

PENERAPAN METODE SUPPORT VECTOR MACHINE DAN NAÏVE BAYES PADA KLASIFIKASI SENTIMEN REVIEW APLIKASI WHATSAPP

Joshua. A. Pratama

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara,
Jln. Letjen S. Parman No.1, Jakarta, 1140, Indonesia
e-mail: joshua.535200028@stu.untar.ac.id

ABSTRAK

Dengan semakin populernya aplikasi berbasis pengguna seperti WhatsApp, terdapat kebutuhan yang semakin besar untuk menganalisis sentimen pengguna secara efektif. Dalam penelitian ini, kami menerapkan metode *Support Vector Machine* (SVM) untuk menganalisis sentiment ulasan pengguna WhatsApp di Google Play Store. Data validasi diolah dan dinormalisasi sebelum diterapkan pada model SVM. Representasi teks dengan teknik TF-IDF digunakan untuk menghasilkan fitur-fitur yang relevan. Eksperimen dilakukan dengan membagi data menjadi set pelatihan dan set pengujian. Hasil evaluasi model menunjukkan efektivitas SVM dalam klasifikasi sentimen, dan akurasi serta metrik evaluasi lainnya memberikan hasil yang memuaskan. Penerapan SVM dalam analisis sentimen ini berkontribusi pada pemahaman mendalam tentang reaksi pengguna terhadap aplikasi WhatsApp dan memberikan wawasan berharga bagi pengembang dan pemangku kepentingan. Eksperimen dilakukan untuk membandingkan performa kedua algoritma, meliputi akurasi, presisi, *recall*, dan skor F1. Hasil eksperimen menunjukkan bahwa kedua algoritma dapat memberikan klasifikasi emosi yang baik, namun kinerjanya dapat bervariasi tergantung pada jenis dataset dan parameter yang digunakan. Studi ini berkontribusi untuk memahami sentimen pengguna terhadap WhatsApp dan memberikan wawasan kepada pengembang untuk meningkatkan kualitas aplikasi mereka. Selain itu, hasil eksperimen memberikan wawasan yang lebih mendalam tentang kelebihan dan kekurangan Naive Bayes dan SVM dalam konteks klasifikasi sentimen dalam tinjauan aplikasi.

Kata kunci: Klasifikasi Sentimen, *Naive Bayes*, *Support Vector Machine*, *Google Play Store*, WhatsApp.

ABSTRACT

With the growing popularity of user-driven applications like WhatsApp, there is a growing need to analyze user sentiment effectively. In this research, we apply the Support Vector Machine (SVM) method to analyze the sentiment of WhatsApp user reviews on the Google Play Store. Validation data is processed and normalized before being applied to the SVM model. Text representation with the TF-IDF technique is used to produce relevant features. Experiments are carried out by dividing the data into a training set and a testing set. The model evaluation results show the effectiveness of SVM in sentiment classification, and accuracy and other evaluation metrics provide satisfactory results. This application of SVM in sentiment analysis contributes to a deep understanding of user reactions to the WhatsApp application and provides valuable insights for developers and stakeholders. Experiments were carried out to compare the performance of the two algorithms, including accuracy, precision, recall and F1 score. Experimental results show that both algorithms can provide good emotion classification, but their performance may vary depending on the type of dataset and parameters used. This study contributes to understanding user sentiment towards WhatsApp and provides insights to developers to improve the quality of their applications. In addition, the experimental results provide deeper insights into the advantages and disadvantages of Naive Bayes and SVM in the context of sentiment classification in application reviews.

Keywords: *Sentiment Classification, Naive Bayes, Support Vector Machine, Google Play Store, WhatsApp.*

1 PENDAHULUAN

Dengan pesatnya pertumbuhan teknologi dan penggunaan aplikasi pesan instan, WhatsApp telah menjadi salah satu aplikasi terpopuler di seluruh dunia. Semakin banyak pengguna berarti semakin beragam pendapat dan perasaan tentang pengalaman pengguna. Analisis sentimen atau

sentiment analysis dalam bahasa Indonesia merupakan sebuah Teknik atau cara yang digunakan untuk mengetahui bagaimana emosi diungkapkan melalui sebuah teks dan bagaimana emosi tersebut dapat digolongkan ke dalam emosi positif atau emosi negatif (Tetsuya Nasukawa, 2003). Oleh karena itu, diperlukan analisis sentimen untuk memahami secara menyeluruh bagaimana pengguna merespons aplikasi tersebut. Dalam konteks ini, metode *Support Vector Machine* (SVM) dan metode naïve bayes merupakan metode yang dipilih untuk analisis sentimen teks. SVM adalah metode *machine learning* yang bekerja atas prinsip *Structural Risk Minimization* (SRM) dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua buah *class* pada *input space* (Cortes, 1995). Naive Bayes adalah klasifikasi linear yang sangat efisien karena merupakan metode pengklasifikasian yang terkenal dengan tingkat keakurasian yang baik. Metode Naive Bayes juga merupakan pengklasifikasian peluang sederhana dengan asumsi antar variabel penjelas saling bebas (*independent*) karena perhitungan cepat dan sederhana. Metode Naive Bayes dapat digunakan untuk membuat aplikasi klasifikasi dalam beberapa kasus (Barokah, 2019). Tujuan dari penelitian ini adalah menerapkan SVM pada review pengguna WhatsApp di Google Play Store untuk mengidentifikasi sentimen positif dan negatif. Oleh karena itu, penelitian ini berkontribusi pada pemahaman mendalam tentang persepsi dan preferensi pengguna terhadap WhatsApp dan memberikan informasi berharga bagi pengembang aplikasi serta penelitian lanjutan di bidang analisis sentimen.

2 TINJAUAN LITERATUR

Berikut merupakan daftar studi literatur terkait dengan penelitian penerapan metode *support vector machine* dan naïve bayes dalam klasifikasi *sentiment review* aplikasi WhatsApp:

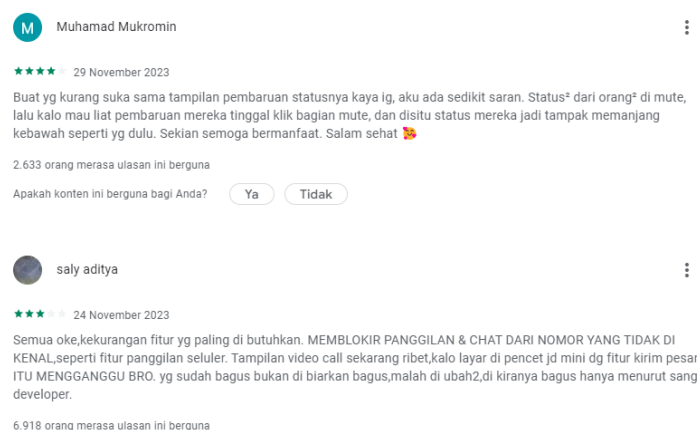
- [1] Analisis Sentimen ulasan pelanggan hotel di trip advisor oleh Wijaya, Willyanto, Dyah Erny Herwindiati, dan Novario Jaya Perdana (Wijaya, 2022).
- [2] Analisis Sentimen Komentar Netizen Twitter Terhadap Kesehatan Mental Masyarakat Indonesia oleh Kenny Yan, Desi Arisandi, dan Tony (Yan Kenny, 2022).
- [3] Sistem Analisis Sentimen Ulasan produk menggunakan metode Naïve Bayes oleh Gunawan, B., Sastypratiwi, H., & Pratama, (Gunawan, 2018).
- [4] Analisis Sentimen Pada Jasa Ojek Online Menggunakan Metode Naive Bayes oleh Nugroho, D. G (Nugroho, 2016)
- [5] Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes oleh Sipayung, E. M., Maharani, H., & Zefanya, I. (Sipayung, 2016)
- [6] Analisis sentimen terhadap opini masyarakat tentang vaksin covid-19 menggunakan algoritma naïve bayes classifier oleh Yulita, W (Yulita, 2021).
- [7] Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berksala Besar Menggunakan Algoritma Support Vector Machine oleh Tuhuteru, H. (Tuhuteru, 2020).
- [8] Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine oleh Styawati, S., Hendrastuty, N., & Isnain, A. R. (Styawati, 2021)
- [9] Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine oleh Novantirani, A., Sabariah, M. K., & Effendy, V (Novantirani, 2015).
- [10] Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine oleh Irfani, F. F., Triyanto, M., & Hartanto, A. D. (Irfani, 2020).
- [11] Analisis sentimen tingkat kepuasan pengguna penyedia layanan telekomunikasi seluler indonesia pada twitter dengan metode Support Vector Machine dan Lexicon Based Features oleh Rofiqoh, U., Perdana, R. S., & Fauzi, M. A. (Rofiqoh, 2017).
- [12] Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma oleh Arsi, P., & Waluyo, R. (Arsi, 2021).
- [13] Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus (Apriyani, 2020).

- [14]. Perbandingan Metode Naïve Bayes, Support Vector Machine Dan Decision Tree Dalam Klasifikasi Konsumsi Obat oleh Sari, C. A. (Sari, 2022)
- [15]. Komparasi Naïve Bayes, Support Vector Machine Dan K-Nearest Neighbor Untuk Mengetahui Akurasi Tertinggi Pada Prediksi Kelancaran Pembayaran TV Kabel oleh Lasulika, M. E. (Lasulika, 2019).

3 METODE PENELITIAN

3.1 Data Ulasan

Pengambilan ulasan pengguna di Google Play Store akan menggunakan bahasa pemrograman *Python* dengan *code editor* Google Colaboratory dengan *library google-play-scraper*. Data yang akan diambil berjumlah 1000 ulasan dengan bahasa Indonesia dan mengambil komentar yang paling masuk akal atau *most relevant*. Untuk data yang akan diambil dapat dilihat di Gambar 1.



Gambar 1 Data ulasan yang akan diambil

3.2 Preprocessing

Kemudian berlanjut dengan proses pemotongan kolom dimana merupakan tahap pertama di *preprocessing*. Data yang sudah diproses menghasilkan hanya kolom komentar dan *score* yang diambil, dengan proses tambahan pelabelan melalui nilai *score* dengan syarat jika *score* yang didapat dari komentar bernilai 4 dan 5 maka akan dilabeli dengan hasil positif namun jika hasil *score* dibawah 3 maka akan dilabeli dengan hasil negatif lalu jika komentar bernilai 3 maka akan dilabeli dengan *none* dan nantinya komentar dengan *score none* akan dihapus dari data yang sudah diambil, hasil kolom yang sudah dipangkas dan dilabeli dapat dilihat di Gambar 2.

	content	score	Label
0	aplikasih yg sangat berguna bgt saat ini. hany...	5	positif
1	aplikasih yg sangat berguna bgt saat ini. hany...	5	positif
2	Untuk fitur lainnya udh bagus, cuman fitur men...	4	positif
3	setiap up video di sw kualitas menjadi buruk, ...	2	negatif
4	Kenapa di chat skrg muncul riwayat panggilan,m...	5	positif
5	Bagus semua juga bagus, tapi kalau bisa tambah...	5	positif
6	Saran aja sih kalo bisa durasi video di story ...	5	positif
7	Terus lanjutkan & tingkatkan kualitasnya d Wha...	5	positif
8	Kalau bisa dipisahkan antara chat personal dan...	4	positif
9	Tolong privasi. Masa pesan yang udah berbulan-...	1	negatif
11	Sama aja dgn sebelumnya, ku pikir akan ada yg ...	2	negatif
12	Aplikasi ini sangatlah bagus cuma kekurangan n...	5	positif
13	Untuk update WA sekarang ok lah, bisa kirim Vi...	5	positif

Gambar 2 Kolom table yang sudah dipangkas dan diberi label

Kemudian tahap selanjutnya yaitu tahap *cleaning data* atau tahap *preprocessing* yaitu dengan tahap awal berupa *case folding* dengan menghapus *link* yang terdapat di komentar, proses penghapusan tanda baca, penghapusan karakter angka, merubah komentar menjadi karakter huruf kecil. dan menghapus semua karakter *emoticon* yang ada di kolom komentar. Untuk hasil *case folding* dapat dilihat di Gambar 3.

index	content	score	Label
0	aplikasih yg sangat berguna bgt saat ini. hanya saja sllu ada fitur terbaru sehingga mau gk mau hrs di perbarui sering minta d perbarui tp gmnr slp video call sllu jelek pdhl signal ok ok aja	5	positif
1	aplikasih yg sangat berguna bgt saat ini. hanya saja sllu ada fitur terbaru sehingga mau gk mau hrs di perbarui sering minta d perbarui tp gmnr slp video call sllu jelek pdhl signal ok ok aja	5	positif
2	Untuk fitur lainnya udi bagus, cuman fitur mengatur ke kualitas yg lebih baik saat meng-upload media GK ad flumya. Sama Saat meng-upload media kualitasnya buram dan settingannya gk bisa diubah So, i hope WhatsApp team read this comment and fix it immediately karna walaupun pake cara apapun tetep hasilnya buram	4	positif
3	setiap up video di sw kualitas menjadi buruk, resolusi nya blur, bener bener menjengkelkan. Kalo Untuk upgrade bagian flumya emang menjadi lengkap, namun keluhan nya yaitu swstory whatsapp tolong kpd developer sekiranya ga penting gausah keseringan di upgrade, ini menjadi sangat buruk, bukan hanya minus disitu saja, kami user andro merasa keberatan kalo disuruh download terus terusan, apalagi masa kadaluarsa nya sebentar.	2	negatif
4	Kenapa di chat skrg muncul riwayat panggilan,menuhi menuhi chat saja,sajamenambah kerjaan menghapus manual,hadeuhh... kan sdh ada riwayat panggilan,kenapa di masukan ke chat lagi...lama lama dimasukan ke status...ada ada aja fitur ga berguna.	5	positif
5	Bagus semua juga bagus, tapi kalau bisa tambahkan fitur filter di vc nva dan tambahn laou biar oas buat status noak ribet noedit	5	positif

Gambar 3 Hasil tahap *case folding*

Kemudian ke tahap selanjutnya yaitu tahap penghapusan *stopwords* atau kata penghubung yang ada di komentar. Tahap ini menghapus kalimat penghubung seperti contoh untuk kalimat di sana dipangkas menjadi sana, untuk gambaran tahap proses *stopword* dapat dilihat di **Gambar 4**. Lalu lanjut ke tahap berikutnya yaitu tahap *tokenizing* atau tahap pemisahan antar kalimat, tahap ini bertujuan untuk memisahkan antar kalimat agar nantinya bisa dihitung bobot TF-IDF dari setiap kalimat tersebut, Untuk contoh tahap *tokenizing* dapat dilihat di Gambar 5.

index	content	score	Label	text_clean	text_Stopword
0	aplikasih yg sangat berguna bgt saat ini. hanya saja sllu ada fitur terbaru sehingga mau gk mau hrs di perbarui sering minta d perbarui tp gmnr slp video call sllu jelek pdhl signal ok ok aja	5	positif	aplikasih yg sangat berguna bgt saat ini hanya saja sllu ada fitur terbaru sehingga mau gk mau hrs di perbarui sering minta d perbarui tp gmnr slp video call sllu jelek pdhl signal ok ok aja	aplikasih yg berguna bgt sllu fitur terbaru gk hrs perbarui d perbarui tp gmnr slp video call sllu jelek pdhl signal ok ok aja
1	Untuk fitur lainnya udi bagus, cuman fitur mengatur ke kualitas yg lebih baik saat meng-upload media GK ad flumya. Sama Saat meng-upload media kualitasnya buram dan settingannya gk bisa diubah So, i hope WhatsApp team read this comment and fix it immediately karna walaupun pake cara apapun tetep hasilnya buram	4	positif	untuk fitur lainnya udi bagus cuman fitur mengatur ke kualitas yg lebih baik saat mengupload media gk ad flumya sama saat mengupload media kualitasnya buram dan settingannya gk bisa diubah so i hope whatsapp team read this comment and fix it immediately karna walaupun pake cara apapun tetep hasilnya buram	fitur udi bagus cuman fitur mengatur kualitas yg mengupload media gk ad flumya mengupload media kualitasnya buram settingannya gk diubah so i hope whatsapp team read this comment and fix it immediately karna apapun tetep hasilnya buram
2	setiap up video di sw kualitas menjadi buruk, resolusi nya blur, bener bener menjengkelkan. Kalo Untuk upgrade bagian flumya emang menjadi lengkap, namun keluhan nya yaitu swstory whatsapp tolong kpd developer sekiranya ga penting gausah keseringan di upgrade, ini menjadi sangat buruk, bukan hanya minus disitu saja, kami user andro merasa keberatan kalo disuruh download terus terusan, apalagi masa kadaluarsa nya sebentar.	2	negatif	setiap up video di sw kualitas menjadi buruk resolusi nya blur bener bener menjengkelkan kalo untuk upgrade bagian flumya emang menjadi lengkap namun keluhan nya yaitu swstory whatsapp tolong kpd developer sekiranya ga penting gausah keseringan di upgrade ini menjadi sangat buruk bukan hanya minus disitu saja kami user andro merasa keberatan kalo disuruh download terus terusan apalagi masa kadaluarsa nya sebentar	up video sw kualitas buruk resolusi nya blur bener bener menjengkelkan kalo upgrade flumya emang lengkap keluhan nya yaitu swstory whatsapp tolong kpd developer ga gausah keseringan upgrade buruk minus disitu user andro keberatan kalo disuruh download terus kadaluarsa nya sebentar
3	Kenapa di chat skrg muncul riwayat panggilan,menuhi menuhi chat saja,sajamenambah kerjaan menghapus manual,hadeuhh... kan sdh ada riwayat panggilan,kenapa di masukan ke chat lagi...lama lama dimasukan ke status...ada ada aja fitur ga berguna.	5	positif	kenapa di chat skrg muncul riwayat panggilanmenuhi menuhi chat sajamenambah kerjaan menghapus manualhadeuhhkan sdh ada riwayat panggilankenapa di masukan ke chat lagi lama lama dimasukan ke statusada ada aja fitur ga berguna	chat skrg muncul riwayat panggilanmenuhi menuhi chat sajamenambah kerjaan menghapus manualhadeuhhkan sdh ada riwayat panggilankenapa di masukan ke chat lagi lama dimasukan ke statusada ada aja fitur ga berguna

Gambar 4. Tahap proses penghapusan kata *stopword*

index	content	score	Label	text_clean	text_Stopword	text_tokens
0	aplikasih yg sangat berguna bgt saat ini. hanya...	5	positif	aplikasih yg sangat berguna bgt saat ini hanya...	aplikasih yg berguna bgt sllu fitur terbaru gk...	[aplikasih, yg, berguna, bgt, sllu, fitur, ter...
1	Untuk fitur lainnya udi bagus, cuman fitur men...	4	positif	untuk fitur lainnya udi bagus cuman fitur meng...	fitur udi bagus cuman fitur mengatur kualitas ...	[fitur, udi, bagus, cuman, fitur, mengatur, ku...
2	setiap up video di sw kualitas menjadi buruk, ...	2	negatif	setiap up video di sw kualitas menjadi buruk r...	up video sw kualitas buruk resolusi nya blur b...	[up, video, sw, kualitas, buruk, resolusi, nya...
3	Kenapa di chat skrg muncul riwayat panggilan m...	5	positif	kenapa di chat skrg muncul riwayat panggilanme...	chat skrg muncul riwayat panggilanmenuhi menuh...	[chat, skrg, muncul, riwayat, panggilanmenuhi menuh...
4	Kenapa di chat skrg muncul riwayat panggilan m...	5	positif	kenapa di chat skrg muncul riwayat panggilanme...	chat skrg muncul riwayat panggilanmenuhi menuh...	[chat, skrg, muncul, riwayat, panggilanmenuhi...
5	Bagus semua juga bagus, tapi kalau bisa tamba...	5	positif	bagus semua juga bagus tapi kalau bisa tamba...	bagus bagus tambahn fitur filter vc nya tamba...	[bagus, bagus, tambahn, fitur, filter, vc, nya...
6	Saran aja sih kalo bisa durasi video di story ...	5	positif	saran aja sih kalo bisa durasi video di story ...	saran aja sih kalo durasi video story wa nya l...	[saran, aja, sih, kalo, durasi, video, story, ...
7	Terus lanjutkan & tingkatkan kualitasnya d...	5	positif	terus lanjutkan tingkatkan kualitasnya d...	lanjutkan tingkatkan kualitasnya d...	[lanjutkan, tingkatkan, kualitasnya, d...

Gambar 5. Tahap *tokenizing*

Kemudian tahap yang terakhir adalah tahap *stemmer* yaitu mengubah kata komentar yang telah dipisah di tahap *tokenizing* diubah menjadi kata dasar seperti contoh kalimat bertanya di *stem* menjadi tanya. Untuk tahap *stemmer* dapat dilihat di Gambar 6.

	content	score	Label	text_clean	text_stopword	text_tokens
0	aplikasi yg sangat berguna bgt saat ini. hany...	5	positif	aplikasi yg sangat berguna bgt saat ini hanya...	aplikasi yg berguna bgt situ fitur terbaru gk...	[aplikasi, yg, berguna, bgt, situ, fitur, ter...
1	Untuk fitur lainnya udh bagus, cuman fitur men...	4	positif	untuk fitur lainnya udh bagus cuman fitur meng...	fitur udh bagus cuman fitur mengatur kualitas ...	[fitur, udh, bagus, cuman, fitur, mengatur, ku...
2	setiap up video di sw kualitas menjadi buruk. ...	2	negatif	setiap up video di sw kualitas menjadi buruk r...	up video sw kualitas buruk resolusi nya blur b...	[up, video, sw, kualitas, buruk, resolusi, nya...
3	Kenapa di chat skrg muncul riwayat panggilan m...	5	positif	kenapa di chat skrg muncul riwayat panggilanme...	chat skrg muncul riwayat panggilanmenuhi menuh...	[chat, skrg, muncul, riwayat, panggilanmenuhi,...
4	Kenapa di chat skrg muncul riwayat panggilan m...	5	positif	kenapa di chat skrg muncul riwayat panggilanme...	chat skrg muncul riwayat panggilanmenuhi menuh...	[chat, skrg, muncul, riwayat, panggilanmenuhi,...
5	Bagus semua juga bagus, tapi kalau bisa tamba...	5	positif	bagus semua juga bagus tapi kalau bisa tamba...	bagus bagus tambahkan fitur filter vc nya lamba...	[bagus, bagus, tambahkan, fitur, filter, vc, ny...
6	Saran aja sih kalo bisa durasi video di story ...	5	positif	saran aja sih kalo bisa durasi video di story ...	saran aja sih kalo durasi video story wa nya t...	[saran, aja, sih, kalo, durasi, video, story, ...
7	Terus lanjutkan & tingkatkan kualitasnya d...	5	positif	terus lanjutkan tingkatkan kualitasnya d...	lanjutkan tingkatkan kualitasnya d...	[lanjutkan, tingkatkan, kualitasnya, d...

Gambar 6. Tahap *stemmer*

3.3 TF-IDF

Metode TF-IDF merupakan metode yang menghitung bobot setiap kata, dan merupakan metode yang paling umum digunakan untuk pengambilan informasi. Cara ini diyakini efisien, sederhana, dan memberikan hasil yang akurat (Robertson, 2004). Metode untuk memperoleh hasil TF – IDF dapat dihitung dengan cara mengalikan nilai dari TF yang mengandung suatu kalimat dari dokumen tertentu dengan IDF, Untuk mencari hasil dari IDF dapat cara logaritma dari banyaknya dokumen n dibagi dengan dokumen df yang mengandung fitur t^{20} untuk rumus lengkap dari tf idf dapat dilihat di bawah ini. Lalu untuk nilai $TF(t,d)$ dapat dihitung dengan cara jumlah kemunculan kata t dalam dokumen d adalah banyaknya kali kata t muncul dalam dokumen d dibagi dengan total kata dalam dokumen d adalah jumlah total kata yang terdapat dalam dokumen d . dan kemudian hasil akhir dari TF IDF dapat didapatkan dengan cara mengalikan nilai dari TF dengan IDF

$$IDF_t = \frac{n}{df(t)} \quad (1)$$

1. $IDF(t)$ adalah *Inverse Document Frequency* untuk *term t*
2. n adalah total dokumen dalam koleksi
3. $df(t)$ adalah jumlah dokumen yang mengandung *term t*

$$W_t = TF_t \times IDF_t \quad (2)$$

3.4 Support Vector Machine

SVM adalah metode *machine learning* yang bekerja atas prinsip *Structural Risk Minimization* (SRM) dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua buah *class* pada *input space* (Cortes, 1995).

3.5 Naïve Bayes

Naive Bayes juga merupakan klasifikasi probabilistik sederhana yang mengasumsikan bahwa variabel penjelas bersifat independen karena cepat dan mudah dihitung. Metode Naive Bayes dapat digunakan untuk membuat aplikasi klasifikasi dalam beberapa kasus (Barokah, 2019). Untuk rumus dari metode Naive bayes dapat dilihat di rumus no (3) (Ismuhamdan, Ismuhamdan, M. E., Prajoko, P., & Apriandari, W. (2023)).

$$P(A|B) = P(B|A). P(A) P(B) \quad (3)$$

A : Data dengan kelas yang belum diketahui

B : Hipotesis dari data B adalah sebuah kelas spesifik

$P(A|B)$: Probabilitas hipotesis A berdasarkan kondisi B

$P(A)$: Probabilitas hipotesis A (probabilitas awal)

$P(B|A)$: Probabilitas B berdasarkan kondisi pada hipotesis A

$P(B)$: Probabilitas B

3.6 *Confusion Matrix*

Confusion matrix adalah tabel yang menunjukkan klasifikasi jumlah data pengujian yang benar dan salah. Dalam data uji ini terdiri dari empat sel utama yaitu *True Positives* (TP), *True Negatives* (TN), *False Positives* (FP), dan *False Negatives* (FN). Dengan bantuan *confusion matrix*, kita dapat menghitung *matrix* evaluasi seperti akurasi, presisi, recall, dan *F1-score*. *Confusion matrix* ini membantu mengukur performa model klasifikasi dalam mengelola kesalahan prediksi. Untuk analogi dari *confusion matrix* dapat dilihat di gambar 7 dan untuk rumus menghitung nilai akurasi, presisi, recall, dan *F1-score* dapat dilihat dibawah ini.

Contoh perhitungan akurasi :

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (4)$$

Contoh perhitungan nilai Presisi :

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (5)$$

Contoh perhitungan nilai *recall* :

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (6)$$

Contoh perhitungan f_1 score

$$f_1 = \frac{2 \times precision \times recall}{precision + recall} \quad (7)$$

4. *True Positives* (TP)

True Positives adalah jumlah data aktual yang diklasifikasikan dengan benar sebagai positif. Pada penyebut rumus ini, kita tertarik pada jumlah semua data yang jika tidak maka akan bernilai negatif.

5. *True Negatives* (TN)

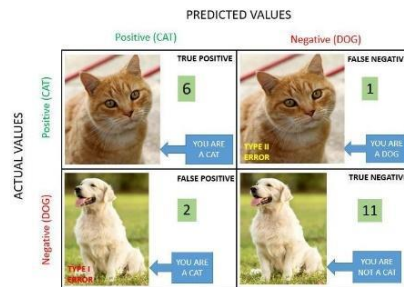
True Negatives adalah jumlah data aktual yang diklasifikasikan dengan benar sebagai negatif. Dalam penyebut rumus ini, kita tertarik pada total semua data yang seharusnya negatif.

6. *False Positives* (FP)

False Positives adalah jumlah data yang seharusnya negatif tetapi salah diklasifikasikan sebagai positif. Dalam penyebut rumus ini, kita tertarik pada total semua data yang seharusnya negatif.

7. *False Negatives* (FN)

False Negatives adalah jumlah data yang seharusnya positif tetapi salah diklasifikasikan sebagai negatif. Dalam penyebut rumus ini, kita tertarik pada total semua data yang seharusnya positif.



Gambar 7 Analogi dari *confusion matrix*

4 HASIL DAN PEMBAHASAN

Dalam hasil pengujiannya telah didapat hasil untuk metode naïve bayes mendapatkan nilai akurasi sebesar 81.71%, lalu nilai *precision* sebesar 82.07%, kemudian untuk nilai *recall* 96.75%, dan untuk nilai *f1 score* sebesar 88.81% dengan *confusion matrix* seperti berikut

$$\begin{bmatrix} 119 & 4 \\ 26 & 15 \end{bmatrix} \quad (8)$$

Kemudian untuk metode SVM mendapatkan nilai akurasi sebesar 83.54%, lalu untuk nilai *precision* sebesar 84.29%, lalu untuk nilai *recall* sebesar 95.93%, dan untuk nilai *f1 score* sebesar 89.73% dengan *confusion matrix* seperti berikut ini

$$\begin{bmatrix} 118 & 5 \\ 22 & 19 \end{bmatrix} \quad (9)$$

Dalam konteks klasifikasi sentimen ulasan pengguna WhatsApp, hasil pengujian menunjukkan bahwa metode *Support Vector Machine* (SVM) mengungguli Naive Bayes. SVM memberikan akurasi lebih tinggi dan keseimbangan yang lebih baik antara mendeteksi ulasan positif dan negatif. Meskipun Naive Bayes memiliki kelebihan dalam mengenali ulasan negatif, perbaikan pada kemampuan mendeteksi sentimen positif perlu diperhatikan. Oleh karena itu, implementasi SVM dapat dianggap sebagai pilihan yang lebih baik untuk tujuan klasifikasi sentimen ulasan WhatsApp, dengan potensi perbaikan pada Naive Bayes untuk meningkatkan performa klasifikasi sentimen positif.

5 KESIMPULAN

Dalam konteks klasifikasi sentimen ulasan pengguna WhatsApp, SVM menunjukkan performa yang lebih baik daripada Naive Bayes. SVM memiliki akurasi yang lebih tinggi dan mampu memberikan keseimbangan yang lebih baik antara mendeteksi ulasan positif dan negatif. Meskipun Naive Bayes memiliki kelebihan dalam mengenali ulasan negatif, perbaikan pada mendeteksi sentimen positif perlu diperhatikan. Oleh karena itu, implementasi SVM dapat menjadi pilihan lebih baik untuk tujuan ini, sementara perbaikan pada Naive Bayes dapat dipertimbangkan untuk meningkatkan performa klasifikasi sentimen positif.

DAFTAR PUSAKA

- [1] Tetsuya Nasukawa, J. Y. (2003). Sentiment analysis: capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on Knowledge capture*, 70-77.
- [2] Cortes, .. V. (1995). Support Vector Machine. *Machine Learning*, 273-297.
- [3] Barokah, K. (2019). Metode Naive Bayes dari teorema Bayes adalah. 10.

- [4] Wijaya, W. D. (2022). Penerapan metode *Support Vector Machine* untuk analisis sentimen pada ulasan pelanggan hotel di *tripadvisor*. *jurnal Ilmu Komputer dan Sistem Informasi*, 10.2
- [5] Yan Kenny, D. A. (2022). Analisis sentimen komentar netizen twitter terhadap kesehatan mental masyarakat Indonesia. *Jurnal Ilmu Komputer Dan Sistem Informasi*, 10.1.
- [6] Gunawan, B., Sastypratiwi, H., & Pratama, E. E. (2018). (Gunawan, 2018). *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 4(2), 113-118.
- [7] Nugroho, D. G. (2016). analisis sentimen pada jasa ojek online menggunakan metode *naive bayes*. *Prosiding seminar sains nasional dan teknologi*, Vol. 1, no. 1.
- [8] Sipayung, E. M., Maharani, H., & Zefanya, I. (2016). Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode *Naive Bayes Classifier*. *Jurnal Sistem Informasi (JSI)*, 8(1), 958-965.
- [9] Yulita, W. (2021). Analisis sentimen terhadap opini masyarakat tentang vaksin covid-19 menggunakan algoritma *naïve bayes classifier*. *jurnal data mining dan sistem informasi*, 1-9.
- [10] Tuhuteru, H. (2020). Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berksala Besar Menggunakan Algoritma *Support Vector Machine*. *Journal Information System Development (ISD)*, 5(2)
- [11] Styawati, S. H. (2021). analisis sentimen masyarakat terhadap program kartu prakerja pada twitter dengan metode *support vector machine*. *jurnal informatika: jurnal pengembangan it*, 150-155.
- [12] Novantirani, A. S. (2015). Analisis sentimen pada twitter untuk mengenai perubahan transportasi umum darat dalam kota dengan metode *support vector machine*. *eProceedings of engineering*, 2(1).
- [13] Irfani, F. F. (2020). analisis sentimen review aplikasi ruang guru menggunakan algoritma *support vector machine*. *JBMI*, 258.
- [14] Rofiqoh, U. P. (2017). analisis sentimen tingkat kepuasan pengguna penyedia layanan telekomunikasi seluler indonesia pada twitter dengan metode *support vector machine* dan *lexicon based features*. *jurnal pengembangan teknologi informasi dan ilmu komputer* , 1725-1732.
- [15] Arsi, P., & Waluyo, R. (2021). Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma *Support Vector Machine (SVM)*. *J. Teknol. Inf. dan Ilmu Komput*, 8(1), 147.
- [16] Apriyani, H. &. (2020). perbandingan metode *naive bayes* dab *support vector machine*. *journal of information technology ampera*, 133-143.
- [17] Sari, C. A. (2022). Perbandingan Metode *Naïve Bayes*, *Support Vector Machine* Dan *Decision Tree* Dalam Klasifikasi Konsumsi Obat. *Jurnal Litbang Edusaintech*, 33-41.
- [18] Lasulika, M. E. (2019). Komparasi *Naïve Bayes*, *Support Vector Machine* Dan *K-Nearest Neighbor* Untuk Mengetahui Akurasi Tertinggi Pada Prediksi Kelancaran Pembayaran Tv Kabel. *ILKOM Jurnal Ilmiah*, 11-16.
- [19] Robertson, S. (2004). Understanding inverse document frequency : on theoretical arguments for *IDF*. *Journal of documentation* 60.5, 503 - 520.
- [20] Ismuhamdan, M. E.-1. (Ismuhamdan, M. E., Prajoko, P., & Apriandari, W. (2023)). Klasifikasi Efektivitas Kebijakan Relokasi Pusat Jajanan Kota Sukabumi Menggunakan Algoritma *Naïve Bayes Classifier*. *Bit (Fakultas Teknologi Informasi Universitas Budi Luhur)*, 20(2), 103-112.