

# Analisis Sentimen Opini Masyarakat Terhadap Stadion Jakarta Internasional Stadium (JIS) Pada Twitter Dengan Perbandingan Metode Naive Bayes Dan Support Vector Machine

Rasiban<sup>1\*</sup>, Sugeng Riyadi<sup>2</sup>

<sup>1,2</sup>Program Studi Sistem Informasi, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Jakarta Timur, Indonesia

<sup>1</sup>rasiban.stikom@yahoo.com, <sup>2</sup>arie.as89@gmail.com

**Abstrak**— Berdasarkan pencarian di media sosial Twitter, Peneliti mengidentifikasi beberapa masalah yang sering disorot oleh masyarakat terkait Isu Stadion Jakarta Internasional Stadium (JIS) Belum berstandar FIFA. Isu ini dilontarkan pertama kali oleh Hunter Jagal, Postingan akun media sosial twitter @hunterjagar3 dengan Konten Tweet “Kaesang pengarep berkomentar terkait pssi yang menyebut Jakarta Internasional Stadium belum berstandar FIFA. Maka Peneliti melakukan Analisis Sentimen Opini Masyarakat Terhadap Stadion Jakarta Internasional Stadium (Jis) Pada Twitter Dengan Perbandingan Metode Naive Bayes Dan Support Vector Machine. Hasil akhir dari Perbandingan dengan dua metode pengujian ini, yaitu hasil prediksi Sentimen Masyarakat Terhadap Isu Stadion Jakarta Internasional Stadium belum berstandar FIFA berdasarkan data yang didapat dari Twitter dan diimplementasikan dengan metode Naive Bayes menunjukkan nilai akurasi sebesar 99.57%. Dari 940 data uji, terprediksi sebesar 892 data sebagai Prediksi Sentimen Negatif dan 48 data sebagai Sentimen Positif dan Metode Support Vector Machine menunjukkan nilai akurasi sebesar 99.68%. Dari 940 data uji, terprediksi sebesar 894 data sebagai Sentimen Negatif dan 46 data sebagai Sentimen Positif.

**Kata Kunci:** *Sentimen, Twitter, JIS, NB, SVM, Rapidminer*

**Abstract**— Based on searches on Twitter social media, researchers identified several problems that are often highlighted by the public regarding the Issue of the Jakarta International Stadium (JIS) Stadium, which does not yet have a FIFA standard. This issue was raised for the first time by Hunter Butcher, posting the social media account Twitter @hunterjaga3 with the Tweet Content "Kaesang Pengarep commented regarding the PSSI which said the Jakarta International Stadium was not yet FIFA standard. So the researchers conducted an analysis of public opinion sentiment towards the Jakarta International Stadium (JIS) on Twitter with a comparison of the Naive Bayes method and the Support Vector Machine. The final result of the comparison with these two testing methods, namely the prediction results of community sentiment towards the issue of the Jakarta International stadium stadium, is not yet standardized. FIFA based on data obtained from Twitter and implemented with the Naive Bayes method shows an accuracy value of 99.57%. Of the 940 test data, it is predicted that 892 data are Negative Sentiment Predictions and 48 data are Positive Sentiment and the Support Vector Machine Method shows an accuracy value of 99.68%. Of the 940 test data, it is predicted that 894 data will be Negative Sentiment and 46 data will be Positive Sentiment

**Keywords:** *Sentiment, Twitter, JIS, NB, SVM, Rapidminer*

## I. PENDAHULUAN

Stadion Jakarta International Stadium (JIS) adalah sebuah stadion yang baru dibangun di Indonesia dan merupakan salah satu stadion terbesar di Asia Tenggara. Sebagai stadion yang baru dibangun, tentu saja stadion ini menjadi perhatian banyak pihak, baik itu dari kalangan penggemar olahraga, pejabat pemerintah, hingga masyarakat umum. Dalam era digital seperti saat ini, media sosial menjadi salah satu tempat yang banyak digunakan masyarakat untuk menyampaikan opini mereka mengenai suatu topik, termasuk stadion JIS. Oleh karena itu, analisis sentimen opini masyarakat terhadap stadion JIS pada Twitter menjadi topik yang menarik untuk diteliti. Dalam analisis sentimen tersebut, terdapat dua metode yang umum digunakan, yaitu metode Naive Bayes dan Support Vector Machine (SVM). Kedua metode ini telah banyak digunakan dalam berbagai penelitian analisis sentimen, termasuk dalam penelitian ini. Dengan membandingkan hasil analisis sentimen yang menggunakan kedua metode tersebut, diharapkan dapat diperoleh pemahaman yang lebih mendalam mengenai sentimen masyarakat terhadap stadion Jakarta Internasional Stadium (JIS) di Twitter,

serta manfaat dari masing-masing metode tersebut dalam melakukan analisis sentimen pada data yang besar seperti data Twitter. Berdasarkan pencarian di media sosial Twitter, Peneliti mengidentifikasi beberapa masalah yang sering disorot oleh masyarakat terkait Isu Stadion Jakarta Internasional Stadium (JIS) Belum berstandar FIFA. Isu ini dilontarkan pertama kali oleh Hunter Jagal, Postingan akun media sosial twitter @hunterjagar3 dengan Konten Tweet “Kaesang pengarep berkomentar terkait pssi yang menyebut Jakarta Internasional Stadium belum berstandar FIFA.

Berdasarkan hasil pengujian dan analisis yang telah dilakukan ditemukan bahwa Social Influence berpengaruh positif terhadap *Perceived Ease of Use*, *Social Influence* berpengaruh terhadap *Perceived Usefulness*, *Perceived Usefulness* berpengaruh terhadap *Attitude Toward Using*, *Perceived Ease of Use* tidak berpengaruh terhadap *Attitude Toward Using*, jadi tingkat *Perceived Ease of Use* seseorang rendah atau tinggi tidak akan mempengaruhi *Attitude Toward Using*. Hal ini terjadi karena sikap seseorang terhadap aplikasi tidak hanya berpengaruh karena kemudahan melainkan orang sekita apakah menggunakan

dan tingkat keamanan dari aplikasi tersebut, *Perceived Usefulness* berpengaruh terhadap *Intention to Use*, *Attitude Toward Using* tidak berpengaruh terhadap *Intention to Use* jadi tingkat *Attitude Toward Using* seseorang rendah atau tinggi tidak akan mempengaruhi *Intention to Use* hal ini terjadi karena sikap orang terhadap aplikasi tidak bisa menarik perhatian orang lain untuk menggunakan karena orang melihat aplikasi dari kegunaannya dan pas dengan orang tersebut atau tidak[2]. Penelitian ini bermaksud untuk mendapatkan informasi sentimen terkait opini masyarakat yang berhubungan dengan penggunaan aplikasi PeduliLindungi, dengan mengaplikasikan algoritma *Support Vector Machine* serta *kernel Radial Basis Function* dan algoritma seleksi fitur yaitu *Particle Swarm Optimization* dalam mengklasifikasikan opini masyarakat terhadap aplikasi PeduliLindungi dari data tweet yang sudah diperoleh serta diberi label sentimen bersifat positif dan label sentimen bersifat negatif. Model *Support Vector Machine* menghasilkan akurasi sebesar 76.24%, *recall (sensitivity)* sebesar 82.14%, presisi sebesar 76.67%, dan *specificity* sebesar 68.89%, sedangkan model *Support Vector Machine* ditambahkan seleksi fitur yaitu *Particle Swarm*

*Optimization* mengalami peningkatan akurasi sebesar 88.12%, *recall (sensitivity)* sebesar 96.43%, presisi sebesar 84.36%, dan *specificity* sebesar 77.78%[3]. Covid-19 adalah penyakit menular yang sudah menyebar ke Indonesia. Pemantauan penyebaran Covid-19 di Indonesia ditangani oleh Kementerian Komunikasi dan Informatika (KOMINFO) dengan membuat aplikasi PeduliLindungi yang dapat ditemukan di Google Play. Pengguna akan memilih aplikasi yang memiliki ulasan yang bagus, tetapi untuk memantau ulasan dari masyarakat tidak mudah sehingga penulis ingin mengetahui analisis review pengguna aplikasi PeduliLindungi berdasarkan komentar pengguna dengan menggunakan algoritma *Support Vector Machine*[4]. Analisis sentimen yang dilakukan dengan mengklasifikasikan ulasan menjadi ulasan positif dan ulasan negatif menggunakan algoritma *Support Vector Machine* dengan seleksi fitur *chi-square*. Pengumpulan data ulasan dilakukan dengan melakukan *scrapping* di google play dengan menggunakan bahasa pemrograman *Python*. Hasil dari klasifikasi sentimen terhadap aplikasi PeduliLindungi menghasilkan performa yang baik dan menghasilkan nilai akurasi sebesar 93%, *recall* sebesar 86%, *precision* sebesar 98%, *specificity* sebesar 98% dan *f1-score* sebesar 92%[5]. Penggunaan metode ini diharapkan mendapatkan akurasi yang tinggi sehingga dengan metode tersebut dapat mengklasifikasi komentar negatif dan komentar positif sehingga mendapatkan evaluasi yang dapat meningkatkan pelayanan kepada masyarakat melalui aplikasi ini. Tahapan-tahapan yang dilakukan dalam penelitian ini menggunakan tahapan *praproses* data seperti *case folding*, *filtering*, *normalisasi* kata, *stopword removal*, *stemming*, dan *tokenisasi*[6]. *The algorithm used is Support Vector Machine and Naive Bayes to classify the data set. The data collection technique*

*is Text Mining and compares the results of the two specified algorithms. The results of this research are Support Vector Machine with TF IDF Vectorizer with 89.05% accuracy followed by Support Vector Machine with Count Vectorizer, Naive Bayes with TF IDF Vectorizer and Naive Bayes with Count Vectorizer*[7]. Dengan melihat ulasan calon pengguna dapat melihat tanggapan pengguna lain yang sudah menggunakan aplikasi lebih dahulu, selain itu ulasan aplikasi dapat dijadikan masukan untuk pengembang. Pengelompokan ulasan dapat menggunakan teknik *text mining* atau penambangan *text*[8]. *During the COVID-19 pandemics, the government has carried out various programs to break the chain of spread, one of which is by creating an application for providing information and services regarding COVID-19, called Peduli Lindungi. This study aims to analyze the public's response to the Peduli Lindungi application using the Naive Bayes Classifier, Support Vector Machine and K-Nearest Neighbors. Results for the classification of sentiments based of SVM obtained 76,5%, the accuracy of the NBC is 72,3%, the accuracy of the KNN is 59,1% and the average sentiments result is Neutral*[9]. Berdasarkan uraian diatas peneliti melakukan analisis sentimen terhadap penggunaan aplikasi PeduliLindungi dengan menggunakan data yang diperoleh dari komentar pengguna di *playstore*. Data yang digunakan sebanyak 321 data yang dikelompokkan kedalam kelas positif dan negatif. Penelitian ini membuat model klasifikasi dengan algoritma *Naive Bayes*[10]. Masyarakat Indonesia memiliki berbagai opini terhadap aplikasi PeduliLindungi, baik opini yang bersifat negatif maupun positif. Opini-opini mengenai aplikasi PeduliLindungi banyak disampaikan melalui media sosial Twitter, yang menjadi aplikasi dengan pengguna harian aktif terbanyak di Indonesia. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat Indonesia mengenai aplikasi PeduliLindungi dari data yang bersumber dari media sosial Twitter dan menentukan apakah sentimen masyarakat terhadap aplikasi ini bersifat *negatif* atau *positif* menggunakan algoritma SVM[11]. kemudian dideteksi apakah *sentiment* tersebut termasuk ke dalam *sentiment positif*, *sentiment negative*, atau *sentiment netral*. Salah satu teknik yang digunakan untuk menggali informasi mendeteksi opini masyarakat adalah dengan menggunakan analisis sentiment.

Penelitian ini terdiri dari beberapa proses untuk melakukan analisis *sentiment*, yaitu mengumpulkan data, melakukan tahapan *pre-processing*, pembobotan kata (ekstraksi fitur), dan proses klasifikasi *sentiment*[12]. Penelitian ini bertujuan membangun sistem yang dapat melakukan analisis sentimen, sehingga ulasan dari masyarakat dapat lebih mudah disaring oleh pihak yang berkepentingan. Analisis *sentiment* akan diimplementasikan menggunakan algoritma *naive bayes classifier*[13]. Aplikasi PeduliLindungi yang diunduh melalui aplikasi play store yang memberikan opini dari berbagai individu masyarakat. Opini tersebut penulis rangkum untuk melakukan penelitian dengan

pengklasifikasian teks review, dimana jumlah data yang dirangkum sebanyak 200 data review, terdiri dari 100 data review positif dan 100 data review negatif, dimana sentiment tersebut berhubungan dengan kalimat: bagus, cepat, kecewa, bodoh, tidak layak[14]. *The data collected were manually labeling into positive and negative class and processed using sentiment analysis with Naive Bayes algorithm, give the result 64.69% positive sentiment and 35.5% negative sentiment regarding PeduliLindungi. The model tested using Naive Bayes algorithm with 10-fold cross validation has the highest performance, the accuracy obtained[15].*

## II. METODE PENELITIAN

Pada bagian ini akan dijelaskan mengenai perancangan sistem dalam penelitian ini. Metode dalam penelitian ini dilakukan melalui beberapa tahap. Secara garis besar, alur penelitian dapat dijelaskan dibawah ini:

### A. Analisis Sentiment

*Analisis sentiment* merupakan proses menentukan opini seseorang yang diwujudkan dalam bentuk teks dan dikategorikan menjadi sentiment positif atau negatif[4]. Pengguna internet banyak menuliskan opini dan segala hal yang menjadi perhatian mereka. Opini tentang apa yang mereka rasakan ini dapat berupa perasaan positif, netral ataupun negatif yang dapat diungkapkan. Preprocessing merupakan tahap persiapan data yang bertujuan agar mempermudah proses pengolahan data. Preprocessing memfokuskan pada data cleaning & cleansing, termasuk menghilangkan noise di data, mengatasi struktur data yang tidak baik, dan informasi yang hilang. Berikut tahapan yang ada dalam preprocessing data:

- Cleansing, yaitu membersihkan data dari noise seperti hashtag, username, url, dan tanda baca.
- Case folding, merupakan tahap mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar yang konsisten secara keseluruhan (dalam hal ini huruf kecil).
- Menghapus Stopword, merupakan tahap untuk membuang kata-kata yang tidak penting seperti “yang”, “di”, “ke” dan seterusnya.
- Stemming, merupakan tahap untuk merubah kata-kata dalam kalimat menjadi kata dasar.
- Tokenisasi, merupakan proses memecah kalimat menjadi kata-kata.
- Filtering, merupakan tahap mengambil kata-kata penting dari hasil token dengan cara membuang kata-kata yang tidak penting.

### B. Naive Bayes

*Naive Bayes Classifier (NBC)* adalah algoritma klasifikasi yang berakar pada *teorema Bayes*. *Naive Bayes Classifier* bekerja sangat baik dibandingkan dengan model *classifier* lain seperti *Decision Tree* atau *Neural Network*. Keuntungan menggunakan metode ini adalah metode ini hanya membutuhkan sedikit data latih untuk menentukan parameter yang dibutuhkan dalam proses klasifikasi[16].

### C. Support Vector Machine (SVM)

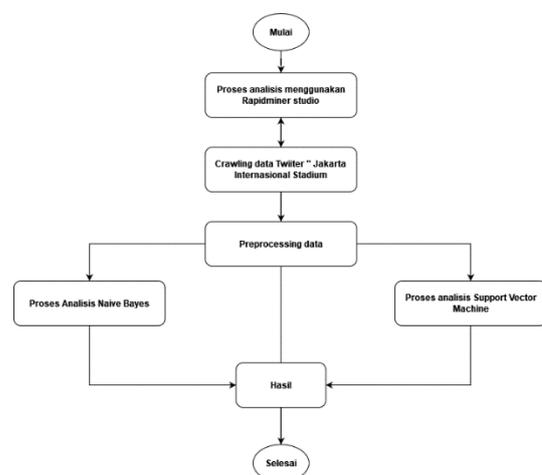
Support Vector Machine (SVM) adalah metode pembelajaran supervised yang menganalisis data dan mengenali pola untuk klasifikasi dan regensi, SVM bekerja menggunakan cara mencari hyperplane terbaik dengan memaksimalkan jarak antar kelas, hyperplane merupakan sebuah fungsi yang digunakan untuk pemisah antar kelas. SVM merupakan sebuah pengklasifikasi maka diberi suatu himpunan pelatihan yang ditandai sebagai milik salah satu dari kelas kategori, algoritma SVM membangun sebuah model yang memprediksi apakah data yang baru diproses merupakan salah satu dalam kategori yang lain[4].

### D. Text Mining

Text Mining dimanfaatkan untuk pengelolaan dokumen yang bermanfaat dari berbagai data dimana sumber datanya yaitu berbentuk teks, yang memiliki format tidak terstruktur. Tahapan pada text mining dalam preprocessing data bertujuan untuk mencari kata yang mewakili isi dari dokumen sehingga mampu melakukan analisis terhubung antar dokumen[14].

### E. Alur Penelitian

Penelitian ini diawali dengan serangkaian tahapan umum dalam penelitian yang terdiri dari tahapan studi literatur untuk dijadikan referensi atau rujukan dalam penelitian kali ini guna memperkuat permasalahan serta sebagai dasar teori dalam melakukan penelitian. Studi literatur didapatkan dari membaca jurnal, artikel dan situs-situs di internet sehingga mendapatkan kumpulan referensi yang relevan dengan masalah dalam penelitian ini. Setelah melakukan studi literatur, peneliti melakukan identifikasi masalah yang akan di teliti untuk menjadi objek penelitian. Hasil atau output yang telah didapat digunakan untuk menarik kesimpulan dari hasil penelitian yang dilakukan. Alur penelitian dapat dilihat pada Gambar 1 dibawah ini:



Gambar 1. Alur Penelitian

### F. Metode Pengumpulan Data

Proses pengumpulan data pada penelitian ini akan menggunakan data set publik. Data set publik didapat dengan Crawling Data Tweet para pengguna Twitter.

### G. Kelas Penelitian

Pada kelas penelitian ini peneliti membuat sebuah sistem untuk melakukan sentimen analisis, adapun dalam

penelitian ini menggunakan algoritma Naïve bayes dan Support Vector Machine. Untuk mempermudah dalam penelitian kali ini, peneliti menggunakan program RapidMiner Studio untuk melakukan Sentimen Analisis. Hasil sentimen analisis akan dibagi 2 yaitu Sentimen Positif dan Sentimen Negatif. Contoh Sentimen Positif dan negatif yang digunakan dalam kelas penelitian ini:

Tabel 1. Kelas Penelitian

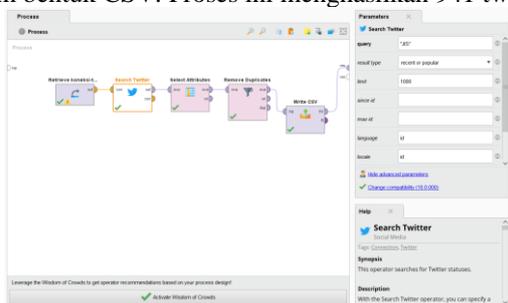
Kelas	Tweet
Sentimen Positif	 <p>(1)</p>
Sentimen Negatif	 <p>(2)</p>

**H. Penerapan Metodologi**

Metodologi yang digunakan dalam penelitian ini menggunakan Metode klasifikasi *naive bayes* dan *Support Vector Machine*. Software yang digunakan dalam melakukan sentimen analisis adalah RapidMiner Studio.

**I. Proses Pengumpulan Data**

Data dikumpulkan dari platform media sosial Twitter dengan menggunakan alat pengumpul data yang disebut API. Data yang dikumpulkan meliputi komentar pengguna pada Stadion Jakarta Stadium Internasional belum Standar FIFA yang berbeda. Data yang diambil adalah 1000 komentar yang diambil dari Tweet Jakarta Stadium Internasional belum Standar FIFA populer. Proses pengumpulan data pada penelitian ini akan menggunakan data set publik. Data set publik didapat dari Tweet para pengguna Twitter dengan menggunakan Operator “Search Twitter” yang ada pada RapidMiner Studio dengan kata kunci “JIS” sebanyak 1000 data yang selanjutnya di filter untuk membuang tweet duplikasi, lalu data di ubah kedalam bentuk CSV. Proses ini menghasilkan 941 tweet.



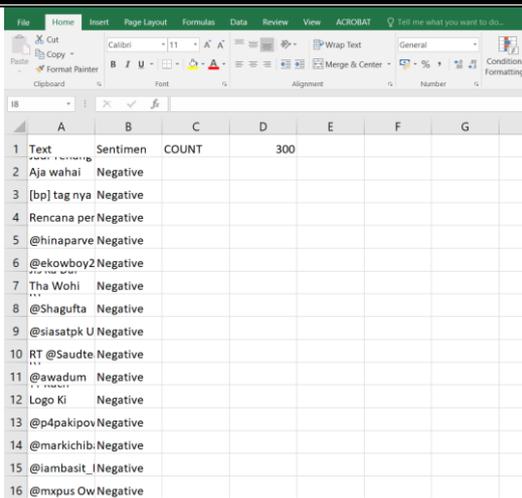
Gambar 2. Proses Pengumpulan Data

a) Retrieve koneksi Twitter pada operator RapidMiner adalah salah satu fitur yang memungkinkan pengguna untuk mengambil data dari Twitter dan melakukan analisis data menggunakan platform RapidMiner. Fitur ini memungkinkan pengguna untuk mengakses data publik di Twitter, seperti tweet, pengguna, dan topik terkait, dan menggunakan data tersebut untuk melakukan analisis data. Untuk menggunakan fitur retrieve koneksi Twitter pada RapidMiner, pengguna

- perlu menghubungkan akun Twitter mereka dengan platform RapidMiner dan mendapatkan akses token untuk mengakses API Twitter. Setelah itu, pengguna dapat mengambil data yang diinginkan dari Twitter menggunakan operator Retrieve Twitter. Operator ini memungkinkan pengguna untuk mengatur parameter pencarian seperti kata kunci, tanggal, lokasi, pengguna tertentu, dan banyak lagi, sehingga pengguna dapat memilih data yang ingin diambil dengan lebih spesifik.
- b) Operator RapidMiner Search Twitter juga memungkinkan pengguna untuk melakukan filter terhadap tweet berdasarkan tanggal, lokasi, jumlah retweet, dan follower. Setelah tweet diambil, pengguna dapat melakukan preprocessing pada data, seperti membersihkan tweet dari karakter yang tidak diinginkan, mengubah huruf kecil menjadi huruf besar, dan menghilangkan stopwords.
- c) Operator RapidMiner Select Attributes adalah operator yang digunakan untuk memilih atribut atau variabel tertentu dari dataset. Dalam pengolahan data, terkadang tidak semua atribut atau variabel pada dataset diperlukan dalam analisis atau pemodelan. Operator Select Attributes membantu dalam mengurangi dimensi data atau feature selection, sehingga dapat meningkatkan performa analisis dan pemodelan data.
- d) Operator RapidMiner "Remove Duplicates" digunakan untuk menghapus baris data duplikat dari dataset yang diberikan. Operator ini sangat berguna ketika peneliti memiliki data yang besar dan ingin memastikan bahwa tidak ada duplikasi dalam dataset Anda. Dengan menggunakan operator "Remove Duplicates", Peneliti dapat menghapus baris data yang identik dari dataset dengan mudah dan cepat.
- e) Operator RapidMiner "Write CSV" adalah operator yang digunakan untuk menulis data hasil pengolahan RapidMiner ke dalam format file CSV (Comma Separated Values). CSV adalah format file yang umum digunakan untuk menyimpan data tabel dalam bentuk teks, di mana setiap baris dalam file mewakili sebuah baris dalam tabel, dan kolom-kolom dipisahkan oleh tanda koma.

**J. Pelabelan Dataset**

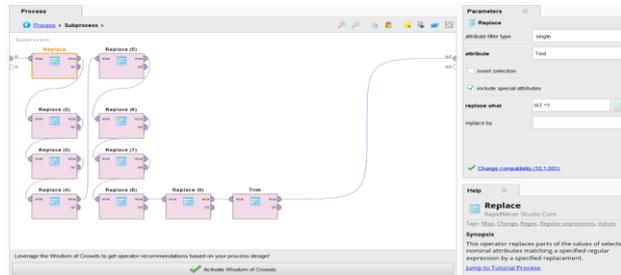
Sebanyak 940 data Pada Gambar 3 akan dilabeli secara manual yang nantinya akan digunakan sebagai data latih. Untuk mengurangi penilaian secara subyektif, proses pelabelan dilakukan dengan bantuan istri penulis. Pelabelan data dibagi menjadi 2 label, yaitu label “Positif” dan label “Negatif”.



	A	B	C	D	E	F	G
1	Text	Sentimen	COUNT	300			
2	Aja wahai	Negative					
3	[bp] tag nya	Negative					
4	Rencana per	Negative					
5	@hinaparve	Negative					
6	@ekowboy2	Negative					
7	Tha Wohi	Negative					
8	@Shagufra	Negative					
9	@siasatpk U	Negative					
10	RT @Saudte	Negative					
11	@awadum	Negative					
12	Logo Ki	Negative					
13	@p4pakipov	Negative					
14	@markichib	Negative					
15	@lambasit U	Negative					
16	@mxps Ow	Negative					

Gambar 3. Tahap Pelabelan

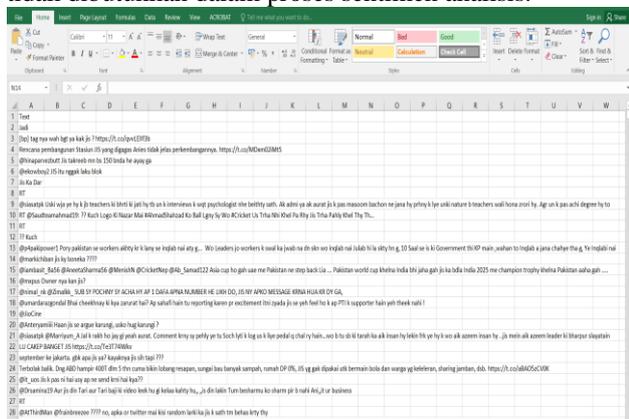
- dilakukan dengan mengatur kondisi filter untuk memeriksa setiap atribut dan mengeliminasi instance yang memiliki missing values.
- c) Operator Subprocess Fungsi komponen yang digunakan untuk mengorganisir dan membagi alur kerja (workflow) menjadi bagian-bagian yang lebih kecil dan terstruktur.



Gambar 6. Tahapan cleaning menggunakan Replace

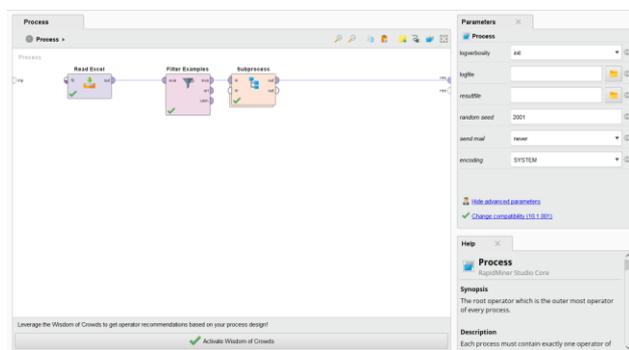
### K. Tahap Cleaning Data

Tahap ini bertujuan untuk membersihkan Tweet dari kata-kata yang tidak diperlukan seperti karakter hastag “#”, mention “@”, membuang url-url atau simbol simbol yang tidak dibutuhkan dalam proses sentimen analisis.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Text																						
2	Text																						
3	Text																						
4	Text																						
5	Text																						
6	Text																						
7	Text																						
8	Text																						
9	Text																						
10	Text																						
11	Text																						
12	Text																						
13	Text																						
14	Text																						
15	Text																						
16	Text																						
17	Text																						
18	Text																						
19	Text																						
20	Text																						
21	Text																						
22	Text																						
23	Text																						
24	Text																						
25	Text																						
26	Text																						
27	Text																						
28	Text																						
29	Text																						

Gambar 4. Tahap Sebelum Cleaning Data



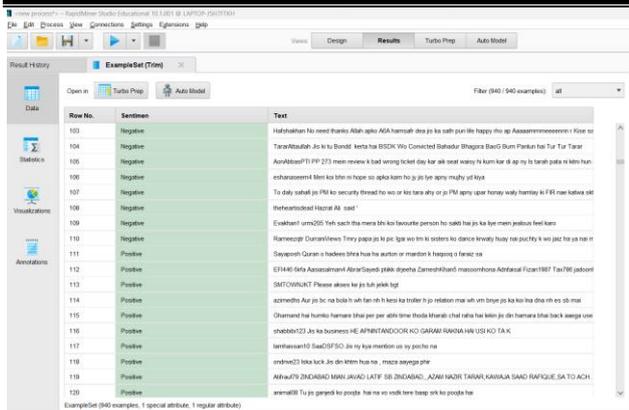
Gambar 5. Operator Cleaning Data

Keterangan Operator pada rapidminer sendiri pada gambar 5 Sebagai Berikut:

- a) Operator Read Excel fungsi nya mengimport data uji yang sudah dilakukan pelabelan
- b) Operator Filter examples untuk menghapus baris atau instance yang memiliki missing values. Ini dapat

Dalam proses ini, penulis akan menggunakan beberapa Operator Replace dan Trim untuk membersihkan Tweet dari kata kata yang tidak diperlukan pada gambar 6 Menggunakan Operator Subprocess dan Replace Sebagai Berikut:

- a) Replace pertama berfungsi untuk menghapus ReTweet yang berada pada depan atau tengah kalimat, masukkan “RT.\*?” pada Parameters.
- b) Replace kedua berfungsi untuk menghapus Mention yang berada pada depan atau tengah kalimat, masukkan “@.\*?” pada Parameters.
- c) Replace ketiga berfungsi untuk menghapus Hastag yang berada pada depan atau tengah kalimat, masukkan “#.\*?” pada Parameters.
- d) Replace keempat berfungsi untuk menghapus Url yang berada pada depan atau tengah kalimat, masukkan “http.\*?” pada Parameters.
- e) Replace kelima berfungsi untuk menghapus ReTweet yang berada pada belakang kalimat, masukkan “RT.\*” pada Parameters.
- f) Replace keenam berfungsi untuk menghapus Mention yang berada pada belakang kalimat, masukkan “@.\*” pada Parameters.
- g) Replace ketujuh berfungsi untuk menghapus Hastag yang berada pada belakang kalimat, masukkan “#.\*” pada Parameters.
- h) Replace kedelapan berfungsi untuk menghapus Url yang berada pada belakang kalimat, masukkan “http.\*” pada Parameters.
- i) Replace kesembilan berfungsi untuk menghapus Simbol yang tidak dibutuhkan dalam kalimat, masukkan “[!]"#\$%&'()\*+/,;:[]?@\|\_{}~]" pada Parameters.
- j) Trims Salah satu operasi yang digunakan untuk menghapus karakter atau spasi ekstra pada awal dan akhir teks dalam atribut. Fungsinya adalah untuk membersihkan dan memformat teks dengan menghilangkan karakter yang tidak diinginkan.



Gambar 7. Setelah Proses Cleaning Data

**L. Tahap Pre-Processing Data**

Tahap Preprocessing data bertujuan untuk mengubah data mentah menjadi data yang siap diolah, adapula tahap ini dibagi menjadi beberapa bagian, yaitu :

- a) Tokenizing, yaitu tahapan untuk membagi teks menjadi kata, seperti teks “sedang mengerjakan skripsi” setelah melewati tahap Tokenizing, akan menjadi 4 kata, yaitu “sedang”, “mengerjakan”, dan “skripsi”. Case Folding, yaitu tahapan untuk mengubah data tweet menjadi lower case (huruf kecil).
- b) Case Folding, yaitu tahapan untuk mengubah data tweet menjadi lower case (huruf kecil).
- c) Stemming, yaitu tahapan untuk membersihkan kata-kata imbuhan awalan dan akhiran yang terdapat dalam teks seperti “mengerjakan” menjadi “kerja”.
- d) Filter Tokens (by Length), yaitu tahapan untuk membuang kata-kata yang kurang dari 2 huruf dan kata-kata yang melebihi 25 huruf.
- e) Filter Stopwords, yaitu tahapan untuk menghapus kata bantu seperti “saya”, “dia”, “aku”, dan “mereka”.



Gambar 8. Tahap Pre-Processing Data

**M. Tahap Pembobotan Kata**

Pada tahap ini, hasil preprocessing akan diolah agar setiap kata memiliki bobot (nilai). Pembobotan kata yang penulis gunakan adalah algoritma TF-IDF. Term Frequency-Inverse Document Frequency atau TF-IDF adalah suatu metode algoritma yang berguna untuk menghitung bobot setiap kata yang umum digunakan. Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat. Metode ini akan menghitung nilai Term Frequency (TF) dan Inverse Document Frequency (IDF) pada setiap token (kata) di setiap dokumen dalam korpus. Secara sederhana, metode TF-IDF digunakan untuk

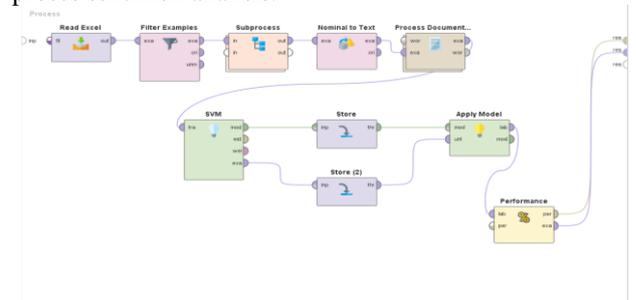
mengetahui berapa sering suatu kata muncul di dalam dokumen.

Row No.	Sentimen	aaaa	aaaaaaa	aaaaabbbbbb	aaaammme	aaap	aaarrghhhhh	aaabmmgami	aaabarrarrah	aaadi
1	Negative	0	0	0	0	0	0	0	0	0
2	Negative	0	0	0	0	0	0	0	0	0
3	Negative	0	0	0	0	0	0	0	0	0
4	Negative	0	0	0	0	0	0	0	0	0
5	Negative	0	0	0	0	0	0	0	0	0
6	Negative	0	0	0	0	0	0	0	0	0
7	Negative	0	0	0	0	0	0	0	0	0
8	Negative	0	0	0	0	0	0	0	0	0
9	Negative	0	0	0	0	0	0	0	0	0
10	Negative	0	0	0	0	0	0	0	0	0
11	Negative	0	0	0	0	0	0	0	0	0
12	Negative	0	0	0	0	0	0	0	0	0
13	Negative	0	0	0	0	0	0	0	0	0
14	Negative	0	0	0	0	0	0	0	0	0
15	Negative	0	0	0	0	0	0	0	0	0
16	Negative	0	0	0	0	0	0	0	0	0
17	Negative	0	0	0	0	0	0	0	0	0

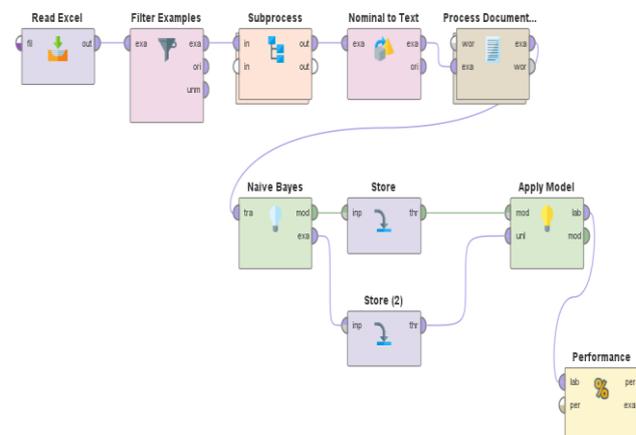
Gambar 9. Hasil Pembobotan Kata dengan TF-IDF

**N. Tahap Pembuatan Model**

Output dari tahapan ini adalah sebuah model klasifikasi dengan metode *Naive Bayes* dan *Support Vector Machine* dan Data Latih yang akan digunakan dalam proses sentimen analisis.



Gambar 10. Pembuatan Model Support Vector Machine

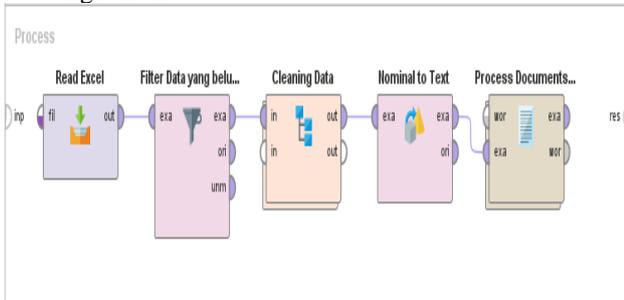


Gambar 11. Pembuatan Model Support Vector Machine

**O. Tahap Mempersiapkan Data Uji**

Pada tahap ini data uji akan diolah kembali agar dapat dilakukan proses selanjutnya, tahapan mengolah data uji antara lain :

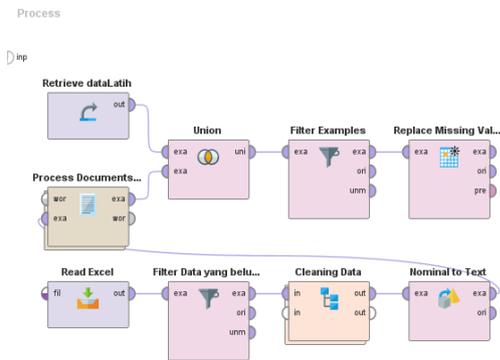
- a) Filter data yang belum memiliki label
- b) Cleaning data untuk menghilangkan kata kata yang tidak dibutuhkan
- c) Preprocessing data dan pembobotan kata dengan algoritma TF-IDF



Gambar 12. Pembuatan Model *Support Vector Machine*

**P. Tahap Union / Pegabungan Data**

Pada tahap ini, data uji dan data latih akan digabungkan menjadi satu data

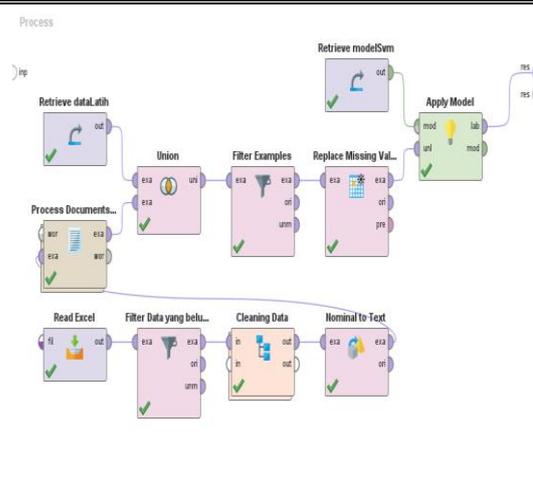


Gambar 13. Pembuatan Model *Support Vector Machine*

**III. HASIL DAN PEMBAHASAN**

**A. Tahap Pengujian**

Pada tahap ini, model yang telah dibuat sebelumnya akan diterapkan untuk memprediksi sentimen pada data uji.



Gambar14. Pembuatan Model *Support Vector Machine*

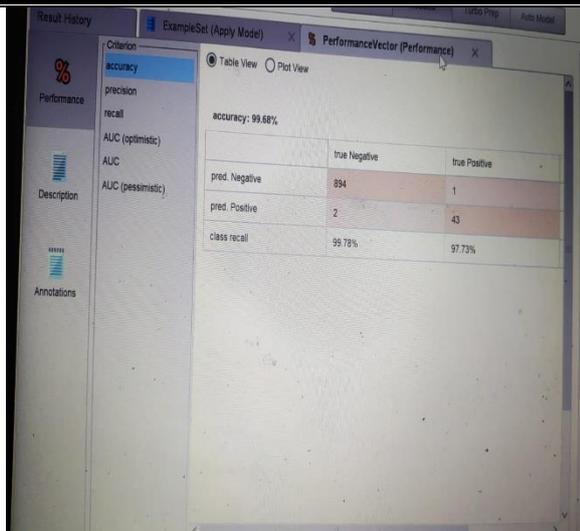
**B. Nilai Akurasi *Support Vector Machine* dan *Naive Bayes***

Hasil perhitungan akurasi data latih dengan menggunakan metode *Support Vector Machine*, didapatkan nilai Accuracy sebesar 99.68, nilai Recall Positif sebesar 97.73%, nilai Recall Negatif sebesar 99.78%, nilai Precision Positif sebesar 100.00%, nilai Precision Negatif sebesar 97.71%. Dari 940 data latih, penulis melabeli sebanyak 46 data sebagai Sentimen Positif dan 894 data sebagai Sentimen Negatif.

Criterion	Value
accuracy	99.68%
precision	
recall	
AUC (optimistic)	
AUC (pessimistic)	
pred. Negative	894
pred. Positive	43
class recall	99.78%

Gambar 15. Hasil Pengujian *Support Vector Machine*

Hasil perhitungan akurasi data latih dengan menggunakan metode *Naive Bayes*, didapatkan nilai Accuracy sebesar 99.57% nilai Recall Positif sebesar 100.00%, nilai Recall Negatif sebesar 99.55%, nilai Precision Positif sebesar 80.00%, nilai Precision Negatif sebesar 100.00%. Dari 940 data latih, penulis melabeli sebanyak 48 data sebagai Sentimen Positif dan 892 data sebagai Sentimen Negatif.



Gambar 16. Hasil Pengujian *Naive Bayes*

### C. Perbandingan Hasil Akurasi Algoritama *Naive Bayes* dan *Support Vector Machine*

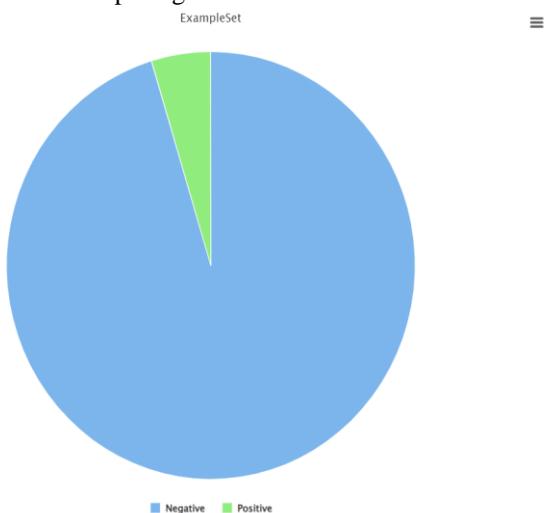
Hasil dari Implementasi yang telah dilakukan, perbandingan tingkat akurasi antara metode *Naive Bayes* dan *Support Vector Machine*.

Tabel 2. Perbandingan Akurasi

Metode	Nilai Akurasi
<i>Naive Bayes</i>	99.57%
<i>Support Vector Machine</i>	99.68%

### D. Hasil Sentimen Analisis

Dari proses sentimen analisis dengan 940 data uji, dihasilkan prediksi sebanyak 896 Sentimen Negatif dan 44 Sentimen Positif, Untuk hasil prediksi dari Sentimen Negatif. Terdapat 940 data terprediksi Negatif dan 1 data yang terprediksi Positif, berikut penulis sajikan data tersebut kedalam Pie Chart dan Wordcloud, berikut data sentiment analis pada gambar 17 dibawah ini:



Gambar 17. Hasil Sentimen Analisis

Hasil akhir dari Perbandingan dengan dua metode pengujian ini, yaitu hasil prediksi Sentimen Masyarakat Terhadap Isu Stadion Jakarta Internasional stadium belum

berstandara FIFA berdasarkan data yang didapat dari Twitter dan diimplementasikan dengan metode SVM (*Support Vector Machine*) menunjukkan nilai akurasi sebesar 99.68%. Dari 940 data uji, terprediksi sebesar 894 data sebagai Sentimen Negatif dan 46 data sebagai Sentimen Positif Untuk hasil prediksi dari Sentimen Negatif, terdapat 892 data terprediksi Negatif dan Metode *Naive bayes* menunjukkan nilai akurasi sebesar 99.57%. Dari 940 data uji, terprediksi sebesar 892 data sebagai Sentimen Negatif dan 48 data sebagai Sentimen Positif Untuk hasil prediksi dari Sentimen Negatif.

### UCAPAN TERIMA KASIH

Saya ingin mengucapkan terima kasih yang sebesar-besarnya atas bimbingan, dukungan, dan dedikasi Bapak Rasiban dalam membantu pembuatan jurnal ini. Tanpa panduan dan dorongan yang berharga dari Anda, pencapaian ini tidak akan mungkin terwujud. Saya sangat berterimakasih atas waktu dan pengetahuan yang Anda luangkan untuk membimbing saya dalam menyusun jurnal ini. Dalam setiap pertemuan dan diskusi, Bapak telah memberikan wawasan yang berharga, saran yang berharga, dan arahan yang tepat. Bimbingan Bapak telah memainkan peran penting dalam membantu saya memahami proses penelitian, mengembangkan metodologi yang tepat, dan menganalisis data dengan benar. Selain itu, terima kasih juga karena telah memberikan koreksi dan umpan balik konstruktif dalam setiap tahap penulisan jurnal. Pengamatan dan penilaian Bapak yang mendalam telah membantu saya untuk meningkatkan kualitas tulisan dan memperbaiki kelemahan yang ada. Saya sangat menghargai kesabaran dan ketelitian yang Anda tunjukkan dalam membimbing saya menuju hasil akhir yang memuaskan. Lebih dari sekadar menjadi pembimbing, Bapak juga menjadi inspirasi bagi saya. Kepedulian dan semangat Bapak terhadap penelitian dan pengembangan ilmu pengetahuan telah mendorong saya untuk terus belajar dan berkembang. Saya merasa sangat beruntung dan bersyukur dapat belajar dari Bapak, seorang akademisi yang berkompeten dan berdedikasi. Terima kasih karena telah memberikan kesempatan kepada saya untuk terlibat dalam penelitian ini dan memperluas pengetahuan serta wawasan saya. Saya meyakini bahwa karya tulis ini tidak hanya akan memberikan manfaat bagi saya secara pribadi, tetapi juga akan memberikan kontribusi yang berarti dalam perkembangan bidang ilmu yang kita geluti.

### IV. REFERENSI

- [1] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode *Random Forest*," vol. 6, no. 9, pp. 4305–4313, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [2] Christian, H. Wijangga, and B. Christian, "Analisa Faktor Minat Penggunaan Aplikasi Pedulilindungi Anak Muda Surabaya Dengan Metode Tam,"



- ZONasi: *Jurnal Sistem Informasi*, vol. 4, no. 2, pp. 1–10, 2022, doi: 10.31849/zn.v4i2.10078.
- [3] I. R. Putra, Y. Widiastiwi, and N. Chamidah, “Pengaruh Seleksi Fitur Particle Swarm Optimization terhadap Sentimen Analisis Aplikasi Pedulilindungi di Twitter dengan Algoritma Support Vector Machine,” *Informatik: Jurnal Ilmu Komputer*, vol. 18, no. 3, p. 208, 2022, doi: 10.52958/iftk.v18i3.4681.
- [4] A. Noviriandini, H. Hermanto, and Y. Yudhistira, “Klasifikasi Support Vector Machine Berbasis Particle Swarm Optimization Untuk Analisa Sentimen Pengguna Aplikasi Pedulilindungi,” *JIKA (Jurnal Informatika)*, vol. 6, no. 1, p. 50, 2022, doi: 10.31000/jika.v6i1.5681.
- [5] C. Chairunnisa, I. Ernawati, and M. M. Santoni, “Klasifikasi Sentimen Ulasan Pengguna Aplikasi Pedulilindungi di Google Play Menggunakan Algoritma Support Vector Machine dengan Seleksi Fitur Chi-Square,” *Informatik: Jurnal Ilmu Komputer*, vol. 18, no. 1, p. 69, 2022, doi: 10.52958/iftk.v17i4.4594.
- [6] G. A. Lustiansyah *et al.*, “Analisis klasifikasi sentimen pengguna aplikasi pedulilindungi berdasarkan ulasan dengan menggunakan metode long short term memory,” *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, pp. 327–336, 2022.
- [7] H. P. Doloksaribu and Yusran Timur Samuel, “Komparasi Algoritma Data Mining Untuk Analisis Sentimen Aplikasi Pedulilindungi,” *Jurnal Teknologi Informasi: Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika*, vol. 16, no. 1, pp. 1–11, 2022, doi: 10.47111/jti.v16i1.3747.
- [8] A. Habib Husaini, R. Mayasari, and U. Singaperbangsa Karawang, “Pengelompokan Ulasan Aplikasi Pedulilindungi Dengan Algoritma K-Medoids Pedulilindungi Application Review Grouping With the K-Medoids Algorithm,” *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 5, no. 2, 2022.
- [9] A. Salma and W. Silfianti, “Sentiment Analysis of User Review on COVID-19 Information Applications Using Naïve Bayes Classifier, Support Vector Machine, and K-Nearest Neighbors,” *International Research Journal of Advanced Engineering and Science*, vol. 6, no. 4, pp. 158–162, 2021.
- [10] A. P. Wibowo, W. Darmawan, and N. Amalia, “Komparasi Metode Naïve Bayes Dan K-Nearest Neighbor Terhadap Analisis Sentimen Pengguna Aplikasi Pedulilindungi,” *IC-Tech*, vol. 17, no. 1, pp. 18–23, 2022, doi: 10.47775/icttech.v17i1.234.
- [11] R. A. Rizka Akmalia, I. Slamet, and H. Pratiwi, “Analisis Sentimen Twitter Berbahasa Indonesia Terhadap Aplikasi Pedulilindungi dengan Algoritma SVM, KNN, dan Regresi Logistik,” *Prosiding Seminar Nasional MIPA UNIPA*, vol. 2022, pp. 150–156, 2022, doi: 10.30862/psnm.v7i1.21.
- [12] N. Putriani, F. R. Umbara, and P. N. Sabrina, “Analisis Sentimen pada Aplikasi Pedulilindungi dengan Menggunakan Metode Improved K-Nearest Neighbor dan Lexicon Based,” vol. 8, no. 1, pp. 350–364, 2022.
- [13] G. K. Locarso, “Analisis Sentimen Review Aplikasi Pedulilindungi Pada Google Play Store Menggunakan NBC,” *Jurnal Teknik Informatika Kaputama (JTIK)*, vol. 6, no. 2, pp. 353–361, 2022.
- [14] P. Astuti and N. Nuris, “Penerapan Algoritma KNN Pada Analisis Sentimen Review Aplikasi Peduli Lindungi,” *Computer Science (CO-SCIENCE)*, vol. 2, no. 2, pp. 137–142, 2022, doi: 10.31294/coscience.v2i2.1258.
- [15] L. Ellyanti, Y. Ruldeviyani, L. E. Pradana, and A. Harjanto, “JURNAL RESTI,” vol. 5, no. 158, pp. 414–421, 2023.
- [16] R. Syahputra, G. J. Yanris, and D. Irmayani, “SVM and Naïve Bayes Algorithm Comparison for User Sentiment Analysis on Twitter,” *Sinkron*, vol. 7, no. 2, pp. 671–678, 2022, doi: 10.33395/sinkron.v7i2.11430.