

ANALISA SENTIMEN MASYARAKAT TERHADAP PENGGUNAAN CHATGPT MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)

Arya Damar Pratama^{*1)}, Hendry²⁾

1. Universitas Kristen Satya Wacana, Salatiga, Indonesia
2. Universitas Kristen Satya Wacana, Salatiga, Indonesia

Article Info

Kata Kunci: *Support Vector Machine (SVM), Confusion Matrix, CRISP-DM, Rapidminer, ChatGPT, Analisa Sentimen*

Keywords: *Support Vector Machine (SVM), Confusion Matrix, CRISP-DM, Rapidminer, ChatGPT, Sentiment Analysis*

Article history:

Received 30 November 2023

Revised 14 December 2023

Accepted 28 December 2023

Available online 1 March 2024

DOI :

<https://doi.org/10.29100/jupi.v9i1.4285>

* Corresponding author.

Arya Damar Pratama

E-mail address:

672019227@student.uksw.edu

ABSTRAK

Bekembangnya teknologi yang semakin modern menyebabkan manusia dimudahkan dalam melakukan aktivitas kehidupannya. Salah satu teknologi yang sedang berkembang untuk memudahkan aktivitas manusia adalah hadirnya AI (Artificial Intelligence) yaitu kecerdasan buatan yang mampu belajar dengan sendirinya karena dibentuk dengan algoritma machine learning didalamnya. Salah satu wujud penerapannya adalah hadirnya chatgpt yaitu sebuah AI yang mampu berinteraksi kepada user melalui inputan user seperti menjawab segala pertanyaan yang diberikan. Dalam penelitian ini akan dilakukan analisa sentimen terhadap penggunaan chatgpt untuk mengetahui pandangan masyarakat apakah positif, negatif dan netral terhadap hadirnya chatgpt dimasyarakat. Untuk pengambilan data berasal dari twitter dengan bantuan application programming language (API) token key yang diintegrasikan dengan tools olah data yang digunakan yaitu rapidminer. Kemudian untuk metode olah data menggunakan metode CRISP-DM sedangkan model algoritma yang digunakan adalah support vector machine (SVM). Data yang diperoleh dalam penelitian ini sebanyak 2000 data tweet namun setelah dilakukan pre processing data menjadi 790 data tweet. Kemudian data hasil preprocessing data tersebut diolah dengan memasukan model algoritmanya dan ditemukan hasil dengan ditunjukkanya confusion matriks yang hasilnya yaitu memiliki hasil performa keakuratan yang cukup baik dengan nilai akurasi sebesar 87,6%. Kemudian untuk sentimen yang lebih dominan adalah sentimen netral dengan nilai precision 96,53% dan recall 85,67%.

ABSTRACT

The development of increasingly modern technology makes it easier for humans to carry out their life activities. One of the technologies being developed to facilitate human activity is the presence of AI (Artificial Intelligence), namely artificial intelligence that is able to learn by itself because it is formed with machine learning algorithms in it. One form of its application is the presence of ChatGpT, which is an AI that is able to interact with users through user input such as answering all questions given. In this study, sentiment analysis will be carried out towards the use of chatgpt to find out whether the public's views are positive, negative and neutral towards the presence of chatgpt in the community. To retrieve data originating from Twitter with the help of the application programming language (API) token key which is integrated with the data processing tools used, namely rapidminer. Then the data processing method uses the CRISP-DM method while the algorithm model used is a support vector machine (SVM). The data obtained in this study were 2,000 tweet data but after pre-processing the data it became 790 tweet data. Then the data pre-processing results are processed by entering the algorithm model and the results are found by showing the confusion matrix, the result of which is that it has a fairly good accuracy performance result with an accuracy value of 87.6%. Then the more dominant sentiment is neutral sentiment with a precision value of 96.53% and a recall of 85.67%.

I. PENDAHULUAN

DENGAN berkembangnya teknologi setiap tahun. Mampu memudahkan manusia untuk melakukan aktivitasnya dalam bidang apapun seperti bidang keamanan, bidang transportasi, bidang kesehatan, bidang

hiburan, bidang komunikasi, dan lain sebagainya. Serta dengan bantuan teknologi mampu memperkecil risiko kesalahan yang dilakukan manusia ketika sedang bekerja atau melakukan aktivitas keseharian ya. Maka dari itu hadirnya teknologi sangat berpengaruh dalam kehidupan manusia. Salah satu teknologi yang sedang berkembang setiap tahunnya yaitu kecerdasan buatan. Kecerdasan buatan adalah teknologi di dalam mesin yang digunakan seperti di dalam handphone, komputer, laptop, website, dan lain sebagainya yang mampu belajar memahami inputan user setiap waktunya untuk memperoleh jawaban yang maksimal. Dikarenakan kecerdasan buatan masih dalam perkembangan terdapat kelebihan dan kekurangan. Untuk kelebihannya adalah meningkatkan efektifitas & efisien pekerjaan, mengoreksi kesalahan bekerja, meningkatkan kualitas bekerja, dan mampu membuat inovasi baru sedangkan kekurangannya adalah belum mampu mengerti seluruh bahasa manusia, belum bisa membedakan informasi keasliannya, keakuratan informasi yang diberikan, dan sumber informasi terpercaya. Dikarenakan hal tersebut penelitian ini dibuat untuk mengetahui pandangan masyarakat tentang kehadiran kecerdasan buatan mampu memberikan dampak positif atau negatif dalam kehidupan manusia untuk seterusnya. Untuk mendukung proses penelitian ini, penulis membutuhkan *tools rapidminer* yaitu aplikasi data mining yang membantu mengelola data menjadi informasi baru menggunakan metode yang sesuai dengan kebutuhan penulis. Dalam hal ini penulis mengambil contoh wujud penerapan kecerdasan buatan yaitu *ChatGPT*. *ChatGPT* merupakan *chatbot* dengan kecerdasan buatan yang dibangun dengan *neural network* sehingga mampu belajar dari informasi yang dicantumkan user. Fungsi dari *ChatGPT* adalah menjadi media baru bagi manusia dalam mencari informasi yang lebih interaktif. Seperti mampu memberikan informasi sesuai dengan bahasa yang diinginkan serta mampu membantu manusia dalam menyusun kalimat atau mendapat jawaban dari sebuah pertanyaan. Terdapat banyak jenis *chatbot* yang digunakan manusia namun penulis memilih *ChatGPT* sebagai subjek penelitian.

Penelitian sebelumnya yang membahas terkait kecerdasan buatan yang sesuai dengan topik penulis masih minim namun penulis memiliki referensi terkait hal tersebut secara garis besar memiliki persamaan pembahasannya. Salah satu manfaat adanya *ChatGPT* hadir dalam sebuah penelitian yang membahas tentang kegunaan *ChatGPT* dalam membantu guru atau pun siswa untuk menghadapinya dengan mengoptimalkan sebagaimana mestinya *ChatGPT*. Dengan menggunakannya secara bijak maka bisa membantu meringankan pekerjaan mereka. Di era Education 4.0 ini, guru/dosen memiliki peluang besar untuk memberi teknologi kepada siswa untuk mendukung pekerjaan mereka, tidak hanya dilatih namun diajarkan agar siswa dapat menggunakan teknologi untuk menemukan jawaban atas permasalahannya (Savitri, 2019) [22]

Dalam penelitian yang membahas tentang manfaat hadirnya AI dalam kehidupan manusia terlebih dalam membantu pekerjaan atau aktivitas manusia. Berbagai bidang mampu dibantu AI untuk membantu menemukan solusi yang dibutuhkan dari permasalahan yang dihadapi. Walaupun keberadaannya sangat membantu dengan terobosan solusinya namun bisa menjadi ancaman bagi kehidupan manusia itu sendiri seperti persaingan kerja sudah bertambah oleh adanya AI yang menjadikan persaingan lebih sulit maka dari itu manusia harus berkembang dan mengasah kemampuannya, terlebih ilmu yang mendalami AI supaya manusia tidak tergantikan sepenuhnya oleh AI. [13]

Dalam penelitian yang bertujuan untuk mengetahui pendapat masyarakat tentang sentimen positif, negatif, dan netral di media Twitter tentang upaya pemerintah mencegah Covid-19 di Indonesia dapat dijadikan indikator keberhasilan pemerintah. Untuk mengilustrasikan hal ini, penambahan teks dilakukan menggunakan metode SVM, Naive Bayes, dan k-NN untuk mengklasifikasikan sintesis dari kumpulan data. Hasil pengujian menunjukkan akurasi Naive Bayes sebesar 84,58%, Support Vector Machine sebesar 92,93%, dan k-NN sebesar 83,70%. Hasil akurasi Naive Bayes sebesar 82,14%, SVM sebesar 95,70 & k-NN sebesar 80,66%. Hasil pengembalian Naive Bayes juga 85,82%, SVM 89,17% dan k-NN 84,13%. Dari sini dapat disimpulkan bahwa Support Vector Machine adalah yang terbaik karena memiliki presisi dan akurasi yang paling tinggi. Kedepannya

kita perlu menggunakan dataset yang lebih besar dan lebih kompleks serta meningkatkan pra-pemrosesan bahasa Indonesia non-standar[14]

Dalam penelitian yang membahas tentang perkembangan vtuber yang diambil datanya dari twitter dalam jangka waktu 1 Juni sampai 23 Desember 2021 yang terkumpul 1221 data dan tersisa 321 data dari hasil preprocessing. Yang diperoleh hasil bahwa Hasil pengujian model menyimpulkan bahwa SVM menghasilkan tingkat akurasi sebesar 88,18%. Dengan teknik SMOTE dapat dibuktikan dapat meningkatkan akurasi algoritma SVM dataset yang ada sedikit. tentang besarnya nilai positif 89,09% pada hasil algoritma SVM yang telah dilakukan dan berdasarkan hasil dan pembahasan menunjukkan bahwa metode SVM dapat diterapkan pada analisis data emosional data Twitter berbahasa Indonesia[18]

Dalam penelitian yang membahas tentang hasil implementasi algoritma Support Vector Machine (SVM). Pada penelitian ini data menunjukkan bahwa tingkat akurasi pinjaman online sebesar 62,00%. Hasil pencarian yang tepat dinilai cukup bagus. Analisis sentimen dengan algoritma Support Vector Machine (SVM). berhasil merangking sentimen masyarakat di Twitter tentang pinjaman online. Hasil klasifikasi menunjukkan bahwa sentimen negatif melebihi sentimen positif, dengan 59% ulasan negatif dan 41% ulasan positif. Dapat disimpulkan bahwa masih banyak orang yang masih takut dengan pinjaman online karena risikonya cukup tinggi. Dan banyaknya korban pinjaman online membuat orang takut untuk mencoba. Di sisi lain, sebagian orang memiliki pendapat positif tentang pinjaman online karena dapat memenuhi kebutuhan keuangannya. [17]

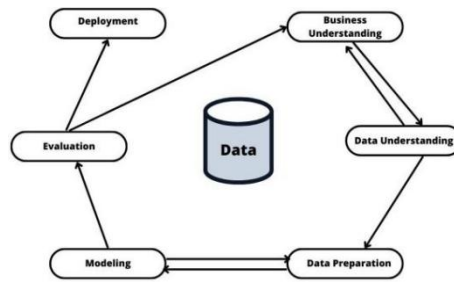
Berdasarkan latar belakang diatas serta dengan penelitian sebelumnya diperoleh susunan masalah yaitu bagaimana model klasifikasi sentimen dengan *Support Vector Machine (SVM)* mengenai sentimen masyarakat Indonesia terhadap penggunaan *chatgpt*. Untuk ruang lingkup penelitian ini menggunakan data *tweet* berbahasa Indonesia pada tahun 2023. Data yang digunakan menggunakan *keyword chatgpt* dan metode klasifikasi dengan *Support Vector Machine (SVM)*. Tujuan adanya penelitian ini adalah untuk mengetahui sentimen yang cenderung dominan terhadap penggunaan *chatgpt* di Indonesia apakah positif, negatif dan netral menggunakan model klasifikasi *Support Vector Machine (SVM)* sebagai model pengukur hasil akurasi penelitian serta *tools* yang digunakan adalah *rapidminer*.

II. METODE PENELITIAN

Jenis tipe analisa sentimen yang digunakan *fine-Grained Sentiment Analysis* yaitu analisa yang fokus terhadap mengelompokkan respon atau pendapat ke dalam beberapa kategori seperti positif, netral, dan negatif. Serta untuk model perhitungannya menggunakan model *supervised learning* yaitu *support vector machine (SVM)*. Dengan model tersebut akan memperoleh hasil olah data analisa sentimen nya. Sehingga akan didapat informasi baru yang diambil dari penelitian ini. Penggunaan metode tersebut digunakan karena ketika data akan diolah dengan Algoritma SVM yang di implementasi kan menggunakan kernel yang mengubah ruang input data menjadi format yang diperlukan. SVM menggunakan teknik yang disebut trik kernel, di mana kernel mengambil ruang masukan berdimensi rendah dan mengubahnya menjadi ruang berdimensi lebih tinggi. Sederhananya, kernel mengubah masalah yang tidak dapat dipisahkan menjadi masalah yang dapat dipisahkan dengan menambahkan lebih banyak dimensi ke dalamnya. Itu membuat SVM lebih efisien, fleksibel dan akurat. Maka dari itu banyak penelitian sebelumnya menggunakan metode *support vector machine (SVM)*.

Dalam analisis sentimen, peneliti menggunakan *tool Rapid Miner* untuk membantu mengelola data, menganalisa dan mampu menemukan informasi terbaru yang diwujudkan berupa bentuk visualisasi dan angka dari data yang telah diolah. Sesuai dengan topik peneliti yaitu analisa sentimen maka dengan bantuan *tools* tersebut mampu mengklasifikasikan dan mengidentifikasi kategori sentimen positif, negatif dan netral manakah yang cenderung lebih banyak jumlahnya yang cukup dengan cara *drag & drop operator* dalam *rapid miner*.

Dalam penelitian ini untuk mendukung proses *data mining* maka digunakan salah satu metode yaitu CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Dikarenakan merupakan metode paling umum dan sering digunakan untuk kebutuhan penelitian. Metode tersebut memiliki beberapa tahap meliputi *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment*. Setiap tahapan memiliki fungsinya masing - masing yang dijelaskan dalam diagram gambar 1 di bawah ini.



Gambar. 1. Metode CRISP-DM (Cross-Industry Standard Process for Data Mining)

A. Business Understanding (Pemahaman Bisnis)

Merupakan tahap melibatkan pemahaman yang mendalam tentang masalah bisnis yang akan diselesaikan dan tujuan yang ingin dicapai melalui analisis data. Dalam topik ini adalah analisa sentimen penggunaan *chatgpt* maka diperlukan pemahaman akan opini masyarakat dalam media *Twitter* supaya hasilnya akan sesuai dengan tujuan yaitu mengetahui tanggapan masyarakat terhadap hadir dan penggunaan *chatgpt* apakah positif, negatif dan netral.

B. Data Understanding (Pemahaman Data)

Merupakan tahap data yang tersedia dikumpulkan, dieksplorasi, dan dipahami secara menyeluruh. Dalam penelitian ini menggunakan *dataset* dari *twitter* yang dicari berdasarkan kata kunci *chatgpt*. Adanya kata kunci yang sesuai mampu mempermudah memperoleh data yang diinginkan.

C. Data Preparation (Persiapan Data)

Merupakan tahap melibatkan persiapan data untuk analisis lebih lanjut dengan melakukan langkah *preprocessing data* untuk merapihkan dan menyeleksi data mentah menjadi data siap pakai. Beberapa tahapan yang dilakukan dalam *rapid miner* terdiri dari *tokenize*, *filter stopwords*, *transform cases*, & *filter tokens*.

D. Modelling (Pemodelan)

Merupakan tahap mengimplementasikan model yang akan digunakan untuk mengelola data. Model *support vector machine (SVM)* terhadap topik analisa sentimen pengaruh *chatgpt*. Hasil dari modeling berupa klasifikasi kategori positif, negatif dan netral yang ditampilkan dalam *confusion matrix* untuk mengetahui hasilnya lebih domain dan akurasi model tersebut.

E. Evaluation (Evaluasi)

Merupakan tahap ini melibatkan evaluasi hasil dari model atau teknik analisis yang diterapkan. Untuk meninjau keakuratan hasil proses pengolahan data berasal dari *Confusion Matrix* yang dihasilkan dalam *Rapid Miner*.

F. Deployment (Penyebaran)

Merupakan tahap dari hasil informasi yang telah didapatkan untuk dilampirkan dalam sebuah laporan sederhana dan artikel tentang hasil penelitian.

III. HASIL DAN PEMBAHASAN

Pada penelitian yang dilakukan, bagian ini adalah tahap hasil dari uji analisis pada penelitian untuk mengimplementasikan metode *support vector machine (SVM)* dalam mengelola dan mengklasifikasi sentimen terkait pengaruh *chatgpt*. Penjelasan hasil analisis dijabarkan dalam bentuk langkah metode CRISP-DM sebagai berikut :

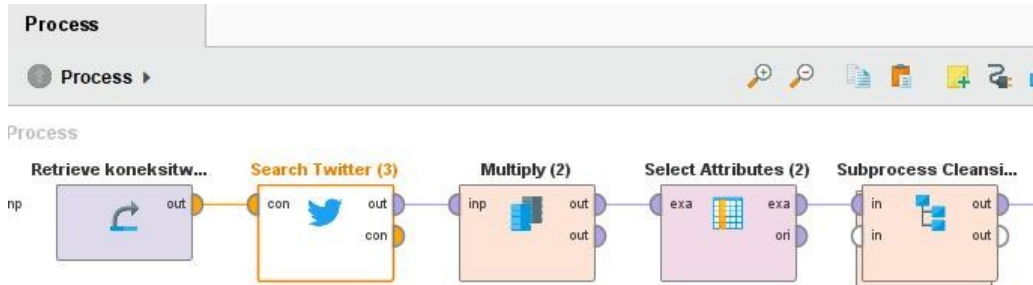
A. Business Understanding

Penelitian analisa sentimen ini digunakan untuk mengetahui opini masyarakat terhadap penggunaan adanya *chatgpt*. Dengan hadirnya AI (*artificial intelligence*) mampu membantu masyarakat dalam aktivitas dalam kehidupannya. Maka dari itu untuk menunjang penelitian tersebut dibutuhkan dataset yang memiliki opini masyarakat yaitu dataset *twitter*. Didalam dataset tersebut akan diolah kembali untuk menilai opini yang bersifat positif, negatif dan netral.

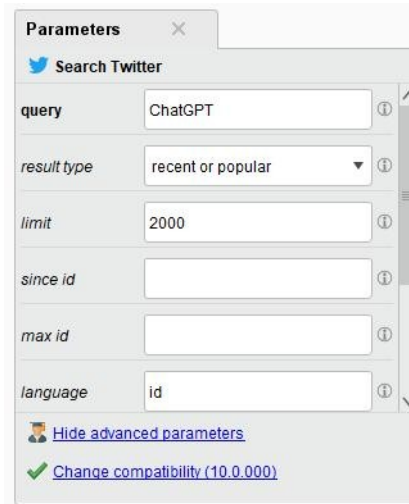
B. Data Understanding

Dataset *twitter* untuk penelitian ini menggunakan Application Programming interface (API) token yang dihubungkan dengan *rapidminer* sebagai tool utama pengolahan data. Didalam *rapidminer* terdapat berbagai jenis operator yang menunjang proses olah data. Operator adalah tools untuk data mining berupa input, output, visualize, preprocessing, & cleansing data. Operator yang digunakan adalah search *twitter* yang parameternya disesuaikan

kembali untuk memperoleh data yang diinginkan namun sebelum itu dibutuhkan adanya membuat koneksi untuk menghubungkan twitter & rapidminer dengan api yang hasilnya berupa kode token yang dikirimkan ke akun user twitter tersebut. Parameter yang diatur kembali dalam search twitter seperti input query untuk mencari kata kunci, language untuk ambil *tweet* berdasarkan bahasa, & jumlah *tweet* yang diambil. Kemudian setelah data didapat maka diextract menjadi file *excel (xlsx)* untuk mempermudah olah data. Data *tweet* yang diperoleh berjumlah sekitar 2000 *tweet*. Lalu diperlukan operator *select attribut* untuk memilih atribut yang akan digunakan. Atribut yang digunakan dalam penelitian ini berupa *text tweet*.



Gambar. 2. Ambil Data *Twitterx*



Gambar. 3. Parameter *Search Twitter*

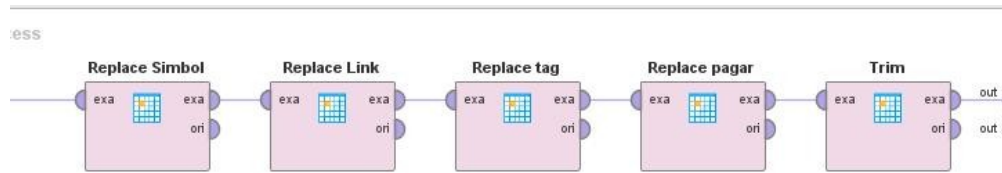
Open in Turbo Prep Auto Model Filter (2,067 / 2,067 examples): all

Row No.	Id	Created-At	From-User	From-User-Id	To-User	To-User-Id	Language	Source	Text
1	1668235243...	Jun 12, 2023 ...	Buletin TV3	155432431	?	-1	in	<a href="http...	KPT Kemer
2	1668296807...	Jun 12, 2023 ...	Xavier Naxa	3286689331	?	-1	in	<a href="http/...	Jangan lupa
3	1668498651...	Jun 13, 2023 ...	Amanz	49794630	?	-1	in	<a href="http...	KPT Tidak M
4	1668569555...	Jun 13, 2023 ...	Fkry	1232434315	?	-1	in	<a href="http/...	RT @Buletin
5	1668568970...	Jun 13, 2023 ...	jemal	1037348989...	?	-1	in	<a href="http/...	RT @Buletin
6	1668568918...	Jun 13, 2023 ...	jaa	3299175139	?	-1	in	<a href="http/...	RT @Buletin
7	1668568912...	Jun 13, 2023 ...	BuzzKini	36328679	?	-1	in	<a href="http/...	"ChatGPT nil
8	1668568185...	Jun 13, 2023 ...	nehiii	376042808	?	-1	in	<a href="http/...	RT @Buletin
9	1668567977...	Jun 13, 2023 ...	shan □	90363577	?	-1	in	<a href="http/...	RT @Buletin
10	1668567755...	Jun 13, 2023 ...	vya	1069911744...	?	-1	in	<a href="http/...	tadi masa sil

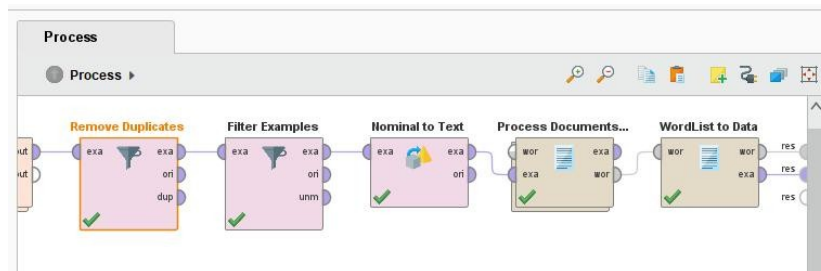
Gambar. 4. Hasil Data Mentah Yang Diperoleh Dari *Twitter*

C. Data Preparation

Setelah data *tweet* terkumpul maka dilanjutkan dengan mengelolah datanya. Diawali dengan *cleansing data* yaitu melakukan pembersihan data *tweet*. Untuk menunjang *cleansing data* maka operator *rapid miner* yang digunakan berupa *replace* untuk menghilangkan tanda baca, simbol, karakter, pagar , link, dan tag serta trim untuk menghapus *space* yang berlebihan dalam 1 *tweet*.

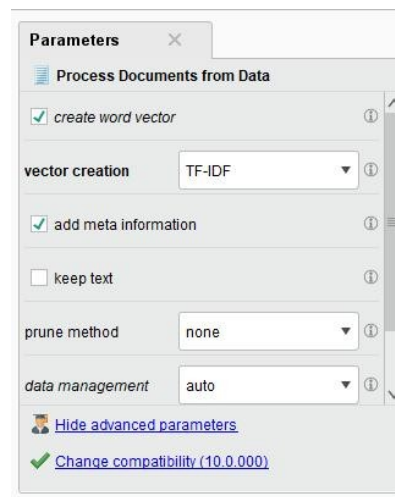


Gambar. 5. Proses *Cleansing Data*



Gambar. 7. Proses *Cleansing Data*

Sebelum menuju proses *pre processing data* dilakukan olah data kembali yaitu dengan operator *remove duplicate* untuk menghapus *tweet* yang sama atau duplikat , *filter examples* untuk memilah atribut *text* yang hanya memiliki nilai didalamnya dan *nominal to text* untuk mengubah jenis nilai atribut normal menjadi text.



Gambar. 7. Parameter dalam *proces document from data*



Gambar. 8. Proses *Pre Processing Data*

Setelah dilakukan cleansing maka dilanjut dengan *pre processing text*. Yaitu proses pengubahan bentuk sebuah data menjadi lebih terstruktur sesuai dengan kebutuhannya dalam proses data mining dan biasanya akan menjadi nilai numerik (Ashari, et al., 2020). Dalam *rapidminer* digunakan *operator tokenize, transform cases, filter stopword & filter tokens* yang diletakan dalam wadah operator *proces document from data* yang parameternya menggunakan TF-IDF (Term Frequency Inverse Document Frequency) merupakan metode yang digunakan untuk menentukan nilai frekuensi sebuah kata di dalam sebuah dokumen atau artikel dan juga frekuensi di dalam banyak dokumen. Perhitungan ini menentukan seberapa relevan sebuah kata di dalam sebuah dokumen (Evan, 2014). Untuk penjelasan oprator didalamnya sebagai berikut :

TABEL I
TABEL DATA DARI PROSES TOKENIZE

Tweet Asli	Tweet Hasil <i>Tokenize</i>
@Yunjhn No comment sih nyets coba kita tanya pendapat chatGPT	["No", "comment", "sih", "nyets", "coba", "kita", "tanya", "pendapat", "chatgpt"]
Duh chatgpt keren bgt deh knp ya km tuh ga exist waktu jaman aku kuliah?!!!!	["Duh", "chatgpt", "keren", "bgt", "deh", "knp", "ya", "km", "tuh", "ga", "exist", "waktu", "jaman", "aku", "kuliah"]
dah kenapa takleh nak register chatgpt ni	["dah", "kenapa", "takleh", "nak", "register", "chatgpt", "ini"]

1. *Tokenize*

Merupakan proses untuk membagi teks yang dapat berupa kalimat, paragraf atau dokumen, menjadi token-token/bagian-bagian tertentu.

TABEL II
TABEL DATA DARI PROSES TRANSFORM CASES

Tweet Asli	Tweet Hasil <i>Transform Cases</i>
Aplikasi ChatGPT Kini Boleh Digunakan Pengguna iPhone Di Malaysia https://t.co/aDb5ZREIMj https://t.co/Arw4PvrjSN	aplikasi chatgpt kini boleh digunakan pengguna iphone di malaysia https://t.co/adb5zrelmj https://t.co/arw4pvrjsn
Ternyata memang belum banyak orang yang menggunakan ChatGPT ... https://t.co/bujpk6ISYA	ternyata memang belum banyak orang yang menggunakan chatgpt ... https://t.co/bujpk6isya
RT @myskill_id: 10 HAL YANG BISA KAMU LAKUIN DI CHATGPT BUAT BANTU KERJAAN sekalian manfaatin teknologi ya khan~ simak gan https://t.co/F...	rt @myskill_id: 10 hal yang bisa kamu lakuin di chatgpt buat bantu kerjaan sekalian manfaatin teknologi ya khan~ simak gan https://t.co/f...

2. *Transform case*

Merupakan proses untuk merubah kalimat menjadi huruf kecil atau huruf besar semua supaya tweet tersebut seragam dengan *tweet* lainnya.

TABEL III
TABEL DATA DARI PROSES FILTER STOPWORD

Tweet Asli	Tweet Hasil <i>Filter Stopword</i>
Duh chatgpt keren bgt deh knp ya km tuh ga exist waktu jaman aku kuliah?!!!!	["chatgpt", "keren", "exist", "waktu", "jaman", "aku", "kuliah"]
@Misaki_Yua Udah dah enak bgt, gas chatgpt wkwkwk	["Enak", "Chatgpt"]
Ternyata chatgpt sangat membantu	["Ternyata", "chatgpt", "sangat", "membantu"]

3. *Filter stopwords*

Merupakan proses untuk memilah kata kata yang tidak memiliki makna atau tidak baku yang kemudian tidak digunakan didalam suatu *tweet* sehingga *tweet* tersebut bisa lebih bersih dan rapi.

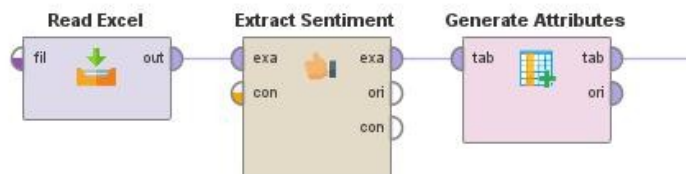
TABEL IV
TABEL DATA DARI PROSES FILTER TOKENS

Tweet Asli	Tweet Hasil <i>Filter Tokens</i>
Duh chatgpt keren bgt deh knp ya km tuh ga exist waktu jaman aku kuliah?!!!!	["chatgpt", "keren", "exist", "waktu", "jaman", "aku", "kuliah"]
Kalo dipikir2 gesrek jg gue pas jadi panitia roro jonggrang kemaren, bikin kata sambutan pake chatgpt HAHA Bikin au bisa pake ChatGPT ga sih anjir gua males bikin fake chat wkwk	["Kalo", "Pikir", "jadi", "panitia", "roro", "jonggrang", "kemaren", "bikin", "kata", "sambutan", "chatgpt"] ["bikin", "bisa", "pake", "chatgpt", "anjir", "males", "bikin", "fake", "chat"]

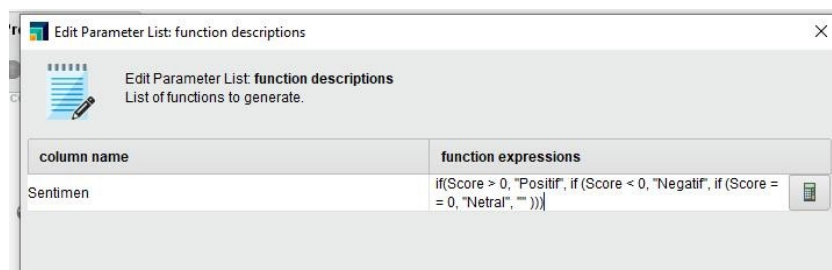
4. Filter tokens

Merupakan proses untuk membatasi jumlah kata yang terbentuk untuk menghindari kata kata tidak baku atau singkatan yang tidak bermakna. Dalam penelitian ini diatur untuk minimal kata berjumlah 4 dan maksimal berjumlah 25.

Kemudian untuk menilai bobot frekuensi kata yang muncul tiap *tweet* digunakan operator *worldlist to data* yaitu Operator yang membuat kumpulan data dari daftar kata. Kumpulan data berisi baris untuk setiap kata dan atribut untuk kata itu sendiri, jumlah dokumen yang terjadi, jumlah dokumen berlabel yang terjadi dan untuk setiap nomor kelas yang terjadi dalam dokumen kelas tersebut.

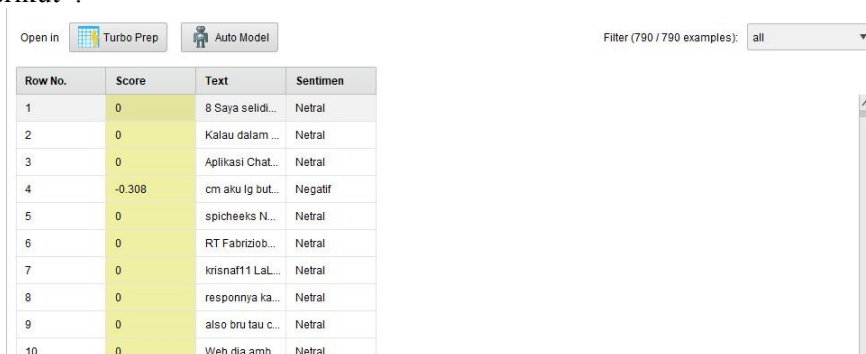


Gambar. 9. Proses *Labeling Data*



Gambar. 10. Proses *Rumus* Penentuan Nilai Sentimen

Sebelum menuju tahap modeling diperlukan tahap labeling data. Yaitu proses data tweet akan diklasifikasikan menjadi positif, negatif dan netral. Diawali dengan menggunakan operator extract sentimen yang menggunakan metode Vader. Metode Vader (Valence Aware Dictionary and sEntiment Reasoner) merupakan lexicon atau analisis sentimen dengan dasar aturan secara khusus yang selaras dengan berbagai sentimen dalam media sosial. Model Vader menggunakan kamus leksikal. Kamus leksikal adalah kumpulan kata-kata dengan informasi tambahan terkait makna, penggunaan, dan karakteristik kata tersebut. Yang kemudian dikuratori secara manual yang berisi kata-kata dengan skor sentimen yang telah ditentukan sebelumnya. Dengan model VADER maka akan menghasilkan skor sentimen yang dihasilkan Untuk memberikan score pada tiap tweet dibantu dengan operator generate attributes untuk memberi nilai positif, negatif dan netral sesuai dengan nilai score yang telah ditentukan rumusnya yaitu $if(Score > 0, "Positif", if (Score < 0, "Negatif", if (Score == 0, "Netral", "")))$. Bila diartikan adalah jika scorenya lebih dari 0 bernilai positif jikalau kurang dari 0 bernilai negatif dan sama dengan 0 bernilai netral. Maka akan tertera hasil labeling sebagai berikut :

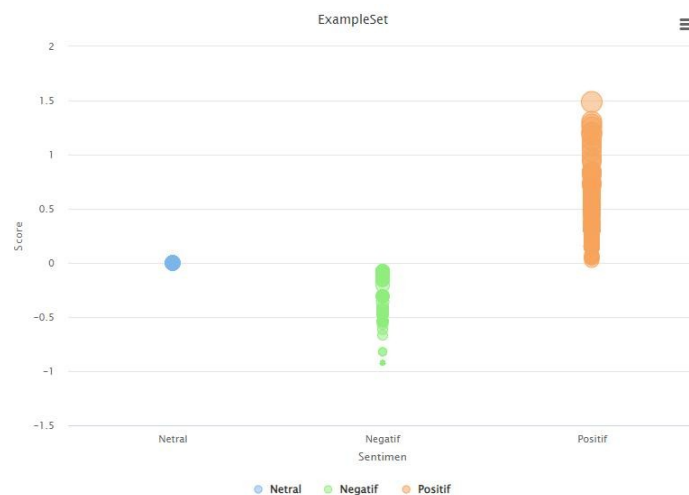


Row No.	Score	Text	Sentimen
1	0	8 Saya selidi...	Netral
2	0	Kalau dalam ...	Netral
3	0	Aplikasi Chat...	Netral
4	-0.308	cm aku lg but...	Negatif
5	0	spicheeks N...	Netral
6	0	RT Fabriziob...	Netral
7	0	krisnaf11 LaL...	Netral
8	0	responnya ka...	Netral
9	0	also oru tau c...	Netral
10	0	Weh dia amb...	Netral

Gambar. 11. Hasil Proses *labeling* data

Contoh hasil data yang telah *labeling*. Diketahui memiliki jumlah hasil 790 data *tweet*, dan 3 atribut yaitu *score* merupakan jumlah nilai sentiment pada atribut *text*, *text* merupakan bentuk tulisan isi dari *tweet*, dan sentiment

merupakan atribut nilai *labeling* yang sudah diproses dengan rumus sebelumnya.



Gambar. 12. Visualisasi Nilai Sentimen

Dari gambar tersebut diketahui bahwa nilai tertinggi untuk nilai sentiment positif bernilai 1.5 , nilai terendah untuk sentiment negative bernilai sekitar -1 dan nilai untuk sentiment netral bernilai 0.

D. Modelling

Metode SVM memungkinkan perhitungan untuk masalah linear dengan menerapkan transformasi matematis untuk ruang belajar menggunakan fungsi kernel [3]. Metode SVM memiliki konsep sentral dalam mengklasifikasikan data, yaitu menentukan hyperplane terbaik untuk memberi jarak atau pemisah antara dua kelas yang telah ditentukan [4]. Cara kerja SVM yaitu dengan menemukan hyperplane optimal yang memberi jarak atau pemisah antar dua kelas. hyperplane yang mempunyai margin maksimum. Jarak antar titik data terdekat dengan hyperplane diklaim sebagai margin. Adapun vector pendukung sebagai titik terdekat dengan hyperplane[5]. Untuk tahapan dalam metode *Support Vector Machine* adalah sebagai berikut:

1. Menentukan kata yang sering muncul dari tiap dokumen atau *tweet* yang digunakan.
2. Menentukan inisialisasi awal untuk nilai $\alpha=0.5, C=1, \lambda=0.5, \text{gamma}=0.5$ dan $\text{epsilon}=0.001$.
3. Menghitung matriks dengan rumus (6) :

$$D_{ij} = Y_i Y_j (K(x_i \rightarrow \cdot x_j \rightarrow) + \lambda^2)$$

Keterangan :

D_{ij} = elemen matriks data ke-ij

Y_i = kelas atau label data ke-I

Y_j = kelas atau label data ke-i

λ = turunan batas teoritis

$(x_i \rightarrow \cdot x_j \rightarrow)$ = fungsi kernel

4. Untuk data ke $n = 1, 2, 3, \dots, n$ gunakan (7),(8),(9) berikut ini :

$$E_i = \sum_{j=1}^n a_i D_{ij} \quad (3)$$

$$\delta a_i = \min\{\max[\gamma(1 - E_i), -a_i], C - a_i\} \quad (8)$$

$$a_i = a_i + \delta a_i \quad (5)$$

Keterangan :

E_i = nilai error data ke-i

Y = tingkat pembelajaran

$max_{(i)} D_{ij}$ = nilai maksimum diagonal matriks hessian

5. Mencari nilai (b) dengan menggunakan persamaan (10)

$$b = -12 [w \cdot x^+ + w \cdot x^-]$$

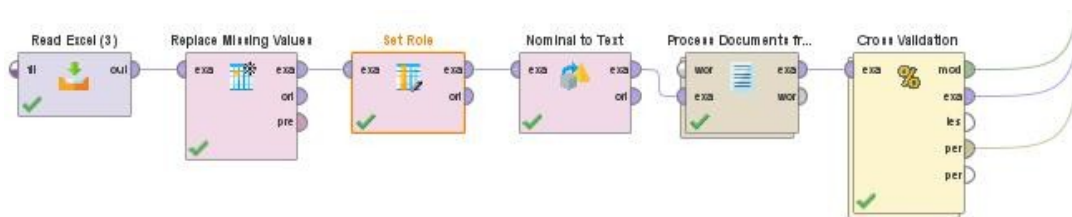
6. Pengujian pada dokumen yang diuji

7. Perhitungan keputusan

Jika berhasil perhitungan keputusan lebih dari sama dengan 0 maka nilai $sign h(x)$ adalah +1, maka termasuk kelas positif dan jika hasil perhitungan keputusan kurang dari 0 maka nilai $sign h(x)$ nilai -1, maka termasuk kelas negatif. Perhitungan keputusan dengan menggunakan persamaan(11).

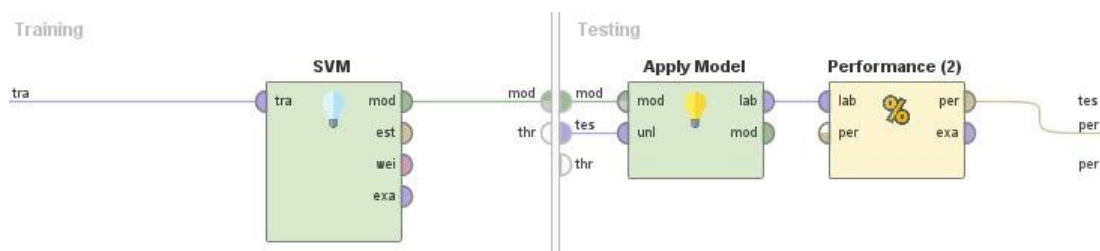
$$h(x) = w \cdot x + b \quad (7) \text{ atau } h(x) = \sum_{i=1}^K \alpha_i y_i K(x, x_i) + b \quad (11)$$

Kemudian dalam proses modeling penerapan model *support vector machine* pada *rapidminer* memerlukan beberapa operator untuk menerapkannya. Diawali dahulu dengan menggunakan data hasil *preprocessing* berupa excel dengan data *tweet* sebanyak 790 *tweet* dari jumlah data mentah sebelumnya sebanyak sekitar 2000 data *tweet* yang dimasukan dalam operator *read excel*. Untuk menghindari *atribut text* yang memiliki nilai *null* maka digunakan *replace missing values*. Dan untuk mengetahui atribut yang diberi akan label dari data *tweet* tersebut maka digunakan *set role*. *Nominal to text* digunakan supaya merubah tipe data angka menjadi text. Kemudian tetap menggunakan *proces document to data* untuk menyeleksi kembali seperti *cleansing data* yang operator didalamnya berupa *tokenize*, *transform cases*, *filter by tokens*, & *stopword*. Untuk awal model *SVM* diawali menggunakan operator *cross validation* yaitu untuk melakukan validasi silang untuk memperkirakan kinerja statistik model pembelajaran.



Gambar. 13. Proses Pembentukan Model SVM

Di dalam operator *cross validation* terdapat dua subproses yaitu *Training* dan *Testing*. Subproses *Training* digunakan untuk melatih model. Model yang dilatih kemudian diterapkan dalam subproses *Testing*. Kinerja model diukur selama fase Pengujian.



Gambar. 14. Penerapan Model SVM

Di dalam *cross validation* terdapat operator algoritma *SVM* didalam *subproses training* dan *subproses testing* terdapat operator *apply model* untuk tujuannya adalah untuk mendapatkan prediksi pada data yang tidak terlihat atau untuk mengubah data dengan menerapkan model *preprocessing* dan *performance* untuk operator ini digunakan untuk evaluasi kinerja. Ini memberikan hasil yang diberikan dari evaluasi kinerja

E. Evaluation

Dalam *evaluasi* terdapat hasil berupa *confusion matriks* yang mengetahui Accuracy, Precision, & Recall. Untuk *accuracy* yaitu rasio prediksi Benar (positif dan negatif) dengan keseluruhan data. *Precision* yaitu rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang di prediksi positif. *Recall* yaitu rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif.

TABEL V
 TABEL HASIL CONFUSION MATRIX SVM

Confusion Matrix				
Pred. Netral	139	4	1	96.53%
Pred. Negatif	0	18	0	100.00%
Pred. Positif	23	0	41	54.06%
Class Recall	85.80%	81.82%	97.62%	

Diketahui dari gambar tersebut bahwa hasil sentimen netral berjumlah 139 dengan *class precision* 96,53% & *class recall* bernilai 85,88%, sentimen positif berjumlah 41 dengan *class precision* 64.05% & *class recall* bernilai 81.82% dan sentimen negatif bernilai 10 dengan *class precision* 100% & *class precision* bernilai 97,62%.

TABEL VI
 TABEL PERFORMA HASIL PERHITUNGAN SVM

Performances		
Criteria	Value	Standard Deviation
Accuracy	87.6%	±2.5%
Classification Error	12.4%	±2.5%
Class Recall	85.80%	81.82%

Pada gambar tersebut diketahui hasil *accuracy* model SVM bernilai 87.6% dengan standard deviasi bernilai kurang lebih 2.5% dan *classification error* bernilai 12.4% dengan standar deviasi kurang lebih 2.5%. Standard deviasi yaitu ukuran penyebaran yang paling baik, karena menggambarkan besarnya penyebaran tiap-tiap unit observasi (Ghozali, 2016) Dan *classification error* adalah kesalahan prediksi yang diperkenalkan, yang membandingkan, untuk setiap baris dalam data uji, nilai sebenarnya dari variabel respons terhadap output model yang dihasilkan ketika nilai baris tersebut untuk variabel penjelas adalah input ke model.

F. Deployment

Evaluasi mendalam terhadap hasil model telah selesai dilakukan, langkah selanjutnya adalah mengimplementasikan keseluruhan model yang telah dikembangkan. Dalam implementasi, penting untuk memastikan bahwa model dapat diintegrasikan dengan sistem atau proses yang ada, dan mampu menghasilkan hasil yang relevan dan bermanfaat sesuai dengan tujuan awal. Selama tahap ini, juga dilakukan penyesuaian terhadap model untuk memastikan bahwa hasil yang dihasilkan sesuai dengan target awal dalam tahap CRISP-DM ini. Terkait pembahasan terdahulu menggunakan metode *Support Vector Machine* memiliki jumlah presisi dan akurasi yang cukup baik sehingga penelitian ini menggunakan metode tersebut. Maka metode SVM mampu di implementasikan sebagai metode klasifikasi dalam melakukan analisa sentimen menggunakan data *Twitter* berbahasa Indonesia

IV. KESIMPULAN

Dari penelitian analisa sentimen yang dilakukan diperoleh hasil bahwa penggunaan metode *support vector machine (SVM)* memiliki hasil performa keakuratan yang cukup baik dengan nilai akurasi sebesar 87,6%. Kemudian untuk sentimen yang lebih dominan adalah sentimen netral dengan nilai *precision* 96,53% dan *recall* 85,67% daripada sentimen negatif dan positif sehingga diketahui bahwa masyarakat Indonesia masih netral terhadap penggunaan *chatgpt* dalam membantu kehidupan aktivitas manusia. Serta penggunaan *chatgpt* masih belum digunakan sebagai *tools* utama yang membantu aktivitas masyarakat Indonesia sehingga masih asing untuk dioptimalkan penggunaannya.

Sebagai saran untuk penelitian berikutnya dengan topik serupa yaitu dilakukan dengan metode yang lebih beragam dan memiliki jumlah data yang lebih banyak supaya mampu menggali lebih mendalam sehingga mampu menemukan informasi terbaru akan topik tersebut yang mampu memberikan informasi kepada pembaca atau masyarakat Indonesia.

DAFTAR PUSTAKA

- [1] Akmalia, Rizka, Isnandar Slamet, Hasih Pratiwi. 2022. Analisis Sentimen Twitter Berbahasa Indonesia Terhadap Aplikasi Peduli Lindungi dengan Algoritma SVM, KNN, dan Regresi Logistik. Prosiding Seminar Nasional MIPA UNIPA 2022 (ISSN -2654-7724)
- [2] A. Rahmansyah, O. Dewi, P. Andini, T. Hastuti, P. Ningrum, and M. E. Suryana, "Membandingkan Pengaruh Feature Selection Terhadap Algoritma Naïve Bayes dan Support Vector Machine," *Semin. Nas. Apl. Teknol. Inf.*, pp. 1–7, 2018.
- [3] Y. Al-Amrani, M. Lazaar, and K. E. El Kadiri, "Sentiment Analysis Using Hybrid Method Of Support Vector Machine And Decision Tree," *J. Theor. adn Appl. Inf. Technol.*, vol. 96, no. 7, pp. 1886–1895, 2018.
- [4] K. A., *Support Vector Machines Succinctly*. 2017.
- [5] S. Rani and J. Singh, "Sentiment Analysis Of Tweets Using Support Vector Machine," *Int. J. Comput. Sci. Mob. Appl.*, vol. 5, no. 10, pp. 83–91, 2017.

- [6] H. – W. A. Social, “Social Media Advertising Audiances.” 2019.
- [7] I. M. B. S. Darma, R. S. Perdana, and Indriati, “Penerapan Sentimen Analisis Acara Televisi Pada Twitter Menggunakan Support Vector Machine dan Algoritma Genetika sebagai Metode Seleksi Fitur,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 3, pp. 998–1007, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [8] I. P. Windasari, F. N. Uzzi, and K. I. Satoto, “Sentiment Analysis on Twitter Posts: An analysis of Positive or Negative Opinion on GoJek,” *Int. Conf. Inf. Tech, Comput. Electr. Eng.*, pp. 266–269, 2017.
- [9] M. Ahmad, S. Aflab, and I. Ali, “Sentiment Analysis of Tweets using SVM,” vol. 177, no. 5, pp. 25–29, 2017.
- [10] I. D. Sholihati, Irmawati, and D. Glory, “Aplikasi Data Mining Berbasis Web Menggunakan Algoritma Apriori untuk Data Penjualan di Apotek,” *Semin. Nas. Teknol. Informasi, Komun. dan Apl.*, vol. 4, no. 2, pp. 121–126, 2019.
- [11] M. Cindo, D. P. Rini, and Ermatita, “Studi Komparatif Metode Ekstraksi Fitur pada Analisis Sentimen Maskapai Penerbangan Menggunakan Support Vector Machine dan Maximum Entropy,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 402–407, 2019.
- [12] Novantika, A., & Sugiman, S. (2022, February). Analisis Sentimen Ulasan Pengguna Aplikasi Video Conference Google Meet menggunakan Metode SVM dan Logistic Regression. In PRISMA, Prosiding Seminar Nasional Matematika (Vol. 5, pp. 808-813).
- [13] Pakpahan, R. (2021). Analisa Pengaruh Implementasi Artificial Intelligence Dalam Kehidupan Manusia. JISICOM (Journal of Information System, Informatics and Computing), 5(2), 506-513.
- [14] Hikmawan, S., Pardamean, A., & Khasanah, S. N. (2020). Sentimen Analisis Publik Terhadap Joko Widodo terhadap wabah Covid-19 menggunakan Metode Machine Learning. *Jurnal Kajian Ilmiah*, 20(2).
- [15] Normah, N., Rifai, B., Vambudi, S., & Maulana, R. (2022). Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE. *Jurnal Teknik Komputer AMIK BSI*, 8(2), 174-180.
- [16] Sari, D. N., Adelia, F., Rosdiana, F., Butar, B. B., & Hariyanto, M. (2020). Analisa Sentimen Terhadap Review Produk Kecantikan Menggunakan Metode Naive Bayes Classifier. *JIKA (Jurnal Informatika)*, 4(3), 109-118.
- [17] Utami, D. S., & Erfina, A. (2021, September). Analisis Sentimen Pinjaman Online di Twitter Menggunakan Algoritma Support Vector Machine (SVM). In *Seminar Nasional Sistem Informasi dan Manajemen Informatika Universitas Nusa Putra* (Vol. 1, No. 01, pp. 299-305).
- [18] Indrayuni, E. (2016). Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization. *EVOLUSI: Jurnal Sains dan Manajemen*, 4(2).
- [19] Bestari, Novina Putri. (14 April 2023). Heboh di Amerika, ChatGPT Sepi Pengguna di Indonesia. <https://www.cnbcindonesia.com/tech/20230414131433-37-429971/heboh-di-amerika-chatgpt-sepi-pengguna-di-indonesia>.
- [20] Binus. (14 Februari 2022). Support Vector Machine Algorithm. <https://sis.binus.ac.id/2022/02/14/support-vector-machine-algorithm/>
- [21] Binus. (18 September 2020). Cross-Industry Standard Process for Data Mining (CRISP-DM). <https://mmsi.binus.ac.id/2020/09/18/cross-industry-standard-process-for-data-mining-crisp-dm/>
- [22] Setiawan, A., & Luthfiyani, U. K. (2023). Penggunaan ChatGPT Untuk Pendidikan di Era Education 4.0: Usulan Inovasi Meningkatkan Keterampilan Menulis. *JURNAL PETISI (Pendidikan Teknologi Informasi)*, 4(1), 49-58.
- [23] Tuhuteru, H. (2020). Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berskala Besar Menggunakan Algoritma Support Vector Machine. *Journal Information System Development (ISD)*, 5(2).