

$$P(c) = \frac{Nc}{N}$$

Diketahui jumlah dokumen yang digunakan untuk data training yaitu 8 data.

Misalnya menghitung Probabilitas Class 1 (Positif):

$$P(Positif) = \frac{4}{8}$$

$$P(Positif) = 0,5$$

Misalnya menghitung Probabilitas Class 0 (Netral):

$$P(Negatif) = \frac{1}{8}$$

$$P(Negatif) = 0,125$$

Misalnya menghitung Probabilitas Class -1 (Negatif):

$$P(Negatif) = \frac{3}{8}$$

$$P(Negatif) = 0,375$$

5) Menghitung Conditional Probability

Conditional Probability menggunakan model Multinomial.

Rumus Multinomial Model:

$$P(w|c) = \frac{\text{count}(w, c) + 1}{\text{count}(c) + |V|}$$

Misalnya Menghitung Multinomial Model Class Positif untuk Fitur Pertama pada Term "tenaga":

$$P(w|1) = \frac{0 + 1}{4 + 10}$$

$$P(w|1) = \frac{1}{14}$$

$$P(w|1) = 0,071428571$$

Misalnya Menghitung Multinomial Model Class Netral untuk Fitur Pertama pada Term "tenaga":

$$P(w|0) = \frac{10 + 1}{1 + 10}$$

$$P(w|0) = \frac{11}{11}$$

$$P(w|0) = 1$$

Misalnya Menghitung Multinomial Model Class Negatif untuk Fitur Pertama pada Term "tenaga":

$$P(w|-1) = \frac{0 + 1}{3 + 10}$$

$$P(w|-1) = \frac{1}{13}$$

$$P(w|-1) = 0,076923077$$

Hasil Perhitungan Conditional Probabilitas:

Tabel 3. 8 Hasil Perhitungan Conditional Probabilitas

TF		IDF								Conditional Probabilitas		
No	Term	D1(1)	D2(1)	D3(1)	D4(1)	D5(1)	D6(1)	D7(1)	D8(1)	P(W 1)	P(W 0)	P(W -1)
1	tenaga	1	1	1	1	1	1	1	1	0,0714285 71	1	0,0769230 77
2	dokter	1	1	0	0	0	0	1	0	0,0714285 71	0,0909090 91	0,3076923 08
3	medis	0	1	1	1	1	1	1	1	0,0714285 71	0,0909090 91	0,6153846 15
4	bijak	0	0	0	1	0	0	0	0	0,1428571 43	0,0909090 91	0,0769230 77
5	kesehatan	1	0	0	0	1	0	1	0	0,2857142 86	0,0909090 91	0,0769230 77
6	penyakit	0	0	0	0	0	1	0	0	0,1428571 43	0,0909090 91	0,0769230 77
7	pemerintah	0	0	0	0	0	0	1	0	0,1428571 43	0,0909090 91	0,0769230 77
8	gejala	0	0	0	0	0	0	0	1	0,0714285 71	0,0909090 91	0,1538461 54
9	fasilitas	0	0	0	0	0	0	1	0	0,0714285 71	0,1818181 82	0,0769230 77
10	semangat	0	0	0	0	0	0	0	0	0,0714285 71	0,0909090 91	0,0769230 77

6) Menghitung Posterior

Perhitungan posterior merupakan proses testing yang digunakan untuk menentukan kelas berdasarkan probabilitas pada setiap kelas. Kelas dengan probabilitas tertinggi merupakan kelas baru untuk data yang sedang diujikan.

Rumus Posterior Naïve Bayes:

$$Posterior = \frac{Likelihood * Prior}{Evidence}$$

Karena pada perhitungan posterior nilai evidence selalu sama untuk setiap kelas maka nilai evidence tidak akan memberikan pengaruh yang signifikan pada perhitungan posterior sehingga untuk mencari nilai posterior cukup dengan menghitung likelihood dengan priornya saja.

Untuk menghitung posterior dibutuhkan nilai prior dan hasil perhitungan conditional probability sebagai likelihoodnya. Berikut adalah perhitungan posterior menggunakan data testing. Berikut rumus yang digunakan:

$$Posterior = Likelihood \times Prior$$

$$P(c|d) = \prod_{w \in d} P(w|c) * P(c)$$

Misalnya menghitung probabilitas dokumen data testing (Dokumen 9) untuk Class 1 (Positif):

$$P(c|d_9) = (P(indonesia|c) * P(meal|c) * P(bungkus|c) * P(konser|c) * P(jangan|c)) \times P(c)$$

$$P(c|d_9) = (0,071428571 * 0,071428571 * 0,071428571 * 0,142857143 * 0,285714286) * 0,071428571$$

$$P(c|d_9) = 1,06248$$

Misalnya menghitung probabilitas dokumen data testing (Dokumen 9) untuk Class 0 (Netral):

$$P(c|d_9) = (P(indonesia|c) * P(meal|c) * P(bungkus|c) * P(konser|c) * P(jangan|c)) \times P(c)$$

$$P(c|d_9) = (1 * 0,090909091 * 0,090909091 * 0,090909091 * 0,090909091) * 1$$

$$P(c|d_9) = 6,83013$$

Misalnya menghitung probabilitas dokumen data testing (Dokumen 9) untuk Class -1 (Negatif):

$$P(c|d_9) = (P(indonesia|c) * P(meal|c) * P(bungkus|c) * P(konser|c) * P(jangan|c)) \times P(c)$$

$$P(c|d_9) = (0,076923077 * 0,307692308 * 0,615384615 * 0,076923077 * 0,076923077) * 0,076923077$$

$$P(c|d_9) = 6,62964$$

Hasil Perhitungan Posterior Data Testing:

Tabel 3. 9 Hasil Perhitungan Posterior Data Testing

No	Doc Ke-1	P(1 Di)	P(0 Di)	P(-1 Di)	Max	Classification
1	9	1,06248E-06	6,83013E-05	6,62964E-06	0	Class -1
2	10	5,31241E-07	1,24184E-05	4,14352E-07	1	Class 0

7) Output Confusion Matrix

Berikut Hasil Confusion Matriks Keseluruhan Data Scraping:Twitter

Tabel 3. 10 Model Confusion Matrix

	Confusion Matrix			
Actual Label	Positif	cell 1	cell 2	cell 3
	Netral	cell 4	cell 5	cell 6
	Negatif	cell 7	cell 8	cell 9
		Positif	Netral	Negatif
	Prediction Label			

Tabel 3. 11 Perhitungan Confusion Matrix

	Confusion Matrix			
Actual Label	Positif	32	5	6
	Netral	15	42	10
	Negatif	10	6	29
		Positif	Netral	Negatif
	Prediction Label			

3.3.2 Algoritma Support Vector Machine

Kebutuhan rumus dalam klasifikasi *Support Vector Machine*

Rumus IDF:

$$idf_t = \log_{10} \frac{N}{df_i}$$

Keterangan:

N = Jumlah Data

dfi = Jumlah Nilai DF

Rumus Bobot:

$$Wtd = TFdt * IDFt$$

Keterangan:

TFdt = Nilai Term Frequent Dokumen ke t

IDFt = Nilai Invers Document Frequent ke t

Rumus Normalisasi TF-IDF:

$$w_{t,d} = \frac{w_{t,d}}{\sqrt{\sum_{t=1}^n w_{t,d}^2}}$$

Rumus Matrix Hessian:

$$D_{ij} = (y_i * y_j) * (K(x_i, x_j) + \lambda^2)$$

Rumus Kernel Polynomial of Degree up to d:

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

Keterangan:

X_i = sebagai data matriks transpose

c = sebagai class

Menghitung Matrix Hessian dari Hasil Perhitungan Kernel Polynomial:

Nilai parameter yang digunakan pada manualisasi algoritme SVM:

c = 1,

d = 2,

cLR = 0,1,

Lambda = 5,

Complexity = 1

Rumus Matrix Hessian:

$$D_{ij} = (y_i * y_j) * (K(x_i, x_j) + \lambda^2)$$

Keterangan:

Y_i = sebagai data matriks transpose

K = sebagai nilai kernel

Rumus Nilai Error:

$$E_i = \sum_{j=1}^n \alpha_j D_{ij}$$

Keterangan:

α_j = sebagai nilai alfa awal

d_{ij} = sebagai nilai jumlah kernel

Rumus Nilai Gamma:

$$\gamma = \frac{cLR}{\max(\text{Matriks Hessian})}$$

cLR = sebagai Nilai Koefisien Learning Rate

Rumus Perubahan Nilai Alfa:

$$\delta\alpha_i = \min\{\max[\gamma(1 - E_i), \alpha_i], C - \alpha_i\}$$

Keterangan:

γ = sebagai nilai gamma

E_i = sebagai nilai error

α_i = sebagai nilai alfa

C = sebagai class

Rumus Nilai Bias:

$$bias_i = -\frac{1}{2} \cdot \left(\sum_{j=1}^n \alpha_j y_j K(x_i, x^-) + \sum_{j=1}^n \alpha_j y_j K(x_i, x^+) \right)$$

Keterangan:

$K(x_i, x^-)$ = sebagai nilai kernel negatif

$K(x_i, x^+)$ = sebagai nilai kernel positif

α = sebagai nilai alfa baru

Rumus Prediksi Sentimen:

$$X_{pred} = (\text{Nilai Kernel} * \text{Nilai Update Alfa} * \text{Class})$$

Rumus Klasifikasi *Support Vector Machine*:

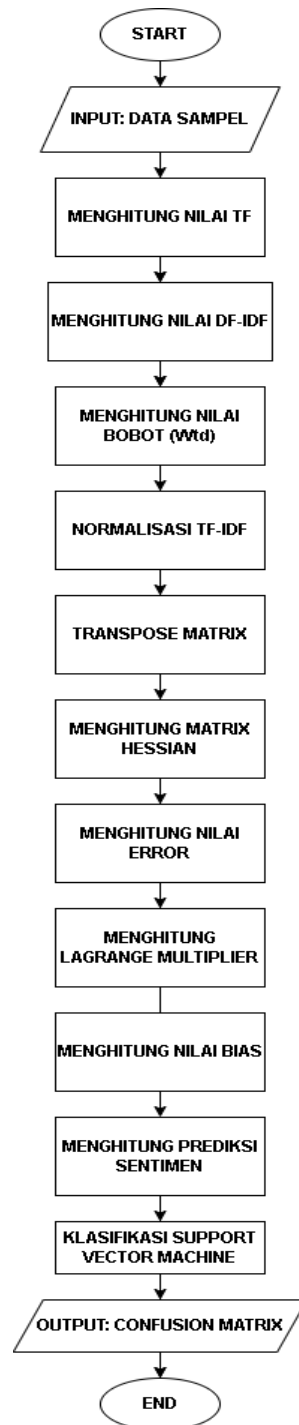
$$f(x_{prediksi}) = \text{Sign} \left(\sum_{i=1}^m \alpha_i y_i K(x_i, x_{pred}) + b_i \right)$$

Keterangan:

K = sebagai nilai kernel

b_i = sebagai nilai bias

Dibawah ini merupakan *flowchart* dari algoritma *Support Vector Machine*



Gambar 3. 5 Flowchart Algoritma Support Vector Machine

Berdasarkan Gambar 3.5 *Flowchart* pada algoritma SVM, terdapat beberapa tahapan untuk mendapatkan hasil klasifikasi, tahap pertama input data yang akan

digunakan, pada perhitungan ini data yang digunakan yaitu data sampel terkait sentimen mengenai petugas medis. Detail mengenai masing-masing tahapan akan dijelaskan sebagai berikut.

1) Input Data Sampel

Tabel 3. 12 Input Data Sampel

No	Stemming	Label
1	mudah pahami mirip tenaga kesehatan dokter	0
2	dokter tenaga medis kasih tahu waktu pagi siang sore masyarakat	-1
3	donasi orang sakit selalu dilakukan secara rutin kepada tenaga medis	-1
4	tenaga medis lebih bijak percaya ilmu sains	1
5	masalah faktor kesehatan pendidikan tenaga medis	1
6	dinas kelautan perikanan peternakan cianjur kurang tenaga medis antisipasi wabah penyakit mulut kuku	1
7	pemerintah siaga dokter anak tenaga medis daerah fasilitas kesehatan	1
8	gejala mirip diagnosa tenaga medis psikiater	-1
9	pemerintah koordinasi dokter anak tenaga medis fasilitas kesehatan	0
10	semangat tinggi tenaga medis	1

2) Menghitung Nilai TF

Pembagian Data:

80% Data Training

20% Data Testing

Tabel 3. 13 Data Term

Dokumen	Term
1	tenaga
2	dokter
3	medis
4	bijak
5	kesehatan
6	penyakit
7	pemerintah
8	gejala
9	fasilitas
10	semangat

Tabel 3. 14 Nilai Term

TF		IDF									
No	Term	D1(1)	D2(1)	D3(1)	D4(1)	D5(1)	D6(1)	D7(1)	D8(1)	D9(0)	D10(0)
1	tenaga	1	1	1	1	1	1	1	1	1	1
2	dokter	1	1	0	0	0	0	1	0	1	0
3	medis	0	1	1	1	1	1	1	1	1	1
4	bijak	0	0	0	1	0	0	0	0	0	0
5	kesehat an	1	0	0	0	1	0	1	0	1	0
6	penyaki t	0	0	0	0	0	1	0	0	0	0
7	pemerin tah	0	0	0	0	0	0	1	0	1	0
8	gejala	0	0	0	0	0	0	0	1	0	0
9	fasilitas	0	0	0	0	0	0	1	0	1	0
10	semang at	0	0	0	0	0	0	0	0	0	1

3) Menghitung Nilai DF-IDF

DF-IDF Data Training:

Rumus IDF:

$$idf_t = \log_{10} \frac{N}{df_i}$$

Menghitung Nilai IDF untuk Term "Indonesia" pada Data Training:

$$idf_t = \log_{10} \frac{N}{df_i}$$

$$idf_t = \log_{10} \frac{8}{8}$$

$$idf_t = \log_{10}(1)$$

$$idf_t = 0$$

IDF Data Training:

Tabel 3. 15 IDF Data Training

TF		IDF								DF	N/df	IDF
No	Term	D1(1)	D2(1)	D3(1)	D4(1)	D5(1)	D6(1)	D7(1)	D8(1)			
1	tenaga	1	1	1	1	1	1	1	1	8	1	0

TF		IDF								DF	N/df	IDF
No	Term	D1(1)	D2(1)	D3(1)	D4(1)	D5(1)	D6(1)	D7(1)	D8(1)			
2	dokter	1	1	0	0	0	0	1	0	3	2,6666 67	0,425968 732
3	medis	0	1	1	1	1	1	1	1	7	1,1428 57	0,057991 947
4	bijak	0	0	0	1	0	0	0	0	1	8	0,903089 987
5	kesehata n	1	0	0	0	1	0	1	0	3	2,6666 67	0,425968 732
6	penyakit	0	0	0	0	0	1	0	0	1	8	0,903089 987
7	pemerint ah	0	0	0	0	0	0	1	0	1	8	0,903089 987
8	gejala	0	0	0	0	0	0	0	1	1	8	0,903089 987
9	fasilitas	0	0	0	0	0	0	1	0	1	8	0,903089 987
10	semang at	0	0	0	0	0	0	0	0	0	0	0

Dalam data nilai term terdapat D1-D10 dimana data nilai term tersebut merupakan keseluruhan data (100%). *Machine learning* membutuhkan sebuah *data learning*. *Data learning* atau *data training* ini adalah dataset yang digunakan untuk pembelajaran atau mencari model yang sesuai untuk sistem analisis sentiment dengan menggunakan metode *Support Vector Machine*. Data nilai term yang digunakan adalah data dari D1-D8 atau 80% data untuk proses training. DF atau Document Frequent merupakan banyaknya dokumen yang muncul berdasarkan frekuensi.

IDF Data Testing:

Tabel 3. 16 IDF Data Testing

TF	IDF			DF	N/df	IDF
No	Term	D9(0)	D10(0)			
1	tenaga	1	1	2	1	0
2	dokter	1	0	1	2	0,301029996
3	medis	1	1	2	1	0
4	bijak	0	0	0	0	0
5	kesehatan	1	0	1	2	0,301029996
6	penyakit	0	0	0	0	0
7	pemerintah	1	0	1	2	0,301029996
8	gejala	0	0	0	0	0
9	fasilitas	1	0	1	2	0,301029996
10	semangat	0	1	1	2	0,301029996

Dalam data nilai term terdapat D1-D10 dimana data nilai term tersebut merupakan keseluruhan data (100%). Untuk mengevaluasi *machine learning* membutuhkan sebuah *data testing*. *Data testing* ini adalah dataset yang digunakan untuk mengetahui performa algoritma yang sudah dilatih pada sistem analisis sentiment dengan menggunakan metode *Support Vector Machine*. Data nilai term yang digunakan adalah data dari D9-D10 atau 20% data untuk proses training.

4) Menghitung Nnilai Bobot (Wtd)

Rumus Bobot:

$$Wtd = TFdt * IDFt$$

Menghitung Nilai Bobot (Wtd) Term "Jaminan" pada Data Training:

$$Wtd = TFdt * IDFt$$

$$Wtd = 1 * 0$$

$$Wtd = 0$$

Wtd Data Training:

Tabel 3. 17 Wtd Data Training

TF		IDF							
	Term	D1(1)	D2(1)	D3(1)	D4(1)	D5(1)	D6(1)	D7(1)	D8(1)
1	tenaga	0	0	0	0	0	0	0	0
2	dokter	0,4259 68732	0,4259 68732	0	0	0	0	0,4259 68732	0
3	medis	0	0,0579 91947	0,0579 91947	0,0579 91947	0,0579 91947	0,0579 91947	0,0579 91947	0,0579 91947
4	bijak	0	0	0	0,9030 89987	0	0	0	0
5	keseha tan	0,4259 68732	0	0	0	0,4259 68732	0	0,4259 68732	0
6	penya kit	0	0	0	0	0	0,9030 89987	0	0
7	pemer intah	0	0	0	0	0	0	0,9030 89987	0
8	gejala	0	0	0	0	0	0	0	0,9030 89987
9	fasilita s	0	0	0	0	0	0	0,9030 89987	0
10	seman gat	0	0	0	0	0	0	0	0

Wtd Data Testing:

Tabel 3. 18 Data Testing

TF	IDF		
No	Term	D9(0)	D10(0)
1	tenaga	0	0
2	dokter	0,301029996	0
3	medis	0	0
4	bijak	0	0
5	kesehatan	0,301029996	0
6	penyakit	0	0
7	pemerintah	0,301029996	0
8	gejala	0	0
9	fasilitas	0,301029996	0
10	semangat	0	0,301029996

5) Normalisasi TF-IDF

Rumus:

$$w_{t,d} = \frac{w_{t,d}}{\sqrt{\sum_{t=1}^n w_{t,d}^2}}$$

Menghitung Nilai Normalisasi (Wtd) Term "indonesia" pada Data Training:

$$w_{t,d} = \frac{0}{\sqrt{0^2}}$$

$$w_{t,d} = 0$$

Normalisasi Data Training:

Tabel 3. 19 Normalisasi Data Training

TF		IDF							
No	Term	D1(1)	D2(1)	D3(1)	D4(1)	D5(1)	D6(1)	D7(1)	D8(1)
1	tenaga	0	0	0	0	0	0	0	0
2	dokter	1	1	0	0	0	0	1	0
3	medis	0	1	1	1	1	1	1	1
4	bijak	0	0	0	1	0	0	0	0
5	kesehatan	1	0	0	0	1	0	1	0
6	penyakit	0	0	0	0	0	1	0	0
7	pemerintah	0	0	0	0	0	0	1	0
8	gejala	0	0	0	0	0	0	0	1

9	fasilitas	0	0	0	0	0	0	1	0
10	semangat	0	0	0	0	0	0	0	0

Normalisasi Data Testing:

Tabel 3. 20 Normalisasi Data Testing

TF	IDF		
No	Term	D9(0)	D10(0)
1	tenaga	0	0
2	dokter	1	0
3	medis	0	0
4	bijak	0	0
5	kesehatan	1	0
6	penyakit	0	0
7	pemerintah	1	0
8	gejala	0	0
9	fasilitas	1	0
10	semangat	0	1

6) Transpose Matrix

Transpose Matrix Data Training:

Tabel 3. 21 Transpose Matrix Data Training

Dokumen	Fitur 1	Fitur 2	Fitur 3	Fitur 4	Fitur 5	Fitur 6	Fitur 7	Fitur 8	Fitur 9	Fitur 10	Label
1	0	1	0	0	1	0	0	0	0	0	1
2	0	1	1	0	0	0	0	0	0	0	-1
3	0	0	1	0	0	0	0	0	0	0	1
4	0	0	1	1	0	0	0	0	0	0	1
5	0	0	1	0	1	0	0	0	0	0	1
6	0	0	1	0	0	1	0	0	0	0	1
7	0	1	1	0	1	0	1	0	1	0	1
8	0	0	1	0	0	0	0	1	0	0	1

Transpose Matrix Data Testing:

Tabel 3. 22 Transpose Matrix Data Testing

Dokumen	Fitur 1	Fitur 2	Fitur 3	Fitur 4	Fitur 5	Fitur 6	Fitur 7	Fitur 8	Fitur 9	Fitur 10	Label
9	0	1	0	0	1	0	1	0	1	0	1
10	0	0	0	0	0	0	0	0	0	1	1

7) Menghitung Matrix Hessian

Menghitung Matrix Hessian menggunakan fungsi kernel Polynomial of Degree up to d.

Rumus Matrix Hessian:

$$D_{ij} = (y_i * y_j) * (K(x_i, x_j) + \lambda^2)$$

Rumus Kernel Polynomial of Degree up to d:

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

Misalnya Menghitung Kernel Dokumen 1 pada Fitur 1 Data Training:

$$K(x_i, x_j) = ((0 * 0) + (1 * 1) + (0 * 0) + (0 * 0) + (1 * 1) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + 1)^2$$

$$K(x_i, x_j) = 3$$

Hasil Perhitungan Kernel Data Training:

Tabel 3. 23 Hasil Perhitungan Kernel Data Training

	1	2	3	4	5	6	7	8	9	10	Jumlah
1	3	3	2	2	2	1	1	1	1	1	17
2	3	3	2	1	1	1	1	1	1	1	15
3	2	2	2	1	1	1	1	1	1	1	13
4	3	3	3	2	1	1	1	1	1	1	17
5	3	3	3	2	2	1	1	1	1	1	18
6	3	3	3	2	2	2	1	1	1	1	19
7	6	6	5	4	4	3	3	2	2	1	36
8	3	3	3	2	2	2	2	2	1	1	21

Hasil Perhitungan Kernel Data Testing:

Tabel 3. 24 Hasil Perhitungan Kernel Data Testing

	1	2	3	4	5	6	7	8	9	10	Jumlah
9	5	5	4	4	4	3	3	2	2	1	33
10	2	2	2	2	2	2	2	2	2	2	20

Menghitung Matrix Hessian dari Hasil Perhitungan Kernel Polynomial:

Nilai parameter yang digunakan pada manualisasi algoritme SVM:

$$c = 1,$$

$d = 2$,
 $cLR = 0,1$,
 $\text{Lambda} = 5$,
 $\text{Complexity} = 1$

Misalnya Menghitung Matrix Hessian Dokumen 1 pada Fitur 1 Data Training:

$$D_{ij} = (1 * 1) * (3 + 5^2)$$

$$D_{ij} = 28$$

Hasil Perhitungan Matrix Hessian Data Training:

Tabel 3. 25 Hasil Perhitungan Matrix Hessian Data Training

	1	2	3	4	5	6	7	8	9	10
1	28	28	27	27	27	26	26	26	26	26
2	28	28	27	26	26	26	26	26	26	26
3	27	27	27	26	26	26	26	26	26	26
4	28	28	28	27	26	26	26	26	26	26
5	28	28	28	27	27	26	26	26	26	26
6	28	28	28	27	27	27	26	26	26	26
7	31	31	30	29	29	28	28	27	27	26
8	28	28	28	27	27	27	27	27	26	26

Hasil Perhitungan Matrix Hessian Data Testing:

Tabel 3. 26 Hasil Perhitungan Matrix Hessian Data Testing:

	1	2	3	4	5	6	7	8	9	10
9	30	30	29	29	29	28	28	27	27	26
10	27	27	27	27	27	27	27	27	27	27

8) Menghitung Nilai Error

Diketahui Nilai Alfa:

Tabel 3. 27 Diketahui Nilai Alfa

	Alfa
1	0
2	0
3	0
4	0

5	0
6	0
7	0
8	0
9	0
10	0

Rumus Nilai Error:

$$E_i = \sum_{j=1}^n \alpha_j D_{ij}$$

Misalnya Menghitung Nilai Error Data ke-1 pada Data Training:

$$E_i = (0 * 28) + (0 * 28) + (0 * 27) + (0 * 27) + (0 * 27) + (0 * 26) + (0 * 26) + (0 * 26) + (0 * 26) + (0 * 26)$$

$$E_i = 0$$

Hasil Perhitungan Nilai Error:

Tabel 3. 28 Hasil Perhitungan Nilai Error

Dokumen ke-	Hasil Nilai Alfa
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0

9) Menghitung Lagrange Multiplier

Perhitungan nilai alfa membutuhkan nilai alfa awal, cLR, gamma, dan complexity. Langkah pertama pada perhitungan nilai alfa adalah menentukan nilai gamma menggunakan nilai cLR dan matriks hessian. Berikut adalah perhitungan untuk menentukan nilai gamma.

a) Menghitung Nilai Gamma

Rumus:

$$\gamma = \frac{cLR}{\max(\text{Matriks Hessian})}$$

Misalnya Menghitung Nilai Gamma:

$$\gamma = \frac{0,1}{31}$$

$$\gamma = 0,003226$$

Hasil Perhitungan Nilai Gamma:

Tabel 3. 29 Hasil Perhitungan Nilai Gamma

Dokumen ke-	Hasil Nilai Gamma
1	0,003226
2	0,003226
3	0,003226
4	0,003226
5	0,003226
6	0,003226
7	0,003226
8	0,003226
9	0,003226
10	0,003226

b) Menghitung Perubahan Nilai Alfa

Rumus:

$$\delta\alpha_i = \min\{\max[\gamma(1 - E_i), \alpha_i], C - \alpha_i\}$$

Misalnya Menghitung Perubahan Nilai Alfa:

$$\delta\alpha_i = ((0 * (1 - 0,003226) * 1 - 0,003226))$$

$$\delta\alpha_i = -0,003226$$

Hasil Perubahan Nilai Alfa:

Tabel 3. 30 Hasil Perhitungan Perubahan Nilai Alfa

Dokumen ke-	Hasil Nilai Alfa
1	-0,003226
2	-0,003226
3	-0,003226
4	-0,003226
5	-0,003226
6	-0,003226
7	-0,003226
8	-0,003226
9	-0,003226
10	-0,003226

10) Menghitung Nilai Bias

Rumus:

$$bias_i = -\frac{1}{2} \cdot \left(\sum_{j=1}^n \alpha_j y_j K(x_i, x^-) + \sum_{j=1}^n \alpha_j y_j K(x_i, x^+) \right)$$

Misalnya Menghitung Nilai Bias Data ke-1 pada Data Training:

$$bias_i = -\frac{1}{2} \cdot ((-0,003226 * 3 * 1) + (-0,003226 * 3 * 1))$$

$$bias_i = 0,053225806$$

Hasil Perhitungan Nilai Bias:

Tabel 3. 31 Hasil Perhitungan Nilai Bias

Bias ke-1	Hasil Nilai Bias
1	0,053225806
2	0,077419355
3	0,074193548
4	0,025806452
5	0,024193548
6	0,022580645
7	-0,00483871
8	0,087096774
9	0,053225806
10	0,053225806

11) Menghitung Prediksi Sentimen

Rumus:

$$X_{pred} = (\text{Nilai Kernel} * \text{Nilai Update Alfa} * \text{Class})$$

Prediksi Data Sentimen pada Dokumen ke-1:

$$X_{pred} = (17 * (-0,003226) * 1)$$

$$X_{pred} = 0$$

Hasil Perhitungan Prediksi Data Sentimen:

Tabel 3. 32 Hasil Perhitungan Pprediksi Data Sentimen

Prediksi	Hasil Prediksi
xpred1	0
xpred2	0,048387097
xpred3	0,041935484
xpred4	-0,05483871
xpred5	-0,058064516
xpred6	-0,061290323
xpred7	-0,116129032
xpred8	0,067741935
xpred9	0
xpred10	0

12) Klasifikasi *Support Vector Machine*

Rumus:

$$f(x_{prediksi}) = \text{Sign} \left(\sum_{i=1}^m \alpha_i y_i K(x_i, x_{pred}) + b_i \right)$$

Misalnya Melakukan Klasifikasi Data Testing Data ke-9:

$$\begin{aligned}
f(x_{Prediksi}) &= \text{Sign} \left((0,053225806 + 0) \right. \\
&\quad + (0,077419355 + 0,048387097) \\
&\quad + (0,074193548 + 0,041935484) \\
&\quad + (0,025806452 + (-0,05483871)) \\
&\quad + (0,024193548 + (-0,058064516)) \\
&\quad + (0,022580645 + (-0,061290323)) \\
&\quad + (-0,00483871 + (-0,116129032)) \\
&\quad + (0,087096774 + 0,067741935) \\
&\quad \left. + (0,053225806 + 0) + (0,053225806 + 0) \right) \\
f(x_{Prediksi}) &= 1
\end{aligned}$$

Hasil Klasifikasi Data Testing:

Tabel 3. 33 Hasil Klasifikasi Data Testing

Dokumen ke-	Hasil Klasifikasi
9	1
10	1

13) Output Confusion Matrix

Berikut Hasil *Confusion Matriks* Keseluruhan Data *Scraping* Twitter:

Model *Confusion Matrix*

Tabel 3. 34 Model Confusion Matrix

	Confusion Matrix			
Actual Label	Positif	cell 1	cell 2	cell 3
	Netral	cell 4	cell 5	cell 6
	Negatif	cell 7	cell 8	cell 9
		Positif	Netral	Negatif
	Prediction Label			

Hasil Confusion Matrix:

Tabel 3. 35 Hasil Confusion Matrix

	<i>Confusion Matrix</i>			
Actual Label	Positif	31	8	4
	Netral	6	50	11
	Negatif	1	11	33
		Positif	Netral	Negatif
	Prediction Label			

Berikut Perhitungan Confusion Matriks Setiap Class:

a) Confusion Matriks Class Positif

$$TP = cell1 = 31$$

$$TN = cell5 + cell6 + cell8 + cell9 = 50 + 11 + 11 + 33 = 105$$

$$FN = cell2 + cell3 = 8 + 4 = 12$$

$$FP = cell4 + cell7 = 6 + 1 = 7$$

Menghitung Nilai Precision Class Positif:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{31}{31 + 7} * 100\%$$

$$Precision = \frac{31}{38} * 100\%$$

$$Precision = 0,82 * 100\%$$

$$Precision = 82\%$$

Menghitung Nilai Recall Class Positif:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{31}{31 + 12} * 100\%$$

$$Recall = \frac{31}{43} * 100\%$$

$$Recall = 0,72 * 100\%$$

$$Recall = 72\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,72 * 0,82)}{(0,72 + 0,82)} * 100\%$$

$$F1\ Score = \frac{1,1808}{1,54} * 100\%$$

$$F1\ Score = 0,77 * 100\%$$

$$F1\ Score = 77\%$$

b) Confusion Matriks Class Netral

$$TP = cell5 = 50$$

$$TN = cell1 + cell3 + cell7 + cell9 = 31 + 4 + 1 + 33 = 69$$

$$FN = cell4 + cell6 = 6 + 11 = 17$$

$$FP = cell2 + cell8 = 8 + 11 = 19$$

Menghitung Nilai Precision Class Netral:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{50}{50 + 19} * 100\%$$

$$Precision = \frac{50}{69} * 100\%$$

$$Precision = 0,72 * 100\%$$

$$Precision = 72\%$$

Menghitung Nilai Recall Class Netral:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{50}{50 + 17} * 100\%$$

$$Recall = \frac{50}{67} * 100\%$$

$$Recall = 0,75 * 100\%$$

$$Recall = 75\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,75 * 0,72)}{(0,75 + 0,72)} * 100\%$$

$$F1\ Score = \frac{1,08}{1,47} * 100\%$$

$$F1\ Score = 0,74 * 100\%$$

$$F1\ Score = 74\%$$

c) Confusion Matriks Class Negatif

$$TP = cell9 = 33$$

$$TN = cell1 + cell2 + cell4 + cell5 = 31 + 8 + 6 + 50 = 95$$

$$FN = cell8 + cell9 = 11 + 33 = 44$$

$$FP = cell3 + cell6 = 4 + 11 = 15$$

Menghitung Nilai Precision Class Negatif:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{33}{33 + 15} * 100\%$$

$$Precision = \frac{33}{48} * 100\%$$

$$Precision = 0,69 * 100\%$$

$$Precision = 69\%$$

Menghitung Nilai Recall Class Negatif:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{33}{33 + 44} * 100\%$$

$$Recall = \frac{33}{77} * 100\%$$

$$Recall = 0,43 * 100\%$$

$$Recall = 43\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,43 * 0,69)}{(0,43 + 0,69)} * 100\%$$

$$F1\ Score = \frac{0,5934}{1,12} * 100\%$$

$$F1\ Score = 0,53 * 100\%$$

$$F1\ Score = 53\%$$

Berikut Perhitungan Confusion Matriks Keseluruhan Class:

Diketahui:

a) Menghitung Akurasi

$$Accuracy = \frac{TP}{N} * 100\%$$

$$Accuracy = \frac{31 + 50 + 33}{155} * 100\%$$

$$Accuracy = \frac{114}{155} * 100\%$$

$$Accuracy = 0,74 * 100\%$$

$$Accuracy = 74\%$$

b) Menghitung Presisi

$$TP = 114$$

$$TN = 269$$

$$FN = 73$$

$$FP = 41$$

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{114}{114 + 41} * 100\%$$

$$Precision = \frac{114}{155} * 100\%$$

$$Precision = 0,74 * 100\%$$

$$Precision = 74\%$$

c) Menghitung Recall

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{114}{114 + 73} * 100\%$$

$$Recall = \frac{114}{187} * 100\%$$

$$Recall = 0,61 * 100\%$$

$$Recall = 61\%$$

d) Menghitung F1 Score

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,61 * 0,74)}{(0,61 + 0,74)} * 100\%$$

$$F1\ Score = \frac{0,9028}{1,35} * 100\%$$

$$F1\ Score = 0,67 * 100\%$$

$$F1\ Score = 67\%$$

BAB 4

IMPLEMENTASI DAN PENGUJIAN SISTEM

4.1 Implementasi Sistem

Implementasi sistem merupakan tahapan yang dilakukan setelah analisis dan perancangan sistem pada bab sebelumnya. Dari implementasi sistem ini akan dihasilkan sebuah aplikasi yang dibuat melalui penyusunan kode-kode pemrograman sehingga dapat digunakan sesuai dengan tujuan dari permasalahan yang akan diselesaikan dalam penelitian ini yaitu terkait analisis sentiment multi social media tentang pelayanan tenaga medis menggunakan metode *Naïve Bayes* dan *Support Vector Machine*.

4.1.1 Spesifikasi Kebutuhan Perangkat Keras

Spesifikasi kebutuhan perangkat keras yang digunakan dalam melakukan implementasi sistem analisis sentiment multi social media tentang pelayanan tenaga medis menggunakan metode *Naïve Bayes* dan *Support Vector Machine* adalah sebagai berikut:

1. Laptop LENOVO LEGION 5
2. Processor AMD Ryzen 7 5800H
3. Memory 16GB
4. Hard Disk 1TB

Spesifikasi kebutuhan perangkat keras yang digunakan akan mempengaruhi keluaran/*output* yang akan dihasilkan melalui proses pengujian sistem. Perangkat keras yang memiliki spesifikasi tinggi akan memproses *output* dengan cepat jika dibandingkan dengan perangkat keras yang memiliki spesifikasi rendah.

4.1.2 Spesifikasi Kebutuhan Perangkat Lunak

Spesifikasi kebutuhan perangkat lunak yang digunakan dalam pengimplementasian sistem aplikasi analisis sentiment multi social media tentang pelayanan tenaga medis menggunakan *Naïve Bayes* dan *Support Vector Machine* adalah sebagai berikut:

1. *Windows 10*
2. *Python*
3. *Sublime Text 3/ Google Collabs*

4.2 Implementasi Sistem Analisis Sentimen

4.2.1 *Scraping*

Berikut ini merupakan proses dari scrool scraping dan hasil scraping dari data di twitter.

```
tweets_data = pd.DataFrame(tweets_data)
tweets_data
# print(tweets_data)

proses Data Number : 1
proses Data Number : 2
proses Data Number : 3
proses Data Number : 4
proses Data Number : 5
proses Data Number : 6
proses Data Number : 7
proses Data Number : 8
proses Data Number : 9
proses Data Number : 10
proses Data Number : 11
proses Data Number : 12
proses Data Number : 13
proses Data Number : 14
proses Data Number : 15
proses Data Number : 16
proses Data Number : 17
proses Data Number : 18
proses Data Number : 19
proses Data Number : 20
proses Data Number : 21
proses Data Number : 22
proses Data Number : 23
proses Data Number : 24
proses Data Number : 25
proses Data Number : 26
proses Data Number : 27
proses Data Number : 28
proses Data Number : 29
proses Data Number : 30
proses Data Number : 31
proses Data Number : 32
proses Data Number : 33
proses Data Number : 34
proses Data Number : 35
proses Data Number : 36
```

Gambar 4. 1 Proses Scroll Scraping Data dari Twitter

	id	name	web	created_at	username	tweet	location
0	1535059785665478656	Diaga	None	2022-06-10 00:42:27	DiagaPaiker	@Muhdzahir @hrdbacot Biar mudah di pahami, mir...	PALEMBANG
1	1535048183424106498	Chalistamall	None	2022-06-09 23:56:21	chalistamall	#Jubahoperasi Jubah operasiIn#Apdgown Apd Gow...	
2	1535034965510934529	Romiaann	None	2022-06-09 23:03:50	romiaann	Tapi kok dokter2/tenaga medis gak jelasin ya? ...	Indonesia
3	1535031570629021696	eL	None	2022-06-09 22:50:20	sundarkeu_	Gimana ya, para dokter atau tenaga medis lain ...	svt • ncl • aспа
4	1535026983469740032	no	None	2022-06-09 22:32:06	aburnin6hill	@mediocrickey semua itu ga bakal terjadi kalo ...	
...
326	1531931873370251264	Rhendy	None	2022-06-01 09:33:15	RhendyW	@bertistuta @Santosoagusbkt @mrshananto Setelu...	Jakarta
327	1531931793296830464	Ge-mi	None	2022-06-01 09:32:58	ge_mi888	@nutrisari_mango @masdandr @tanyakanr! Sorry r...	she/her, 21+
328	1531920794321707008	★	None	2022-06-01 08:49:13	vnerasanentur	@AREAJULID Kode etiknya loh, kode etiknyaaa 🤔 ...	
329	1531920725174431749	kang_nk	None	2022-06-01 08:48:57	EnkaDhee	@afikmi jadi, apakah dokter & tenaga medis...	Sleman, Indonesia
330	1531919197814784000	Ndah	https://t.co/rqzv4uk64P	2022-06-01 08:42:53	ndahdien	Ketika tenaga medis menjadikan pasien sebagai ...	Indonesia

331 rows x 7 columns

Gambar 4. 2 Scraping Data dari Twitter

4.2.2 Preprocessing

Dibawah ini merupakan tahap dari *preprocessing*

```
data = pd.read_excel('Scraping Gabung.xlsx')
data.to_numpy()
data
```

	Username	Komen
0	DiagaPaiker	@Muhdzahir @hrdbacot Biar mudah di pahami, mir...
1	chalistamall	#Jubahoperasi Jubah operasiIn#Apdgown Apd Gow...
2	romiaann	Tapi kok dokter2/tenaga medis gak jelasin ya? ...
3	sundarkeu_	Gimana ya, para dokter atau tenaga medis lain ...
4	aburnin6hill	@mediocrickey semua itu ga bakal terjadi kalo ...
...
1941	jojo_duren88	d jktim ad ga min
1942	MS4wLjABAAAjUJt-zinMzURP7MM9h1QRJM2fJpJb0yQ8m...	di jakarta ada ga pak
1943	_semesta04	Lampung menunggu...kapan suhu... 🤔 🤔
1944	callme123lisa	Lok mana si
1945	hodijah155	kapan buka d bogor

1946 rows x 2 columns

Gambar 4. 3 Dataset

Dibawah ini menampilkan data yang telah dilakukan *preprocessing* dan menampilkan hasil dari data negatif, positive dan netral

```
[ ] df1 = pd.DataFrame(dataset)
df1.to_csv('dataset_labeling_finish.csv')
df1.head()
```

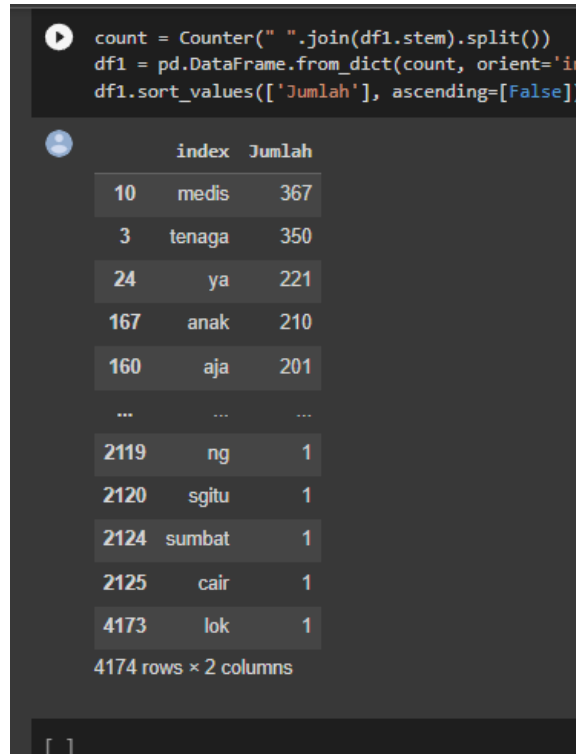
	user	tweet	stem	label	tokenizing
0	DiagaPaiker	@Muhdzahir @hrdbacot Biar mudah di pahami, mir...	biar mudah paham tenaga sehat pns dokter pppk ...	1	[biar, mudah, di, pahami, mirip, seperti, tena...
1	chalistamall	#Jubahoperasi Jubah operasiIn#Apdgown Apd Gow...	jubah operasi apd gown surgical gown jubah ope...	-1	[jubah, operasi, apd, gown, surgical, gown, ju...
2	romiaann	Tapi kok dokter2/tenaga medis gak jelasin ya? ...	dokter tenaga medis jelasin ya orang kritis hm...	-1	[tapi, kok, dokter, tenaga, medis, gak, jelasi...
3	sundarkeu_	Gimana ya, para dokter atau tenaga medis lain ...	gimana ya dokter tenaga medis tuh ngasih tau p...	-1	[gimana, ya, para, dokter, atau, tenaga, medis...
4	aburnin6hill	@mediocrickey semua itu ga bakal terjadi kalo ...	ga kalo indonesia bijak safe abortions kaya uk...	1	[semua, itu, ga, bakal, terjadi, kalo, indones...

```
print('Total Label Dari Semua data :')
print(total_label)
```

Total Label Dari Semua data :
{ 'Negative': 515, 'Positive': 407, 'Neutral': 934 }

Gambar 4. 4 Dataset Setelah Preprocessing

Dibawah ini menampilkan kata yang muncul secara *descending* (dari yang terbesar ke yang terkecil)



```
count = Counter(" ".join(df1.stem).split())
df1 = pd.DataFrame.from_dict(count, orient='index')
df1.sort_values(['Jumlah'], ascending=[False])
```

	index	Jumlah
10	medis	367
3	tenaga	350
24	ya	221
167	anak	210
160	aja	201
...
2119	ng	1
2120	sgitu	1
2124	sumbat	1
2125	cair	1
4173	lok	1

4174 rows × 2 columns

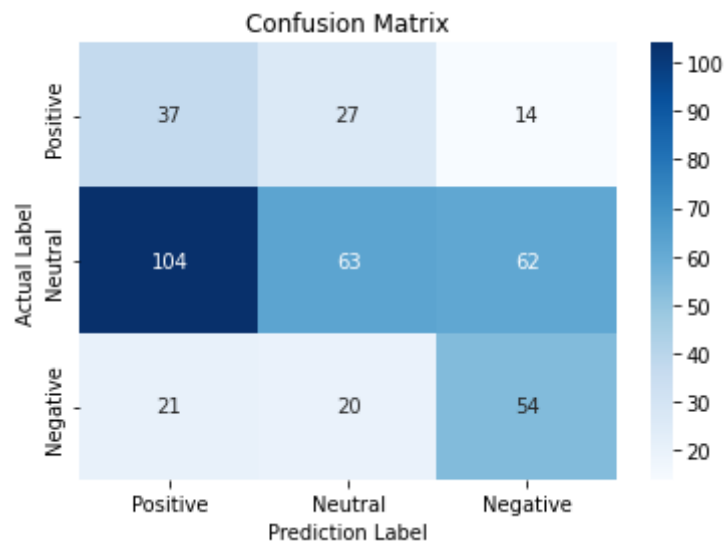
Gambar 4. 5 Menampilkan kata yang muncul secara descending

4.3 Pengujian Sistem

Pengujian sistem dilakukan untuk memeriksa sistem yang sudah dibangun. Maka akan dilakukan pengujian sistem dari facebook, tiktok dan twitter. Dibawah ini merupakan beberapa pengujian dari 3 sosial media diatas.

4.3.1 Confussion Matriks Facebook

a. Confusion Matriks *Naïve Bayes Platform* Facebook



Accuracy Score = 0.38
 Precision = 0.38
 Recall = 0.38
 F-Score = 0.38

```

# Menghitung nilai akurasi dari klasifikasi naive bayes
from sklearn.metrics import classification_report
print(classification_report(y_test,y_pred))
  
```

```

              precision    recall  f1-score   support

-1         0.42         0.57         0.48         95
0          0.57         0.28         0.37        229
1          0.23         0.47         0.31         78

accuracy          0.38         402
macro avg         0.41         0.44         0.39         402
weighted avg      0.47         0.38         0.38         402
  
```

Gambar 4. 6 Pengujian Confusion Matrix Naïve Bayes Facebook

Tabel 4. 1 Hasil Confusion Matrix

Confusion Matrix				
Actual Label	Positif	37	27	14
	Netral	104	63	62
	Negatif	21	20	54
		Positif	Netral	Negatif
Prediction Label				

Berikut Perhitungan Confusion Matriks Setiap Class:

a) Confusion Matriks Class Positif

$$TP = cell1 = 37$$

$$TN = cell5 + cell6 + cell8 + cell9 = 63 + 62 + 20 + 54 = 199$$

$$FN = cell2 + cell3 = 27 + 14 = 41$$

$$FP = cell4 + cell7 = 104 + 21 = 125$$

Menghitung Nilai Precision Class Positif:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{37}{37 + 125} * 100\%$$

$$Precision = \frac{37}{156} * 100\%$$

$$Precision = 0,23 * 100\%$$

$$Precision = 23\%$$

Menghitung Nilai Recall Class Positif:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{37}{37 + 41} * 100\%$$

$$Recall = \frac{37}{78} * 100\%$$

$$Recall = 0,47 * 100\%$$

$$Recall = 47\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,47 * 0,23)}{(0,47 + 0,23)} * 100\%$$

$$F1\ Score = \frac{0,2163}{0,7} * 100\%$$

$$F1\ Score = 0,31 * 100\%$$

$$F1\ Score = 31\%$$

b) Confusion Matriks Class Netral

$$TP = cell5 = 63$$

$$TN = cell1 + cell3 + cell7 + cell9 = 37 + 14 + 21 + 54 = 126$$

$$FN = cell4 + cell6 = 104 + 52 = 166$$

$$FP = cell2 + cell8 = 27 + 29 = 47$$

Menghitung Nilai Precision Class Netral:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{63}{63 + 47} * 100\%$$

$$Precision = \frac{63}{110} * 100\%$$

$$Precision = 0,57 * 100\%$$

$$Precision = 57\%$$

Menghitung Nilai Recall Class Netral:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{63}{63 + 166} * 100\%$$

$$Recall = \frac{63}{229} * 100\%$$

$$Recall = 0,28 * 100\%$$

$$Recall = 28\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,28 * 0,57)}{(0,28 + 0,57)} * 100\%$$

$$F1\ Score = \frac{0,3192}{0,85} * 100\%$$

$$F1\ Score = 0,37 * 100\%$$

$$F1\ Score = 37\%$$

c) Confusion Matriks Class Negatif

$$TP = cell9 = 54$$

$$TN = cell1 + cell2 + cell4 + cell5 = 37 + 27 + 104 + 63 = 231$$

$$FN = cell8 + cell9 = 20 + 54 = 74$$

$$FP = cell3 + cell6 = 14 + 62 = 76$$

Menghitung Nilai Precision Class Negatif:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{54}{54 + 76} * 100\%$$

$$Precision = \frac{54}{130} * 100\%$$

$$Precision = 0,42 * 100\%$$

$$Precision = 42\%$$

Menghitung Nilai Recall Class Negatif:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{54}{54 + 74} * 100\%$$

$$Recall = \frac{54}{128} * 100\%$$

$$Recall = 0,42 * 100\%$$

$$Recall = 42\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,42 * 0,42)}{(0,42 + 0,42)} * 100\%$$

$$F1\ Score = \frac{0,3528}{0,84} * 100\%$$

$$F1\ Score = 0,42 * 100\%$$

$$F1\ Score = 42\%$$

Berikut Perhitungan Confusion Matriks Keseluruhan Class:

Diketahui:

a) Menghitung Akurasi

$$Accuracy = \frac{TP}{N} * 100\%$$

$$Accuracy = \frac{37 + 63 + 54}{155} * 100\%$$

$$Accuracy = \frac{154}{402} * 100\%$$

$$Accuracy = 0,38 * 100\%$$

$$Accuracy = 38\%$$

b) Menghitung Presisi

$$TP = 154$$

$$TN = 556$$

$$FN = 281$$

$$FP = 248$$

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{154}{154 + 248} * 100\%$$

$$Precision = \frac{154}{402} * 100\%$$

$$Precision = 0,38 * 100\%$$

$$Precision = 38\%$$

c) Menghitung Recall

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{154}{154 + 281} * 100\%$$

$$Recall = \frac{154}{435} * 100\%$$

$$Recall = 0,35 * 100\%$$

$$Recall = 35\%$$

d) Menghitung F1 Score

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

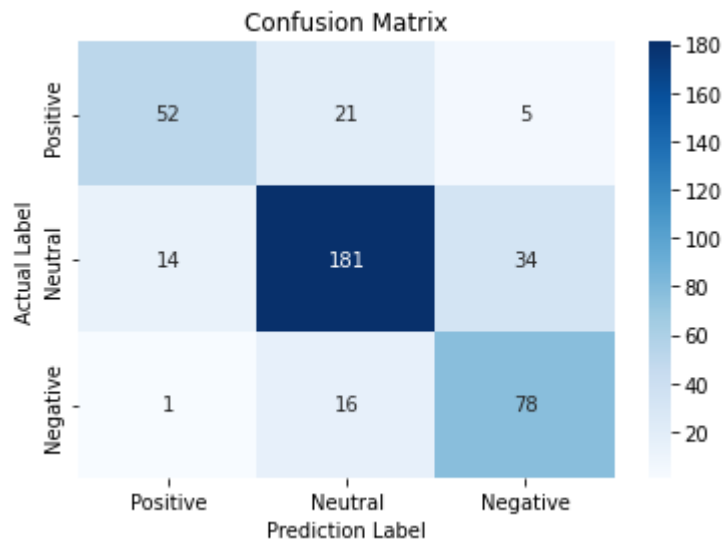
$$F1\ Score = \frac{(2 * 0,35 * 0,38)}{(0,35 + 0,38)} * 100\%$$

$$F1\ Score = \frac{0,266}{0,73} * 100\%$$

$$F1\ Score = 0,37 * 100\%$$

$F1\ Score = 37\%$

b. Confusion Matriks SVM Platfrom Facebook



Accuracy Score = 0.77

Precision = 0.77

Recall = 0.77

F-Score = 0.77

```
# Menghitung nilai akurasi dari klasifikasi SVM
from sklearn.metrics import classification_report
print(classification_report(y_test, pred))
```

	precision	recall	f1-score	support
-1	0.67	0.82	0.74	95
0	0.83	0.79	0.81	229
1	0.78	0.67	0.72	78
accuracy			0.77	402
macro avg	0.76	0.76	0.75	402
weighted avg	0.78	0.77	0.77	402

Gambar 4. 7 Pengujian Confusion Matrix SVM Facebook

Hasil Confusion Matrix:

Tabel 4. 2 Hasil Confusion Matrix

	Confusion Matrix			
Actual Label	Positif	52	21	5
	Netral	14	181	34
	Negatif	1	16	78
		Positif	Netral	Negatif
	Prediction Label			

Berikut Perhitungan Confusion Matriks Setiap Class:

a) Confusion Matriks Class Positif

$$TP = cell1 = 52$$

$$TN = cell5 + cell6 + cell8 + cell9 = 181 + 34 + 16 + 78 = 309$$

$$FN = cell2 + cell3 = 21 + 5 = 26$$

$$FP = cell4 + cell7 = 14 + 1 = 15$$

Menghitung Nilai Precision Class Positif:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{52}{52 + 15} * 100\%$$

$$Precision = \frac{52}{67} * 100\%$$

$$Precision = 0,78 * 100\%$$

$$Precision = 78\%$$

Menghitung Nilai Recall Class Positif:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{52}{52 + 26} * 100\%$$

$$Recall = \frac{52}{78} * 100\%$$

$$Recall = 0,67 * 100\%$$

$$Recall = 67\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,67 * 0,78)}{(0,67 + 0,78)} * 100\%$$

$$F1\ Score = \frac{1,0452}{1,45} * 100\%$$

$$F1\ Score = 0,72 * 100\%$$

$$F1\ Score = 72\%$$

b) Confusion Matriks Class Netral

$$TP = cell5 = 181$$

$$TN = cell1 + cell3 + cell7 + cell9 = 52 + 5 + 16 + 78 = 136$$

$$FN = cell4 + cell6 = 14 + 34 = 48$$

$$FP = cell2 + cell8 = 21 + 16 = 37$$

Menghitung Nilai Precision Class Netral:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{181}{181 + 37} * 100\%$$

$$Precision = \frac{181}{218} * 100\%$$

$$Precision = 0,83 * 100\%$$

$$Precision = 83\%$$

Menghitung Nilai Recall Class Netral:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{181}{181 + 48} * 100\%$$

$$Recall = \frac{181}{229} * 100\%$$

$$Recall = 0,81 * 100\%$$

$$Recall = 81\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,81 * 0,83)}{(0,81 + 0,83)} * 100\%$$

$$F1\ Score = \frac{1,3446}{1,64} * 100\%$$

$$F1\ Score = 0,81 * 100\%$$

$$F1\ Score = 81\%$$

c) Confusion Matriks Class Negatif

$$TP = cell9 = 78$$

$$TN = cell1 + cell2 + cell4 + cell5 = 51 + 21 + 14 + 181 = 268$$

$$FN = cell8 + cell9 = 16 + 78 = 94$$

$$FP = cell3 + cell6 = 5 + 34 = 39$$

Menghitung Nilai Precision Class Negatif:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{78}{78 + 39} * 100\%$$

$$Precision = \frac{78}{117} * 100\%$$

$$Precision = 0,67 * 100\%$$

$$Precision = 67\%$$

Menghitung Nilai Recall Class Negatif:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{78}{78 + 94} * 100\%$$

$$Recall = \frac{78}{172} * 100\%$$

$$Recall = 0,45 * 100\%$$

$$Recall = 45\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,45 * 0,67)}{(0,45 + 0,67)} * 100\%$$

$$F1\ Score = \frac{0,603}{1,12} * 100\%$$

$$F1\ Score = 0,54 * 100\%$$

$$F1\ Score = 54\%$$

Berikut Perhitungan Confusion Matriks Keseluruhan Class:

Diketahui:

a) Menghitung Akurasi

$$Accuracy = \frac{TP}{N} * 100\%$$

$$Accuracy = \frac{52 + 181 + 78}{402} * 100\%$$

$$Accuracy = \frac{311}{402} * 100\%$$

$$Accuracy = 0,77 * 100\%$$

$$Accuracy = 77\%$$

b) Menghitung Presisi

$$TP = 311$$

$$TN = 713$$

$$FN = 168$$

$$FP = 91$$

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{311}{311 + 91} * 100\%$$

$$Precision = \frac{311}{402} * 100\%$$

$$Precision = 0,77 * 100\%$$

$$Precision = 77\%$$

c) Menghitung Recall

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{311}{311 + 168} * 100\%$$

$$Recall = \frac{311}{479} * 100\%$$

$$Recall = 0,65 * 100\%$$

$$Recall = 65\%$$

d) Menghitung F1 Score

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,65 * 0,77)}{(0,65 + 0,77)} * 100\%$$

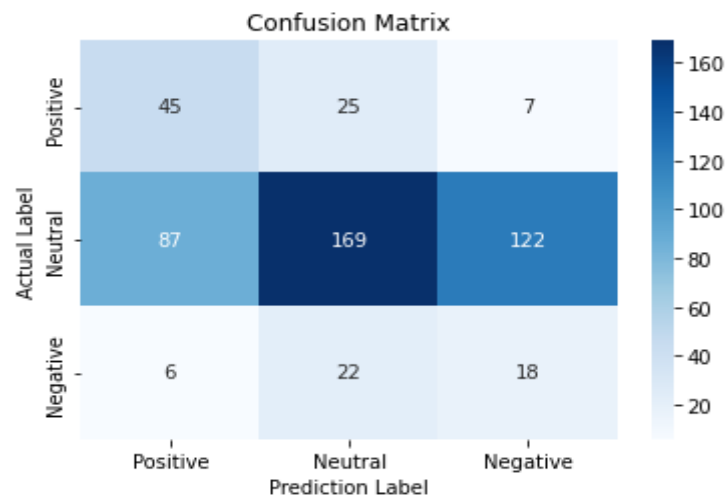
$$F1\ Score = \frac{1,001}{1,42} * 100\%$$

$$F1\ Score = 0,71 * 100\%$$

$$F1\ Score = 71\%$$

4.3.2 Confussion Matriks Tiktok

a. Confusion Matriks Naïve Bayes Platform Tiktok



```
:
_accuracy = accuracy_score(y_test, y_pred)
_precision = precision_score(y_test, y_pred, average='micro')
_recall = recall_score(y_test, y_pred, average='micro')
_fscore = f1_score(y_test, y_pred, average='micro')

print('Accuracy Score =', "{:.2f}".format(_accuracy))
print('Precision =', "{:.2f}".format(_precision))
print('Recall =', "{:.2f}".format(_recall))
print('F-Score =', "{:.2f}".format(_fscore))
```

```
Accuracy Score = 0.46
Precision = 0.46
Recall = 0.46
F-Score = 0.46
```

```
: # Menghitung nilai akurasi dari klasifikasi naive bayes
from sklearn.metrics import classification_report
print(classification_report(y_test,y_pred))
```

	precision	recall	f1-score	support
-1	0.12	0.39	0.19	46
0	0.78	0.45	0.57	378
1	0.33	0.58	0.42	77
accuracy			0.46	501
macro avg	0.41	0.47	0.39	501
weighted avg	0.65	0.46	0.51	501

Gambar 4. 8 Pengujian Confusion Matrix Naïve Bayes Tiktok

Hasil Confusion Matrix:

Tabel 4. 3 Hasil *Confusion Matrix*

	Confusion Matrix			
Actual Label	Positif	45	25	7
	Netral	87	169	122
	Negatif	6	22	18
		Positif	Netral	Negatif
	Prediction Label			

Berikut Perhitungan Confusion Matriks Setiap Class:

a) Confusion Matriks Class Positif

$$TP = \text{cell1} = 45$$

$$TN = \text{cell5} + \text{cell6} + \text{cell8} + \text{cell9} = 169 + 122 + 22 + 18 = 331$$

$$FN = \text{cell2} + \text{cell3} = 25 + 7 = 32$$

$$FP = \text{cell4} + \text{cell7} = 87 + 6 = 93$$

Menghitung Nilai Precision Class Positif:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{45}{45 + 93} * 100\%$$

$$Precision = \frac{45}{138} * 100\%$$

$$Precision = 0,33 * 100\%$$

$$Precision = 33\%$$

Menghitung Nilai Recall Class Positif:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{45}{45 + 32} * 100\%$$

$$Recall = \frac{45}{77} * 100\%$$

$$Recall = 0,58 * 100\%$$

$$Recall = 58\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,58 * 0,33)}{(0,58 + 0,33)} * 100\%$$

$$F1\ Score = \frac{0,3828}{0,91} * 100\%$$

$$F1\ Score = 0,42 * 100\%$$

$$F1\ Score = 42\%$$

b) Confusion Matriks Class Netral

$$TP = cell5 = 169$$

$$TN = cell1 + cell3 + cell7 + cell9 = 45 + 7 + 6 + 18 = 76$$

$$FN = cell4 + cell6 = 87 + 122 = 209$$

$$FP = cell2 + cell8 = 25 + 22 = 47$$

Menghitung Nilai Precision Class Netral:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{169}{169 + 47} * 100\%$$

$$Precision = \frac{169}{216} * 100\%$$

$$Precision = 0,78 * 100\%$$

$$Precision = 78\%$$

Menghitung Nilai Recall Class Netral:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{169}{169 + 209} * 100\%$$

$$Recall = \frac{169}{378} * 100\%$$

$$Recall = 0,45 * 100\%$$

$$Recall = 45\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,45 * 0,78)}{(0,45 + 0,78)} * 100\%$$

$$F1\ Score = \frac{0,702}{1,23} * 100\%$$

$$F1\ Score = 0,57 * 100\%$$

$$F1\ Score = 57\%$$

c) Confusion Matriks Class Negatif

$$TP = cell9 = 18$$

$$TN = cell1 + cell2 + cell4 + cell5 = 45 + 25 + 87 + 169 = 326$$

$$FN = cell8 + cell9 = 22 + 18 = 40$$

$$FP = cell3 + cell6 = 7 + 122 = 129$$

Menghitung Nilai Precision Class Negatif:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{18}{18 + 129} * 100\%$$

$$Precision = \frac{18}{147} * 100\%$$

$$Precision = 0,12 * 100\%$$

$$Precision = 12\%$$

Menghitung Nilai Recall Class Negatif:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{18}{18 + 40} * 100\%$$

$$Recall = \frac{18}{58} * 100\%$$

$$Recall = 0,31 * 100\%$$

$$Recall = 31\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,31 * 0,12)}{(0,31 + 0,12)} * 100\%$$

$$F1\ Score = \frac{0,0744}{0,43} * 100\%$$

$$F1\ Score = 0,18 * 100\%$$

$$F1\ Score = 18\%$$

Berikut Perhitungan Confusion Matriks Keseluruhan Class:

Diketahui:

a) Menghitung Akurasi

$$Accuracy = \frac{TP}{N} * 100\%$$

$$Accuracy = \frac{45 + 169 + 18}{501} * 100\%$$

$$Accuracy = \frac{232}{501} * 100\%$$

$$Accuracy = 0,46 * 100\%$$

$$Accuracy = 46\%$$

b) Menghitung Presisi

$$TP = 232$$

$$TN = 733$$

$$FN = 281$$

$$FP = 269$$

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{232}{232 + 269} * 100\%$$

$$Precision = \frac{232}{501} * 100\%$$

$$Precision = 0,46 * 100\%$$

$$Precision = 46\%$$

c) Menghitung Recall

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{232}{232 + 281} * 100\%$$

$$Recall = \frac{232}{513} * 100\%$$

$$Recall = 0,45 * 100\%$$

$$Recall = 45\%$$

d) Menghitung F1 Score

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

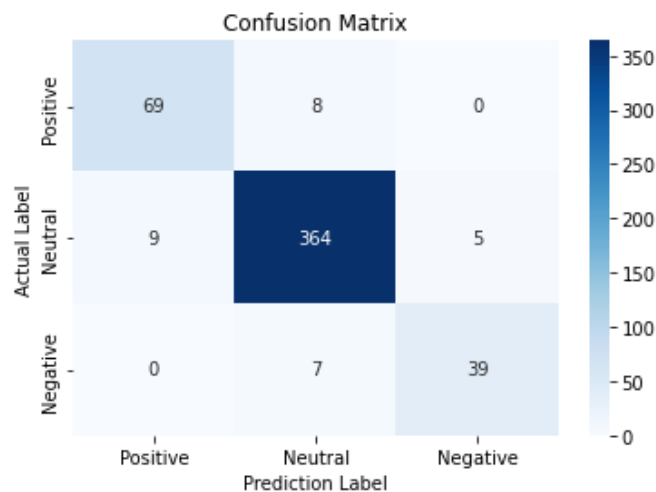
$$F1\ Score = \frac{(2 * 0,45 * 0,46)}{(0,45 + 0,46)} * 100\%$$

$$F1\ Score = \frac{0,414}{0,91} * 100\%$$

$$F1\ Score = 0,46 * 100\%$$

$$F1\ Score = 46\%$$

b. *Confusion Matriks SVM Platform Tiktok*



Accuracy Score = 0.94

Precision = 0.94

Recall = 0.94

F-Score = 0.94

```
# Menghitung nilai akurasi dari klasifikasi SVM
from sklearn.metrics import classification_report
print(classification_report(y_test, pred))
```

	precision	recall	f1-score	support
-1	0.89	0.85	0.87	46
0	0.96	0.96	0.96	378
1	0.88	0.90	0.89	77
accuracy			0.94	501
macro avg	0.91	0.90	0.91	501
weighted avg	0.94	0.94	0.94	501

Gambar 4. 9 Pengujian Confusion Matrix SVM Tiktok

Hasil Confusion Matrix:

Tabel 4. 4 Hasil Confusion Matrix

	Confusion Matrix			
Actual Label	Positif	69	8	0
	Netral	9	364	5
	Negatif	0	7	39
		Positif	Netral	Negatif
	Prediction Label			

Berikut Perhitungan Confusion Matriks Setiap Class:

a) Confusion Matriks Class Positif

$$TP = cell1 = 69$$

$$TN = cell5 + cell6 + cell8 + cell9 = 364 + 5 + 7 + 39 = 415$$

$$FN = cell2 + cell3 = 8 + 0 = 8$$

$$FP = cell4 + cell7 = 9 + 0 = 9$$

Menghitung Nilai Precision Class Positif:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{69}{69 + 9} * 100\%$$

$$Precision = \frac{69}{78} * 100\%$$

$$Precision = 0,88 * 100\%$$

$$Precision = 88\%$$

Menghitung Nilai Recall Class Positif:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{69}{69 + 8} * 100\%$$

$$Recall = \frac{69}{77} * 100\%$$

$$Recall = 0,90 * 100\%$$

$$Recall = 90\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,90 * 0,88)}{(0,90 + 0,88)} * 100\%$$

$$F1\ Score = \frac{1,584}{1,78} * 100\%$$

$$F1\ Score = 0,89 * 100\%$$

$$F1\ Score = 89\%$$

b) Confusion Matriks Class Netral

$$TP = cell5 = 364$$

$$TN = cell1 + cell3 + cell7 + cell9 = 69 + 0 + +0 + 39 = 108$$

$$FN = cell4 + cell6 = 9 + 5 = 14$$

$$FP = cell2 + cell8 = 8 + 7 = 15$$

Menghitung Nilai Precision Class Netral:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{364}{364 + 15} * 100\%$$

$$Precision = \frac{364}{379} * 100\%$$

$$Precision = 0,96 * 100\%$$

$$Precision = 96\%$$

Menghitung Nilai Recall Class Netral:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{364}{364 + 14} * 100\%$$

$$Recall = \frac{364}{378} * 100\%$$

$$Recall = 0,96 * 100\%$$

$$Recall = 96\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,96 * 0,96)}{(0,96 + 0,96)} * 100\%$$

$$F1\ Score = \frac{1,8432}{1,92} * 100\%$$

$$F1\ Score = 0,96 * 100\%$$

$$F1\ Score = 96\%$$

c) Confusion Matriks Class Negatif

$$TP = cell9 = 39$$

$$TN = cell1 + cell2 + cell4 + cell5 = 69 + 8 + 9 + 364 = 450$$

$$FN = cell8 + cell9 = 7 + 39 = 46$$

$$FP = cell3 + cell6 = 0 + 5 = 5$$

Menghitung Nilai Precision Class Negatif:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{39}{39 + 5} * 100\%$$

$$Precision = \frac{39}{44} * 100\%$$

$$Precision = 0,89 * 100\%$$

$$Precision = 89\%$$

Menghitung Nilai Recall Class Negatif:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{39}{39 + 46} * 100\%$$

$$Recall = \frac{39}{85} * 100\%$$

$$Recall = 0,46 * 100\%$$

$$Recall = 46\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,46 * 0,89)}{(0,46 + 0,89)} * 100\%$$

$$F1\ Score = \frac{0,8188}{1,35} * 100\%$$

$$F1\ Score = 0,60 * 100\%$$

$$F1\ Score = 60\%$$

Berikut Perhitungan Confusion Matriks Keseluruhan Class:

Diketahui:

a) Menghitung Akurasi

$$Accuracy = \frac{TP}{N} * 100\%$$

$$Accuracy = \frac{69 + 364 + 39}{501} * 100\%$$

$$Accuracy = \frac{472}{501} * 100\%$$

$$Accuracy = 0,94 * 100\%$$

$$Accuracy = 94\%$$

b) Menghitung Presisi

$$TP = 472$$

$$TN = 973$$

$$FN = 68$$

$$FP = 29$$

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{472}{472 + 29} * 100\%$$

$$Precision = \frac{472}{501} * 100\%$$

$$Precision = 0,94 * 100\%$$

$$Precision = 94\%$$

c) Menghitung Recall

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{472}{472 + 68} * 100\%$$

$$Recall = \frac{472}{540} * 100\%$$

$$Recall = 0,87 * 100\%$$

$$Recall = 87\%$$

d) Menghitung F1 Score

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,87 * 0,95)}{(0,87 + 0,95)} * 100\%$$

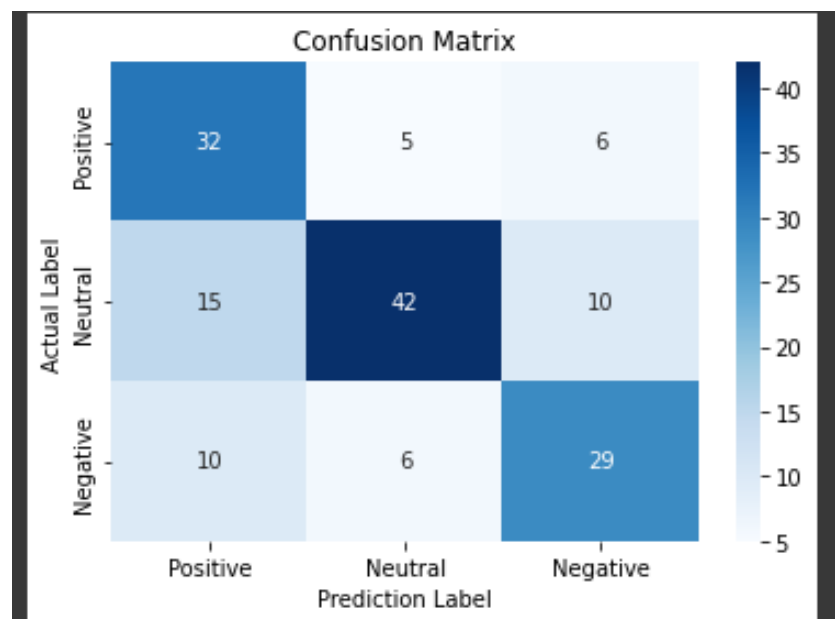
$$F1\ Score = \frac{1,6356}{1,81} * 100\%$$

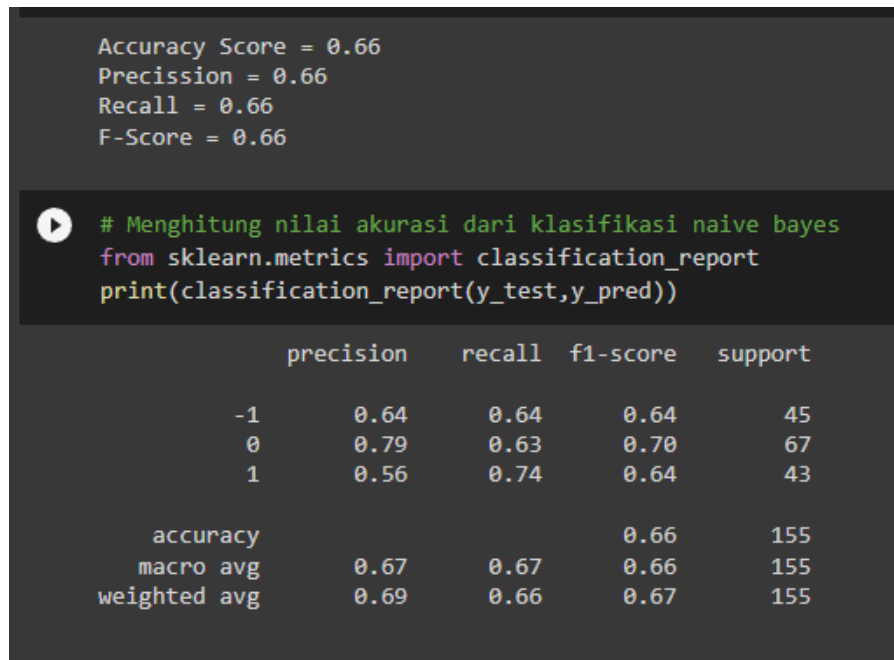
$$F1\ Score = 0,91 * 100\%$$

$$F1\ Score = 91\%$$

4.3.3 Confussion Matriks Twitter

a. Confusion Matriks *Naïve Bayes Platform* Twitter





Gambar 4. 10 Pengujian Confusion Matrix Naïve Bayes Twitter

Tabel 4. 5 Perhitungan Confusion Matrix

	Confusion Matrix			
Actual Label	Positif	32	5	6
	Netral	15	42	10
	Negatif	10	6	29
		Positif	Netral	Negatif
	Prediction Label			

Berikut Perhitungan Confusion Matriks Setiap Class:

a) Confusion Matriks Class Positif

$$TP = \text{cell1} = 32$$

$$TN = \text{cell5} + \text{cell6} + \text{cell8} + \text{cell9} = 42 + 10 + 6 + 29 = 87$$

$$FN = \text{cell2} + \text{cell3} = 5 + 6 = 11$$

$$FP = \text{cell4} + \text{cell7} = 15 + 10 = 25$$

Menghitung Nilai Precision Class Positif:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{32}{32 + 25} * 100\%$$

$$Precision = \frac{32}{57} * 100\%$$

$$Precision = 0,56 * 100\%$$

$$Precision = 56\%$$

Menghitung Nilai Recall Class Positif:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{32}{32 + 11} * 100\%$$

$$Recall = \frac{32}{43} * 100\%$$

$$Recall = 0,74 * 100\%$$

$$Recall = 74\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,74 * 0,56)}{(0,74 + 0,56)} * 100\%$$

$$F1\ Score = \frac{0,8288}{1,3} * 100\%$$

$$F1\ Score = 0,64 * 100\%$$

$$F1\ Score = 64\%$$

b) Confusion Matriks Class Netral

$$TP = cell5 = 42$$

$$TN = cell1 + cell3 + cell7 + cell9 = 36 + 6 + 10 + 29 = 77$$

$$FN = cell4 + cell6 = 15 + 10 = 25$$

$$FP = cell2 + cell8 = 5 + 6 = 11$$

Menghitung Nilai Precision Class Netral:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{42}{42 + 11} * 100\%$$

$$Precision = \frac{42}{53} * 100\%$$

$$Precision = 0,79 * 100\%$$

$$Precision = 79\%$$

Menghitung Nilai Recall Class Netral:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{42}{42 + 25} * 100\%$$

$$Recall = \frac{42}{67} * 100\%$$

$$Recall = 0,63 * 100\%$$

$$Recall = 63\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,63 * 0,79)}{(0,63 + 0,79)} * 100\%$$

$$F1\ Score = \frac{0,9954}{1,42} * 100\%$$

$$F1\ Score = 0,70 * 100\%$$

$$F1\ Score = 70\%$$

c) Confusion Matriks Class Negatif

$$TP = cell9 = 29$$

$$TN = cell1 + cell2 + cell4 + cell5 = 32 + 5 + 15 + 42 = 94$$

$$FN = cell8 + cell9 = 6 + 29 = 35$$

$$FP = cell3 + cell6 = 6 + 10 = 16$$

Menghitung Nilai Precision Class Negatif:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{29}{29 + 16} * 100\%$$

$$Precision = \frac{29}{45} * 100\%$$

$$Precision = 0,64 * 100\%$$

$$Precision = 64\%$$

Menghitung Nilai Recall Class Negatif:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{29}{29 + 35} * 100\%$$

$$Recall = \frac{29}{64} * 100\%$$

$$Recall = 0,45 * 100\%$$

$$Recall = 45\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,45 * 0,64)}{(0,45 + 0,64)} * 100\%$$

$$F1\ Score = \frac{0,576}{1,09} * 100\%$$

$$F1\ Score = 0,53 * 100\%$$

$$F1\ Score = 53\%$$

Berikut Perhitungan Confusion Matriks Keseluruhan Class:

Diketahui:

a) Menghitung Akurasi

$$Accuracy = \frac{TP}{N} * 100\%$$

$$Accuracy = \frac{32 + 42 + 29}{155} * 100\%$$

$$Accuracy = \frac{103}{155} * 100\%$$

$$Accuracy = 0,66 * 100\%$$

$$Accuracy = 66\%$$

b) Menghitung Presisi

$$TP = 103$$

$$TN = 258$$

$$FN = 71$$

$$FP = 52$$

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{103}{103 + 52} * 100\%$$

$$Precision = \frac{103}{155} * 100\%$$

$$Precision = 0,66 * 100\%$$

$$Precision = 66\%$$

c) Menghitung Recall

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{103}{103 + 71} * 100\%$$

$$Recall = \frac{103}{174} * 100\%$$

$$Recall = 0,59 * 100\%$$

$$Recall = 59\%$$

d) Menghitung F1 Score

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

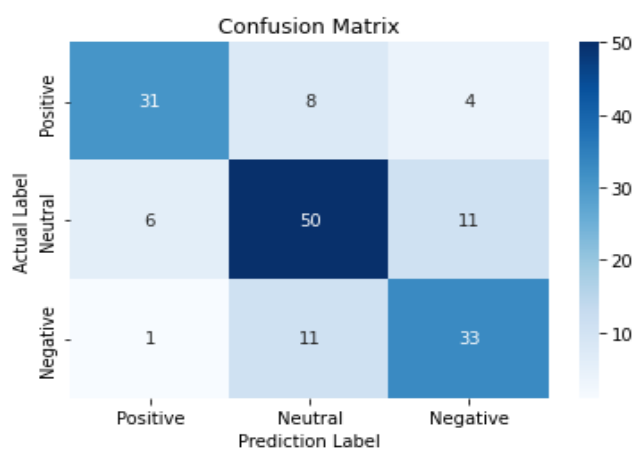
$$F1\ Score = \frac{(2 * 0,59 * 0,66)}{(0,59 + 0,66)} * 100\%$$

$$F1\ Score = \frac{0,7788}{1,25} * 100\%$$

$$F1\ Score = 0,63 * 100\%$$

$$F1\ Score = 63\%$$

b. Confusion Matriks *Support Vector Machine Platform Twitter*



```

Accuracy Score = 0.74
Precision = 0.74
Recall = 0.74
F-Score = 0.74

[ ] # Menghitung nilai akurasi dari klasifikasi SVM
from sklearn.metrics import classification_report
print(classification_report(y_test, pred))

```

	precision	recall	f1-score	support
-1	0.69	0.73	0.71	45
0	0.72	0.75	0.74	67
1	0.82	0.72	0.77	43
accuracy			0.74	155
macro avg	0.74	0.73	0.74	155
weighted avg	0.74	0.74	0.74	155

Gambar 4. 11 Pengujian Confusion Matrix Support Vector Machine
Twitter

Hasil Confusion Matrix:

Tabel 4. 6 Hasil Confusion Matrix

Actual Label	Confusion Matrix			
	Positif	31	8	4
	Netral	6	50	11
	Negatif	1	11	33
		Positif	Netral	Negatif
Prediction Label				

Berikut Perhitungan Confusion Matriks Setiap Class:

a) Confusion Matriks Class Positif

$$TP = cell1 = 31$$

$$TN = cell5 + cell6 + cell8 + cell9 = 50 + 11 + 11 + 33 = 105$$

$$FN = cell2 + cell3 = 8 + 4 = 12$$

$$FP = cell4 + cell7 = 6 + 1 = 7$$

Menghitung Nilai Precision Class Positif:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{31}{31 + 7} * 100\%$$

$$Precision = \frac{31}{38} * 100\%$$

$$Precision = 0,82 * 100\%$$

$$Precision = 82\%$$

Menghitung Nilai Recall Class Positif:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{31}{31 + 12} * 100\%$$

$$Recall = \frac{31}{43} * 100\%$$

$$Recall = 0,72 * 100\%$$

$$Recall = 72\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,72 * 0,82)}{(0,72 + 0,82)} * 100\%$$

$$F1\ Score = \frac{1,1808}{1,54} * 100\%$$

$$F1\ Score = 0,77 * 100\%$$

$$F1\ Score = 77\%$$

b) Confusion Matriks Class Netral

$$TP = cell5 = 50$$

$$TN = cell1 + cell3 + cell7 + cell9 = 31 + 4 + 1 + 33 = 69$$

$$FN = cell4 + cell6 = 6 + 11 = 17$$

$$FP = cell2 + cell8 = 8 + 11 = 19$$

Menghitung Nilai Precision Class Netral:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{50}{50 + 19} * 100\%$$

$$Precision = \frac{50}{69} * 100\%$$

$$Precision = 0,72 * 100\%$$

$$Precision = 72\%$$

Menghitung Nilai Recall Class Netral:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{50}{50 + 17} * 100\%$$

$$Recall = \frac{50}{67} * 100\%$$

$$Recall = 0,75 * 100\%$$

$$Recall = 75\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,75 * 0,72)}{(0,75 + 0,72)} * 100\%$$

$$F1\ Score = \frac{1,08}{1,47} * 100\%$$

$$F1\ Score = 0,74 * 100\%$$

$$F1\ Score = 74\%$$

c) Confusion Matriks Class Negatif

$$TP = cell9 = 33$$

$$TN = cell1 + cell2 + cell4 + cell5 = 31 + 8 + 6 + 50 = 95$$

$$FN = cell8 + cell9 = 11 + 33 = 44$$

$$FP = cell3 + cell6 = 4 + 11 = 15$$

Menghitung Nilai Precision Class Negatif:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{33}{33 + 15} * 100\%$$

$$Precision = \frac{33}{48} * 100\%$$

$$Precision = 0,69 * 100\% = 69\%$$

Menghitung Nilai Recall Class Negatif:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{33}{33 + 44} * 100\%$$

$$Recall = \frac{33}{77} * 100\%$$

$$Recall = 0,43 * 100\%$$

$$Recall = 43\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,43 * 0,69)}{(0,43 + 0,69)} * 100\%$$

$$F1\ Score = \frac{0,5934}{1,12} * 100\%$$

$$F1\ Score = 0,53 * 100\%$$

$$F1\ Score = 53\%$$

Berikut Perhitungan Confusion Matriks Keseluruhan Class:

Diketahui:

a) Menghitung Akurasi

$$Accuracy = \frac{TP}{N} * 100\%$$

$$Accuracy = \frac{31 + 50 + 33}{155} * 100\%$$

$$Accuracy = \frac{114}{155} * 100\%$$

$$Accuracy = 0,74 * 100\%$$

$$Accuracy = 74\%$$

b) Menghitung Presisi

$$TP = 114$$

$$TN = 269$$

$$FN = 73$$

$$FP = 41$$

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{114}{114 + 41} * 100\%$$

$$Precision = \frac{114}{155} * 100\%$$

$$Precision = 0,74 * 100\%$$

$$Precision = 74\%$$

c) Menghitung Recall

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{114}{114 + 73} * 100\%$$

$$Recall = \frac{114}{187} * 100\%$$

$$Recall = 0,61 * 100\% = 61\%$$

d) Menghitung F1 Score

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,61 * 0,74)}{(0,61 + 0,74)} * 100\%$$

$$F1\ Score = \frac{0,9028}{1,35} * 100\%$$

$$F1\ Score = 0,67 * 100\%$$

$$F1\ Score = 67\%$$

d) Confusion Matriks Class Negatif

$$TP = cell9 = 33$$

$$TN = cell1 + cell2 + cell4 + cell5 = 31 + 8 + 6 + 50 = 95$$

$$FN = cell8 + cell9 = 11 + 33 = 44$$

$$FP = cell3 + cell6 = 4 + 11 = 15$$

Menghitung Nilai Precision Class Negatif:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

$$Precision = \frac{33}{33 + 15} * 100\%$$

$$Precision = \frac{33}{48} * 100\%$$

$$Precision = 0,69 * 100\%$$

$$Precision = 69\%$$

Menghitung Nilai Recall Class Negatif:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

$$Recall = \frac{33}{33 + 44} * 100\%$$

$$Recall = \frac{33}{77} * 100\%$$

$$Recall = 0,43 * 100\%$$

$$Recall = 43\%$$

Menghitung Nilai F1 Score:

$$F1\ Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} * 100\%$$

$$F1\ Score = \frac{(2 * 0,43 * 0,69)}{(0,43 + 0,69)} * 100\%$$

$$F1\ Score = \frac{0,5934}{1,12} * 100\%$$

$$F1\ Score = 0,53 * 100\%$$

$$F1\ Score = 53\%$$

Hasil dari seluruh pengujian analisis sentiment pada *platform* facebook menggunakan *Naïve Bayes* menghasilkan akurasi sebesar 37% dan menggunakan SVM menghasilkan akurasi sebesar 71%, pengujian analisis sentiment pada *platform* tiktok menggunakan *Naïve Bayes* menghasilkan akurasi sebesar 46% dan

menggunakan SVM menghasilkan akurasi sebesar 91% sedangkan pengujian analisis sentiment pada *platform* twitter menggunakan *Naïve Bayes* menghasilkan akurasi sebesar 63% dan menggunakan SVM menghasilkan akurasi sebesar 67%. Berdasarkan hasil pengujian yang didapat, lebih mudah mencrawling data twitter dibanding facebook dan tiktok.

Berdasarkan hasil pengujian tersebut dapat disimpulkan bahwa algoritma SVM memiliki akurasi yang lebih baik daripada algoritma *Naïve Bayes*, dimana perbedaan akurasi keduanya terlampau jauh terdapat pada *platform* tiktok yaitu algoritma *Naïve Bayes* menghasilkan akurasi sebesar 46% sedangkan SVM menghasilkan akurasi sebesar 91%.

Berdasarkan hasil token yang didapat yang diperoleh dari hasil TFIDF, Running time yang tercepat ialah *Naïve Bayes* dibandingkan SVM, tetapi tingkat akurasi yang terbaik ialah SVM.

Berdasarkan jumlah data yang didapat, pengguna Twitter menyebutkan 29.33% Positif, 28.52% Negatif, dan 42.13% Neutral. Adapun pengguna Tiktok menyebutkan 14.94% Positif, 9.69% Negatif, dan 75.36% Neutral. Sedangkan pengguna Facebook menyebutkan 19.19% Positif, 25.91% Negatif, dan 54.89% Neutral.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan pengujian yang telah dilakukan pada analisis sentimen tentang pelayanan tenaga medis menggunakan metode *Naïve Bayes* dan *Support Vector Machine*, maka kesimpulan yang didapatkan oleh penulis adalah sebagai berikut:

1. Penelitian ini berhasil mengetahui kecenderungan opini dari banyaknya komentar yang ada di media sosial terkait pelayanan tenaga medis, dengan melakukan *scraping* kemudian melakukan analisis sentiment pada data komentar dari berbagai media sosial yang dilakukan secara otomatis dan cepat.
2. *Preprocessing* pada komentar efektif menghasilkan kalimat yang penting terhadap proses analisis sentiment.
3. Pengujian analisis sentiment pada *platform* facebook menggunakan *Naïve Bayes* menghasilkan akurasi sebesar 37% dan menggunakan SVM menghasilkan akurasi sebesar 71%.
4. Pengujian analisis sentiment pada *platform* tiktok menggunakan *Naïve Bayes* menghasilkan akurasi sebesar 46% dan menggunakan SVM menghasilkan akurasi sebesar 91%.
5. Pengujian analisis sentiment pada *platform* twitter menggunakan *Naïve Bayes* menghasilkan akurasi sebesar 63% dan menggunakan SVM menghasilkan akurasi sebesar 67%

5.2 Saran

Adapun saran yang dapat diberikan oleh penulis setelah melakukan penelitian ini terkait perbaikan dan untuk pengembangan yang dapat dilakukan kepada penelitian selanjutnya adalah dapat menggunakan algoritma yang lain sebagai perbandingan agar dapat menghasilkan hasil pengklasifikasian analisis sentiment yang lebih baik dan lebih akurat.

DAFTAR PUSTAKA

- Ailiyya, Sabrah. 2020. Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Tokopedia Menggunakan *Support Vector Machine*. Skripsi. Universitas Islam Negeri Syarif Hidayatullah.
- D.V. Nagarjuna Devi,dkk 2016. A Feature Based Approach for Sentiment Analysis by Using *Support Vector Machine* Diakses dari A Feature Based Approach for Sentiment Analysis by Using Support Vector Machine | IEEE Conference Publication | IEEE Xplore Diakses pada tanggal 5 Maret 2022.
- Fikri, M. I., Sabrilla, T. S., & Azhar, Y. (2020). Perbandingan Metode *Naive Bayes* dan *Support Vector Machine* Pada Analisis Sentimen Twitter. *SMATIKA Jurnal*.
- Gunawan, D., & Riana, D. (2020). Komparasi Algoritma SVM dan *Naive Bayes* dengan Algoritma Genetika pada Analisis Sentimen Calon Gubernur Jabar 2018- 2023. *Jurnal Teknik Komputer AMIK BSI*.
- Herlinawati, N., & Yuliani, Y. (2020). Analisis Sentimen Zoom Cloud Meetings di PlayStore Menggunakan *Naive Bayes* dan SVM. *Journal of Computer Engineering System and Science*.
- Liu, B. 2010. *Handbook of Natural Language Processing, chapter SentimentAnalysis and Analysis*, 2nd Edition. Chapman & Hall / CRC Press.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.
- Nikolay Butakov,dkk,2018. Unified domain-specific language for collecting and processing data of social media Diakses dari [Unified domain-specific language for collecting and processing data of social media | SpringerLink](#) Diakses pada tanggal 5 Maret 2022.
- Osaldi, Gregorius Bryan. 2021. Analisis Sentimen Pembelajaran Daring Pada Media Sosial Twitter Menggunakan Multinomial *Naive Bayes* dan *Support Vector Machine*. Skripsi. Universitas Sanata Dharma.
- Saepulrohman, A., Saepudin, S., & Gustian, D. (2021). Analisis Sentimen Kepuasan Pengguna Aplikasi Whatsapp Menggunakan Algoritma *Naive Bayes* dan

Support Vector Machine. Accounting Information Systems and Information Technology Business Enterprise.

Samsir, Ambiyar,dkk. (2021). Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 157-163.

Sidiq dkk. 2020. Sentimen Analisis Komentar Toxic pada Grup Facebook Game Online Menggunakan Klasifikasi Naïve Bayes. *Jurnal Informatika Universitas Pamulang*, 5(3), 356-363.

Wiguna, Cahya Daksa. (2012). *Analisis Pengaruh Kualitas Pelayanan, Kualitas Tenaga Medis dan Fasilitas Terhadap Kepuasan Pasien Rawat Inap di RSIA UMMU Hani Purbalingga*. Skripsi. Universitas Diponegoro. Hal. 1-2

Zaenal Abidin, D. N. (2021). Penerapan Algoritma *Naive Bayes* untuk Analisis Sentimen Review Data Twitter BMKG Nasional. *Jurnal TEKNO KOMPAK*, 131-145.