

BAB II

TINJAUAN PUSTAKA

A. Hasil Penelitian Terdahulu

Penelitian yang dilakukan (Tineges et al., 2020) dengan judul “Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi *Support Vector Machine* (SVM)” mempunyai tujuan yaitu untuk mendapatkan model klasifikasi sentimen menggunakan *Support Vector Machine* (SVM) dan untuk mengetahui seberapa besar akurasi yang dihasilkan oleh metode SVM yang diterapkan pada analisis sentiment terhadap layanan Indihome, serta untuk mengetahui seberapa puas pengguna layanan Indihome berdasarkan Twitter. Pada penelitian ini menggunakan metode *Support Vector Machine* (SVM) serta menggunakan klasifikasi kelas berupa positif dan negatif. Hasil dari penelitian ini menunjukkan tingkat *accuracy* 87%, *precision* 86%, *recall* 95%, eror rate 13%, dan *f1-score* 90%.

Penelitian yang dilakukan (Alhaq et al., 2021) dengan judul “Penerapan Metode *Support Vector Machine* Untuk Analisis Sentimen Pengguna Twitter” mempunyai tujuan yaitu untuk mengetahui tanggapan atau sentimen melalui media sosial Twitter mengenai marketplace Bukalapak. Tanggapan atau pendapat pengguna Bukalapak di media sosial Twitter diklasifikasikan menggunakan metode SVM dengan klasifikasi positif, negatif, dan netral. Hasil klasifikasi menggunakan metode *Support Vector Machine* (SVM) diperoleh tingkat akurasi sebesar 93%.

penelitian yang dilakukan (Pravina et al., 2019) dengan judul “Analisis sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme *Support Vector Machine* (SVM)” mempunyai tujuan yaitu untuk mengklasifikasikan suatu *tweet* yang berisikan opini termasuk ke dalam kelas apa sehingga membantu perusahaan maskapai penerbangan dalam hal evaluasi peningkatan pelayanan serta dapat membantu masyarakat dalam memilih maskapai penerbangan yang tepat. Pada penelitian ini terdapat kelas positif dan negatif serta menggunakan fitur *Lexicon Based* yang dapat menerima opini berbahasa lain selain Bahasa Indonesia untuk melakukan analisis sentimen. Algoritma *support vector machine* diterapkan untuk melakukan klasifikasi. Hasil dari penelitian ini menunjukkan parameter optimal dan pengaruh penggunaan *Lexicon Based Features*. Dengan digunakan parameter C bernilai 10 dan learning rate bernilai 0,03 serta digunakan *Lexicon Based Features* dengan iterasi sebanyak 50 kali memberikan hasil *accuracy* sebesar 40%, *precision* 40%, 100% *recall*, dan *f-measure* sebesar 57,14%.

Penelitian yang dilakukan (Putri et al., 2022) dengan judul “Analisis Sentimen *E-Commerce* Lazada pada jejaring Sosial Twitter Menggunakan Algoritma *Support Vector Machine*” mempunyai tujuan yaitu untuk menganalisis sentimen masyarakat terkait pendapat dan komentar tentang *e-commerce* Lazada pada media sosial Twitter dalam rangka meningkatkan kualitas pelayanan perusahaan. Pada penelitian ini

menggunakan Algoritma *Support Vector Machine* dengan klasifikasi positif, negatif, dan netral serta menggunakan dataset berjumlah 300 *tweet*. Hasil dari penelitian ini menunjukkan Performa algoritma *Support Vector Machine* dalam proses *training* dan *testing* telah menghasilkan nilai *accuracy*, *precision*, *recall* dan *F1 score* diatas 80 %. Hasil analisis sentimen dari penelitian ini dapat digunakan sebagai tolak ukur untuk peningkatan layanan perusahaan dan memberikan wawasan kepada masyarakat.

Penelitian yang dilakukan (Sari & Haranto, 2019) dengan judul “Implementasi *Support Vector Machine* Untuk Analisis Sentimen pengguna Twitter Terhadap Pelayanan Telkom dan Biznet” mempunyai tujuan yaitu untuk mengetahui presentase positif dan negatif dari pengguna Telkom dan Biznet pada media sosial Twitter sehingga kedua perusahaan tersebut bisa meningkatkan pelayanan yang lebih baik. Pada penelitian ini menggunakan data dengan jumlah dataset sebanyak 500 *tweet* yang berasal dari *crawling* data twitter, terdapat 250 *tweet* yang dijadikan dataset pada masing-masing objek. Sejumlah data tersebut akan dipergunakan untuk data training serta data testing dalam proses pembuatan model menggunakan algoritma *Support Vector Machine*. Hasil dari penelitian ini menunjukkan nilai *accuracy* 79,6%, *precision* 76,5%, *recall* 72,8% , dan *F1-score* 74,6% untuk Telkom, serta *accuracy* 83,2%, *precision* 78,8%, *recall* 71,6%, dan *F1-score* 75% untuk Biznet.

Penelitian yang dilakukan (Darwis et al., 2021) dengan judul “Penerapan Algoritma Naïve Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional” mempunyai tujuan yaitu melakukan pencarian komentar negatif, positif, dan netral dengan review data Twitter BMKG Nasional menggunakan algoritma naive bayes, lalu melakukan perhitungan data negatif, positif, dan netral menggunakan algoritma naïve bayes, serta dapat menghasilkan klasifikasi berdasarkan nilai yang didapat dengan algoritma *naive bayes*. Hasil Penelitian ini masuk kedalam *fined grained sentiment analysis* yaitu analisis pada suatu kalimat komentar. Data tersebut akan diproses menggunakan text mining, kemudian dilanjutkan dengan mengklasifikasikan *Tweet* ke dalam tiga kelas, yaitu positif, negatif, dan netral. Klasifikasi ini menggunakan algoritma *naive bayes*. Hasil uji akurasi pada metode *naive bayes* untuk klasifikasi yaitu 69.97%.

Penelitian yang dilakukan (Gifari et al., 2022) dengan judul “Analisis Sentimen *Review* Film Menggunakan TF-IDF dan *Support Vector Machine*” mempunyai tujuan yaitu untuk menjelaskan analisis sentimen pada dokumen film. Metode yang digunakan pada penelitian ini adalah TF-IDF (*Term Frequency-Inverse Document Frequency*) dan SVM (*Support Vector Machine*). Penelitian dilakukan menggunakan data yang diperoleh dari twitter berjumlah 200 *tweet*, yang didapat melalui proses *crawling* dengan batasan hanya *tweet* yang berbahasa indonesia saja. Dari uji skenario yang dilakukan, diketahui bahwa algoritma TF-IDF dan SVM

dapat digunakan untuk kasus *review* film dengan nilai *Accuracy* 85%, nilai *Precision* 100%, nilai *Recall* 70%, dan nilai *F1-Score* sebesar 82%.

Penelitian yang dilakukan (Hidayat et al., 2021) dengan judul “Analisis Sentimen Pemindahan Ibu Kota Pada Twitter Dengan Metode *Support Vector Machine*” mempunyai tujuan yaitu untuk mengklasifikasikan opini publik tentang topik pemindahan Ibu Kota Indonesia, data penelitian berasal dari ribuan tweet yang telah dikumpulkan dan disaring. Data *Tweet* dari 22-29 Maret 2020 telah di proses menjadi 992 tweet dan terdiri dari 221 data dengan label positif dan 771 data negatif. Hasil Penelitian ini menggunakan dengan metode SVM yang memiliki akurasi 77,72% dan dikombinasikan dengan TFIDF yang meningkatkan akurasinya menjadi 78,33%.

Penelitian yang dilakukan oleh (Bourequat & Mourad, 2021) dengan judul “*Sentiment Analysis Approach for Analyzing iPhone Release using Support Vector Machine*” mempunyai tujuan yaitu untuk mengetahui respon masyarakat terkait perilisan produk baru dari Apple apakah mendapatkan respon yang baik dari masyarakat atau tidak. Data penelitian berasal dari Twitter yang berupa *Tweet* atau cuitan dengan jumlah 2000 data. Penelitian ini menggunakan metode support vector machine dan menghasilkan tingkat akurasi, presisi, recall, dan skor F1 menghasilkan 89,21%, 92,43%, 95,53%, dan 93,95.

Penelitian yang dilakukan oleh (Sontayasara et al., 2021) dengan judul “*Twitter Sentiment Analysis of Bangkok Tourism During COVID-19*

Pandemic Using Support Vector Machine Algorithm” mempunyai tujuan yaitu untuk mengidentifikasi tanggapan wisatawan atau bagaimana perasaan wisatawan asing terhadap Wilayah metropolitan Bangkok pada awal pandemi di Thailand, dengan menganalisis ulasan wisatawan melalui media sosial Twitter. Data diambil menggunakan bahasa Inggris dengan keyword "Bangkok" dimulai dari tanggal 3 April 2020 sampai 30 April 2020, Analisis sentimen dilakukan dengan menggunakan algoritma *support vector machine*. Hasilnya menunjukkan akurasi klasifikasi 71,03% berdasarkan tiga klasifikasi sentimen positif, negatif, dan netral.

Penelitian yang dilakukan oleh (Rahardi et al., 2022) dengan judul “*Sentiment Analysis of Covid-19 Vaccination using Support Vector Machine in Indonesia*” mempunyai tujuan yaitu Penelitian ini bertujuan untuk mengetahui respon masyarakat terhadap penerimaan program vaksinasi. Pengambilan data dilakukan melalui Twitter pada tahun 2021. Selanjutnya data tersebut diolah menggunakan klasifikasi *Support Vector Machine*. Hasil klasifikasi dievaluasi dengan *confusion matrix*. Hasil penelitian menghasilkan 56,80% positif, 33,75% netral, dan 9,45% negatif. Keakuratan model tertinggi diperoleh kernel RBF sebesar 92%, kernel linier dan polinomial memperoleh akurasi 90%, dan kernel sigmoid memperoleh akurasi 89%.

Tabel 2.1 Penelitian Terdahulu

No	Judul Artikel	Metode	Hasil	Penelitian yang Akan Dilakukan
1.	Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi <i>Support Vector Machine</i> (SVM) (2020)	<i>Support Vector Machine</i> (SVM)	Hasil dari penelitian ini menunjukkan tingkat <i>accuracy</i> 87%, <i>precision</i> 86%, <i>recall</i> 95%, eror rate 13%, dan <i>fi-score</i> 90%	Analisis Sentimen Masyarakat Terhadap <i>E-Commerce</i> Pada Media Sosial Twitter Studi Kasus (Shopee dan Tokopedia) Dengan tujuan untuk mengklasifikasikan nilai sentimen dan menentukan kelas dari dataset cuitan (<i>tweets</i>) dari pengguna media sosial Twitter yang berkaitan dengan topik <i>E-Commerce</i> Shopee dan Tokopedia
2.	Penerapan Metode <i>Support Vector Machine</i> Untuk Analisis Sentimen Pengguna Twitter (2021)	<i>Support Vector Machine</i>	Pada penelitian ini menggunakan klasifikasi positif, negatif, dan netral. Hasil klasifikasi menggunakan metode <i>Support Vector Machine</i> (SVM) diperoleh tingkat akurasi sebesar 93%.	
3.	Analisis sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme <i>Support Vector Machine</i> (SVM) (2019)	<i>Support Vector Machine</i> (SVM)	Hasil dari penelitian ini menunjukkan parameter optimal dan pengaruh penggunaan <i>Lexicon Based Features</i> . Dengan digunakan parameter C bernilai 10 dan learning rate bernilai 0,03 serta digunakan <i>Lexicon Based Features</i> dengan iterasi sebanyak 50 kali memberikan hasil <i>accuracy</i> sebesar 40%, <i>precision</i> 40%, 100% <i>recall</i> , dan <i>f-measure</i> sebesar 57,14%.	

No	Judul Artikel	Metode	Hasil	Penelitian yang Akan Dilakukan
4.	Analisis Sentimen <i>E-Commerce</i> Lazada pada jejaring Sosial Twitter Menggunakan Algoritma <i>Support Vector Machine</i> (2022)	<i>Support Vector Machine</i>	Hasil dari penelitian ini menunjukkan Performa algoritma <i>Support Vector Machine</i> dalam proses <i>training</i> dan <i>testing</i> telah menghasilkan nilai <i>accuracy</i> , <i>precision</i> , <i>recall</i> dan <i>F1 score</i> diatas 80 %. Hasil analisis sentimen dari penelitian ini dapat digunakan sebagai tolak ukur untuk peningkatan layanan perusahaan dan memberikan wawasan kepada masyarakat.	menggunakan algoritma <i>Support Vector Machine</i> (SVM) menggunakan klasifikasi positif dan negatif.
5.	Implementasi <i>Support Vector Machine</i> Untuk Analisis Sentimen pengguna Twitter Terhadap Pelayanan Telkom dan Biznet (2019)	<i>Support Vector Machine</i>	Hasil dari penelitian ini menunjukkan nilai <i>accuracy</i> 79,6%, <i>precision</i> 76,5%, <i>recall</i> 72,8% , dan <i>F1-score</i> 74,6% untuk Telkom, serta <i>accuracy</i> 83,2%, <i>precision</i> 78,8%, <i>recall</i> 71,6%, dan <i>F1-score</i> 75% untuk Biznet.	

No.	Judul Artikel	Metode	Hasil	Penelitian yang akan dilakukan
6.	Penerapan Algoritma Naïve Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional (2021)	Naïve Bayes	Hasil Penelitian ini masuk kedalam fined grained sentiment analysis yaitu analisis pada suatu kalimat komentar. Data tersebut akan diproses menggunakan text mining, kemudian dilanjutkan dengan mengklasifikasikan Tweet ke dalam tiga kelas, yaitu positif, negatif, dan netral. Klasifikasi ini menggunakan algoritma naive bayes. Hasil uji akurasi pada metode naive bayes untuk klasifikasi yaitu 69.97%.	
7.	Analisis Sentimen <i>Review</i> Film Menggunakan TF-IDF dan <i>Support Vector Machine</i> (2022)	<i>Support Vector Machine</i>	Dari uji skenario yang dilakukan, diketahui bahwa algoritma TF-IDF dan SVM dapat digunakan untuk kasus <i>review</i> film dengan nilai <i>Accuracy</i> 85%, nilai <i>Precision</i> 100%, nilai <i>Recall</i> 70%, dan nilai <i>F1-Score</i> sebesar 82%.	
8.	Analisis Sentimen Pemindahan Ibu Kota Pada Twitter Dengan Metode <i>Support Vector Machine</i> (2021)	<i>Support Vector Machine</i>	Hasil Penelitian ini menggunakan dengan metode SVM yang memiliki akurasi 77,72% dan dikombinasikan dengan TFIDF yang meningkatkan akurasinya menjadi 78,33%.	

No.	Judul Artikel	Metode	Hasil	Penelitian yang akan dilakukan
9.	<i>Sentiment Analysis Approach for Analyzing iPhone Release using Support Vector Machine (2021)</i>	<i>Support Vector Machine</i>	Penelitian ini menggunakan metode support vector machine dan menghasilkan tingkat akurasi, presisi, recall, dan skor F1 menghasilkan 89,21%, 92,43%, 95,53%, dan 93,95.	
10.	<i>Twitter Sentiment Analysis of Bangkok Tourism During COVID-19 Pandemic Using Support Vector Machine Algorithm (2021)</i>	<i>Support Vector Machine</i>	Analisis sentimen dilakukan dengan menggunakan algoritma <i>support vector machine</i> . Hasilnya menunjukkan akurasi klasifikasi 71,03% berdasarkan tiga klasifikasi sentimen positif, negatif, dan netral.	
11.	<i>Sentiment Analysis of Covid-19 Vaccination using Support Vector Machine in Indonesia (2022)</i>	<i>Support Vector Machine</i>	Hasil penelitian menghasilkan 56,80% positif, 33,75% netral, dan 9,45% negatif. Keakuratan model tertinggi diperoleh kernel RBF sebesar 92%, kernel linier dan polinomial memperoleh akurasi 90%, dan kernel sigmoid memperoleh akurasi 89%	

B. Landasan Teori

1. Analisis Sentimen

Sentimen analisis dapat dikenal sebagai penambangan opini atau proses memahami data teks untuk mendapatkan informasi. Analisis sentimen dapat dimanfaatkan untuk beragam hal seperti mengetahui apa yang salah pada suatu produk, pendapat masyarakat terhadap suatu fitur aplikasi atau suatu topik perbincangan tertentu (Pane et al., 2021). Menurut (Yasid, 2019) analisis sentimen ini digunakan untuk melihat kecenderungan opini terhadap suatu objek apakah beropini positif, negatif atau netral. Secara istilah, analisis sentimen adalah deteksi sikap-sikap (*attitudes*) terhadap objek atau orang. Dari miliaran data cuitan di twitter, dapat dilakukan analisis sentimen untuk menemukan berapa presentase sentimen positif dan berapa presentase sentimen negatif terhadap seseorang, perusahaan, institusi, kelompok, atau sebuah situasi tertentu (Alhaq et al., 2021).

2. E-Commerce

E-Commercee (Electronic Commercee) merupakan proses pembelian dan penjualan jasa atau produk antara dua belah pihak melalui internet (*commerce net*) dan sejenis mekanisme bisnis elektronik dengan fokus pada transaksi bisnis berbasis individu dengan menggunakan internet sebagai media pertukaran barang atau jasa baik antar instansi atau individu dengan instansi (Sri et al., 2011).

3. Media Sosial

Media sosial adalah sebuah media *online*, dengan para penggunanya bisa dengan mudah berpartisipasi, berbagi, dan menciptakan isi meliputi blog, jejaring sosial, wiki, forum dan dunia virtual. Blog, jejaring sosial dan wiki merupakan bentuk media sosial yang paling umum digunakan oleh masyarakat di seluruh dunia. Media sosial juga dapat diartikan sebagai media online yang mendukung interaksi sosial dan media sosial menggunakan teknologi berbasis web yang mengubah komunikasi menjadi dialog interaktif (Basri, 2017). Media sosial merupakan *platform* digital yang memudahkan penggunanya untuk berkomunikasi, berpartisipasi atau membagikan konten berupa tulisan, foto maupun video dan dapat dilakukan dengan cara posting pada sebuah blog, *tweet*, membagikan foto ataupun video pada Instagram, membagikan video pada youtube yang dapat dilihat secara langsung oleh orang lain.

4. Twitter

Twitter adalah Jejaring Sosial yang membatasi penggunaannya untuk mengirim sebuah *tweet* dengan batas 140 Kata, Tidak lebih. Twitter dengan Facebook mempunyai kesamaan dan perbedaan. Kesamaannya ialah Twitter dan Facebook sama-sama layanan Jejaring Sosial yang berguna untuk saling menghubungkan antara pengguna satu dengan pengguna lainnya (Basri, 2017). Sedangkan menurut (Zukhrufillah, 2018) Twitter adalah layanan jejaring sosial

dan microblog daring yang memungkinkan penggunaanya untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter, yang dikenal dengan sebutan kicauan (tweet). Twitter didirikan pada bulan Maret 2006 oleh Jack Dorsey, dan situs jejaring sosialnya diluncurkan pada bulan Juli. Sejak diluncurkan, Twitter telah menjadi salah satu dari sepuluh situs yang paling sering dikunjungi di Internet, dan dijuluki dengan “pesan singkat dari Internet”. Di Twitter, pengguna tak terdaftar hanya bisa membaca kicauan, sedangkan pengguna terdaftar bisa menulis kicauan melalui antarmuka situs web, pesan singkat (SMS), atau melalui berbagai aplikasi untuk perangkat seluler.

Fitur – fitur yang terdapat pada Twitter adalah sebagai berikut :

- a. *Tweet*, Fitur *Tweet* ini ialah fitur utama di Twitter. *Tweet* ialah kicauan ,yaitu untuk mengirim dan melihat kicauan setiap pengguna twitter.
- b. *Following*, Fitur *Following* adalah fitur untuk mengikuti teman atau kerabat di Twitter, Fitur ini merupakan salah satu Fitur utama di twitter.
- c. *Followers*, Fitur yang satu ini adalah fitur untuk melihat siapa yang mengikuti anda di twitter.
- d. *Bio*, Fitur *Bio* adalah fitur yang digunakan untuk mengetahui Pesan Akun Twitter anda yang terdapat di Profil.
- e. *Profil*, merupakan fitur utama dari Twitter yang berfungsi untuk melihat Avatar Twitter, Bio twitter dan lainnya.

5. Klasifikasi

Klasifikasi merupakan proses yang bertujuan untuk menentukan suatu obyek kedalam suatu kelas atau kategori yang sudah ditentukan sebelumnya. Klasifikasi data ataupun dokumen juga dapat dimulai dari membangun aturan klasifikasi tertentu menggunakan data training yang sering disebut sebagai tahapan pembelajaran dan pengujian digunakan sebagai data testing (Parapat et al., 2018). Klasifikasi juga dapat diartikan sebagai proses untuk menemukan sekumpulan model atau fungsi yang mendeskripsikan dan membedakan kelas data. Klasifikasi mempunyai tujuan yaitu memprediksi kelas dari suatu objek yang kelasnya tidak diketahui. Klasifikasi yang sering digunakan adalah seperti model *decision tress*, *naïve baiyes*, *Support Vector Machine*, *K-nearest-neighbourhood classifier*, *neural network*.

6. Pre-Processing

Preprocessing adalah tahapan dimana data atau informasi diatur dari data tekstual semi terstruktur dan tidak terstruktur menjadi sebuah format yang lebih terstruktur yang nantinya akan mempermudah pemakaian. Tujuan dari *preprocessing* adalah melakukan seleksi komentar yang mengandung kata atau hashtag (#), karena Twitter terdapat fungsi retweet, yaitu memberikan komentar terhadap *tweet* komentar seseorang, karena komentar tweet akan mengganggu dalam proses analisis Sentimen tweet. Jadi dalam

preprocessing ini komentar tweet di hapus (Buntoro, 2017). Tahapan–
tahap *preprocessing* adalah sebagai berikut :

- a. *Cleansing* merupakan tahapan membersihkan dokumen dari komponen-komponen yang tidak memiliki hubungan dengan informasi yang ada pada dokumen, seperti URL, *hashtag* (#), *username* (@username), email, *emoticon* (:@, :* , :D), tanda baca seperti koma (,), titik (.) dan juga tanda baca lainnya.
- b. *Case folding* dapat mengubah bentuk kata menjadi bentuk dasarnya agar sebuah karakter dapat seragam (*lower case*) atau dapat disebut juga penyeragaman bentuk huruf.
- c. *Tokenizing* dapat memisahkan data teks menjadi beberapa token. *Tokenizing* secara garis besar memecah sekumpulan karakter dalam suatu teks menjadi suatu kata, bagaimana membedakan karakter-karakter tertentu yang dapat diperlakukan sebagai pemisah kata atau bukan.
- d. Normalisasi dapat mengubah kata yang tidak baku atau sesuai ejaan, pada tahap ini dilakukan dengan menggunakan database kamus kata bahasa baku dan tidak baku yang dibuat sendiri berdasarkan dari data komentar yang digunakan (Putri et al., 2022).
- e. *Filtering* merupakan proses membuang kata yang tidak penting dari proses *tokenizing* sebelumnya. Kata akan di *stopword* berdasarkan jenis kata yang akan di *stopword* seperti

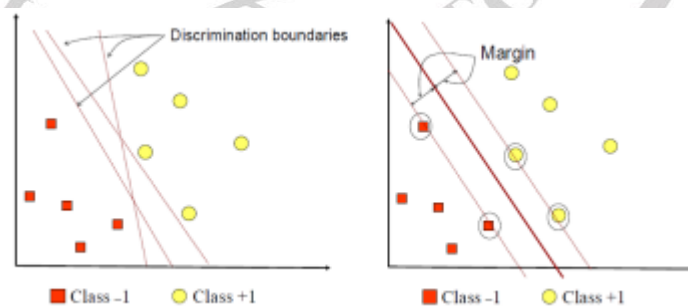
kata keterangan, kata seru, kata ganti, kata depan dan kata hubung. Selain itu juga dilakukan penghapusan tanda baca dengan menggunakan fungsi regexp yang bisa menghilangkan tanda baca seperti berikut “\.,|:|;|!|\?|\(|\)|\||\+|'|"|"’|'“|”|\’|...|-|-|—|\\$|&|*|>|<” (Nurhafida & Sembiring, 2021).

- f. *Stemming* yaitu melakukan proses mencari kata dasar dari setiap kata hasil proses *filtering* sebelumnya. *Stemming* melakukan proses pengembalian berbagai bentukan kata menjadi bentukan kata dasar dengan menghilangkan imbuhan (Putri et al., 2022).

7. *Support Vector Machine (SVM)*

Support Vector Machine merupakan salah satu metode klasifikasi dengan menggunakan metode *machine learning (supervised learning)* yang memprediksi kelas berdasarkan pola dari hasil proses training yang diciptakan oleh Vladimir Vapnik. Klasifikasi dilakukan dengan garis pembatas (*hyperlane*) yang memisahkan antara kelas opini positif dan opini negatif. Secara intuitif, suatu garis pembatas yang baik adalah yang memiliki jarak terbesar ke titik data pelatihan terdekat dari setiap kelas, karena pada umumnya semakin besar margin, semakin rendah error generalisasi dari pemilah. Margin adalah jarak dari suatu titik vektor di suatu kelas terhadap *hyperplane* (Santoso et al., 2017).

Hal yang mendasari untuk memahami klasifikasi dengan *Support Vector Machine* adalah mencari garis (*hyperplane*) yang optimal. Hal ini berfungsi untuk memisahkan dua kelas data yang berbeda yaitu data positif (+1) dan data negatif (-1). Pada Gambar 2.1 untuk data positif (+1) disimbolkan dengan warna kuning dan data negatif (-1) disimbolkan dengan warna merah. Secara umum, proses *Support Vector Machine* diilustrasikan pada gambar 2.1. Pada gambar 2.1 mengilustrasikan tentang beberapa kemungkinan garis pemisah (*discrimination boundaries*) pada SVM untuk melakukan set data. Sedangkan grafik sebelah kanan pada Gambar 2.1 mengilustrasikan *discrimination boundaries* dengan margin yang maksimum. Margin atau disebut juga batas pemisah adalah jarak antara dua kelas data terdekat pada bidang *hyperplane*. *Hyperplane* dengan margin yang maksimum adalah *hyperplane* terbaik yang akan memberikan generalisasi untuk mendapatkan hasil klasifikasi yang lebih baik (Handayani et al., 2020).



Gambar 2. 1 proses SVM dalam menemukan hyperline sumber (Handayani et al., 2020).

Data yang tersedia dinotasikan sebagai $x \in \mathbb{R}^d$ sedangkan label dinotasikan $y_i \in \{-1, +1\}$ untuk $i = 1, 2, \dots, l$, yang mana l adalah banyaknya data. Diasumsikan kedua kelas -1 dan $+1$ dapat terpisah secara sempurna oleh *hyperplane* berdimensi d , yang didefinisikan :

$$w \cdot x + b = 0 \dots\dots\dots(1)$$

Pola x_i yang termasuk kelas -1 (sampel negatif) dapat dirumuskan sebagai pola memenuhi pertidaksamaan:

$$w \cdot x + b \leq -1 \text{ untuk } y_i = -1 \dots\dots\dots(2)$$

Sedangkan pola yang termasuk kelas $+1$ (sampel positif) dirumuskan :

$$w \cdot x + b \leq +1 \text{ untuk } y_i = +1 \dots\dots\dots(3)$$

Keterangan :

X_i = data ke $-i$

W = Nilai bobot Support Vector yang tegak lurus dengan *hyperline*

b = Nilai bias

Y_i = kelas data ke $-i$

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara *hyperplane* dan titik terdekatnya, yaitu $1/\|w\|$. Dapat dirumuskan sebagai masalah pemrograman kuadratik (QP), yaitu mencari titik minimum persamaan dengan batasan persamaan:

$$\min \tau(w) = \frac{1}{2} \|w\|^2 \dots\dots\dots(4)$$

$$y_i(x_i \cdot w + b) - 1 \geq 0 \dots\dots\dots(5)$$

8. Term Frequency – Inverse Document Frequency (TF-IDF)

Term Frequency – Inverse Document Frequency (TF-IDF)

merupakan suatu proses pembobotan setiap kata agar dapat mengoptimalkan kemampuan analisis sentimen pada proses *text mining* (Gifari et al., 2022). Berikut ini adalah beberapa metode pembobotan menggunakan TF-IDF:

- a. *Term Frequency* merupakan perhitungan jumlah kemunculan tiap term pada keseluruhan dokumen (data *tweet*). Rumus *Term Frequency* terdapat pada persamaan 6 berikut ini:

$$TF = \frac{\text{Frekuensi kemunculan term pada dokumen}}{\text{Total Term pada dokumen}} \dots\dots\dots(6)$$

Keterangan:

TF = Frekuensi kemunculan kata pada satu dokumen teks.

- b. *Invers Document Frequency (IDF)* merupakan frekuensi kemunculan term pada seluruh dokumen teks. *Term* yang jarang muncul pada keseluruhan dokumen teks memiliki nilai *Invers Document Frequency* lebih besar dibandingkan dengan *term* yang sering muncul. Dengan kata lain *Document Frequency* merupakan jumlah *tweet* yang mengandung setiap *term*, sedangkan *Invers Document Frequency* merupakan *inverse* dari nilai DF. Rumus *Invers Document Frequency (IDF)* terdapat pada persamaan 7. sebagai berikut:

$$IDF = \log \frac{N}{DFt} \dots\dots\dots(7)$$

Keterangan :

N = Jumlah Dokumen

DFt = Nilai TF

c. *Term Frequency – Inverse Document Frequency* (TF-IDF) merupakan perhitungan yang digunakan untuk proses pembobotan. Setelah TF dan IDF didapatkan dengan persamaan yang terdapat pada nomor persamaan 6 dan 7 maka selanjutnya mencari TF-IDF. TF-IDF didapat dengan cara mengkalikan nilai TF dan IDF untuk memperoleh nilai bobot tiap *tweet*. Rumus TF-IDF terdapat pada persamaan 8 berikut ini :

$$W_t = TF \times IDF \dots\dots\dots(8)$$

9. *Confusion Matrix*

Confusion Matrix merupakan teknik yang digunakan untuk mengevaluasi klasifikasi model untuk memperkirakan objek yang benar atau salah. Sebuah matriks dari prediksi akan dibandingkan dengan kelas asli yang berisi informasi aktual dan prediksi nilai klasifikasi (Pravina et al., 2019). Tabel 2.2 menunjukkan *confusion matrix* dengan dataset terdiri dari 2 kelas yaitu dimana kelas yang satu dianggap sebagai positif dan lainnya yaitu negatif yang digunakan untuk membantu dalam perhitungan sistem evaluasi (Ramadhan & L, 2018). Pengujian menggunakan *confusion matrix* empat kondisi yaitu : *True Positive* (TP), *True negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Tabel 2.3 menunjukkan tabel *confusion matrix* dengan dataset yang terdiri dari kelas yaitu positif, negatif, dan netral.

Tabel 2. 2 *Confusion Matrix* Dua Kelas

	Kelas <i>Positive</i> Prediksi	Kelas <i>Negative</i> Prediksi
<i>Positive</i>	<i>True Positive</i> (TP)	<i>False Negative</i> (FN)
<i>Negative</i>	<i>False Positive</i> (FP)	<i>True Negative</i> (TN)

Tabel 2. 3 *Confusion Matrix* Tiga Kelas

	Kelas <i>Positive</i> Prediksi	Kelas <i>Negative</i> Prediksi	Kelas <i>Neutral</i> prediksi
<i>Positive</i>	<i>True Positive</i> (TP)	<i>False Negative</i> (FN)	<i>False Neutral</i> (FNt)
<i>Negatif</i>	<i>False Positive</i> (FP)	<i>True Negative</i> (TN)	<i>False Neutral</i> (FNt)
<i>Neutral</i>	<i>False Positive</i> (FP)	<i>False Negative</i> (FN)	<i>True Neutral</i> (TNt)

True Positive (TP), merupakan nilai kelas yang diprediksi bernilai *positive* dan faktanya *positive*. *False Positive* (FP), merupakan nilai kelas yang diprediksi bernilai *positive* dan faktanya *negative* atau *neutral*. *True Negative* (TN), merupakan nilai kelas yang diprediksi bernilai *negative* dan faktanya *negative*. *False Negative* (FN), merupakan nilai kelas yang diprediksi bernilai *negative* dan

faktanya *positive* atau *neutral*. *True Neutral (TNt)* merupakan nilai kelas yang diprediksi bernilai *neutral* dan faktanya *neutral*. *False neutral (FNt)* merupakan nilai yang kelas yang diprediksi bernilai *neutral* dan faktanya *positive* dan *negative*.

Evaluasi menggunakan *confusion matrix* menghasilkan nilai *accuracy*, *recall*, *precision* dan *f1-score*. Nilai *accuracy* merupakan persentase dari suatu kelas terprediksi dengan benar oleh model yang sudah dibuat. (Pintoko & L., 2018). Nilai *accuracy* dapat diperoleh dengan persamaan (9).

$$\text{Akurasi} = \frac{(TP+TN+TNt)}{(TN + FNt + FP + FN + TNt + FP + FN + FNt + TP)} \times 100\% \dots(9)$$

Nilai *Recall* adalah persentase sebuah program memprediksi sebuah data ke bukan kelas aktualnya, *recall* biasa disebut juga dengan sensitifitas. Perhitungan *recall* dapat dilakukan dengan menggunakan persamaan (10) (Huang et al., 2005).

$$\text{Recall} = \frac{(TP)}{(TP+FN)} \times 100\% \dots\dots\dots(10)$$

Nilai *Precision* adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Perhitungan *precision* dapat dilakukan dengan menggunakan persamaan (11) (Huang et al., 2005).

$$\text{Precision} = \frac{(TP)}{(TP+FP)} \times 100\% \dots\dots\dots(11)$$

Nilai *F1-Score* merupakan salah satu perhitungan evaluasi dalam informasi yang mengkombinasikan *recall* dan *precision*

(Huang et al., 2005). Nilai *F1-Score* menunjukkan perbandingan rata-rata *recall* dan *precision* yang dibobotkan. perhitungan *F1-Score* dapat dilakukan dengan menggunakan persamaan (12).

$$F1-Score = \frac{2 \cdot recall \cdot precision}{recall + precision} \times 100\% \dots\dots\dots(12)$$

10. Python

Python merupakan bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. *Python* diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif. *Python* didukung oleh komunitas yang besar. *Python* mendukung multi paradigma pemrograman, utamanya namun tidak dibatasi pada pemrograman berorientasi objek, pemrograman imperatif, dan pemrograman fungsional. Salah satu fitur yang tersedia pada *python* adalah sebagai bahasa pemrograman dinamis yang dilengkapi dengan manajemen memori otomatis. Seperti halnya pada bahasa pemrograman dinamis lainnya, *python* umumnya digunakan sebagai bahasa *script* meski pada praktiknya penggunaan bahasa ini lebih luas mencakup konteks pemanfaatan yang umumnya tidak dilakukan dengan menggunakan bahasa *script*. *Python* dapat digunakan untuk berbagai keperluan pengembangan perangkat lunak dan dapat berjalan di berbagai *platform* sistem operasi (Syahrudin & Kurniawan, 2018)