



Analisis Sentimen Terhadap Aplikasi Satu Sehat Pada Twitter Menggunakan Algoritma Naive Bayes Dan Support Vector Machine

Francis Matheos Sarimole^{1*}, Kudrat²

^{1,2}Program Studi Sistem Informasi, Sekolah Tinggi Ilmu Komputer Cipta Karya informatika, Jakarta Timur, Indonesia

¹matheosfrancis.s@gmail.com, ²kudrattullah1@gmail.com

Abstrak— Perbandingan Klasifikasi dan Nilai Akurasi Sentimen Analisis yang dilakukan dengan Algoritma Naive Bayes dan Support Vector Machine dalam menganalisa label sentimen positif dan negative Dapat mengetahui sentimen masyarakat mengenai Informasi dan edukasi Aplikasi SATUSEHAT dengan lebih efisien dari segi waktu dan tenaga. Mengetahui faktor-faktor yang mempengaruhi nilai Akurasi dari Algoritma Naive Bayes dan Support Vector Machine dalam menganalisa label sentimen positif dan negatif. Pengambilan data dilakukan lewat media sosial Twitter dengan menggunakan API Twitter untuk pengambilan data untuk pengambilan data Twitter dengan kata kunci yang digunakan yaitu " SatuSehat" Total jumlah sampel yang digunakan dalam penelitian ini minimal 1046 data komentar atau lebih. Sampel data yang akan diambil memiliki rentang waktu setahun yaitu pada tahun 2023. Hasil akhir dari Perbandingan dengan dua metode pengujian ini, yaitu hasil prediksi Sentimen Masyarakat Terhadap Isu Aplikasi SatuSehat berdasarkan data yang didapat dari Twitter dan diimplementasikan dengan metode SVM (Support Vector Machine) menunjukkan nilai akurasi sebesar 87.95 %. Dari 1080 data uji, terprediksi 132 data sebagai Sentimen Positif dan 947 data sebagai Sentimen Negatif. Untuk hasil prediksi dari Sentimen Negatif, terdapat 1080 data terprediksi Negatif dan 1 data yang terprediksi Positif. dan Metode Naive bayes menunjukkan nilai akurasi sebesar 81.65%. Dari 1081 data uji, terprediksi sebesar 947 data sebagai Sentimen Negatif dan 132 data sebagai Sentimen Positif Untuk hasil prediksi dari Sentimen Negatif, terdapat 947 data terprediksi Negatif dan 1 data yang terprediksi Positif.

Kata Kunci: Sentimen, Twitter, SatuSehat, NB, SVM, Rapidminer

Abstract— Comparison of Classification and Accuracy Values of Sentiment Analysis carried out using the Naive Bayes Algorithm and Support Vector Machine in analyzing positive and negative sentiment labels Can find out public sentiment regarding SATUSEHAT application information and education more efficiently in terms of time and energy. Know the factors that influence the accuracy value of the Naive Bayes Algorithm and Support Vector Machine in analyzing positive and negative sentiment labels. Data collection is done via social media Twitter by using the Twitter API to retrieve data crawling to retrieve Twitter data with the keyword used, namely "SatuSehat." The total number of samples used in this research is a minimum of 1046 comment data or more. The data samples to be taken have a time span of a year in 2023. Results The end of the comparison with these two testing methods, namely the results of predicting Community Sentiment towards the SatuSehat Application Issue based on data obtained from Twitter and implemented using the SVM (Support Vector Machine) method shows an accuracy value of 87.95%. Of the 1080 test data, 132 data were predicted as Positive Sentiment and 947 data as Negative Sentiment. For prediction results from Negative Sentiment, there are 1080 data predicted to be Negative and 1 data predicted to be Positive. and the Naive Bayes method shows an accuracy value of 81.65%. Of the 1081 test data, 947 data were predicted as Negative Sentiment and 132 data were Positive Sentiment. For the prediction results from Negative Sentiment, there were 947 data predicted as Negative and 1 data predicted as Positive

Keywords: Sentiment, Twitter, One Healthy, NB, SVM, Rapidminer

I. PENDAHULUAN

SATUSEHAT sebagai nama Indonesia Health Services yang merupakan platform integrasi dan standarisasi layanan data kesehatan di Indonesia, Menjadi salah satu upaya Kemenkes RI dalam mentransformasikan layanan kesehatan melalui digitalisasi, SATUSEHAT merupakan sebuah platform konektivitas data, analisis, dan layanan untuk mendukung integrasi antar aplikasi dan fasilitas pelayanan kesehatan (*fasyankes*). Nantinya, pasien tidak perlu berulang-ulang mengisi formulir baru saat berpindah *fasyankes*.

Melalui SATUSEHAT, pasien juga bisa mendapatkan informasi mengenai kondisi kesehatannya secara lebih transparan. Karena resume rekam medis di rumah sakit dicatat dan direkam secara digital dengan aman melalui persetujuan (consent) pemilik data. Hal ini sejalan dengan rencana Kemenkes RI dalam mentransformasikan PeduliLindungi menjadi aplikasi kesehatan masyarakat.

Jadi manfaatnya akan semakin berkembang dari yang semula untuk penanganan pandemi bertransformasi menjadi lebih luas lagi.

Adapun nama SATUSEHAT dipilih berdasarkan hasil sayembara yang pendaftarannya dibuka kepada publik sejak 6 Juli lalu. Dari sayembara tersebut, terpilih tiga pemenang utama. Karya nama dari pemenang utama tersebut yang akhirnya ditetapkan dan diumumkan sebagai nama resmi pengganti IHS.

Berdasarkan pencarian di media sosial *Twitter*, saya mengidentifikasi beberapa masalah yang sering disorot oleh masyarakat terkait aplikasi SATUSEHAT antara lain Tweet Pengguna Kesulitan dalam mengakses aplikasi, Banyak pengguna mengeluhkan kesulitan dalam mengakses aplikasi SATUSEHAT. Beberapa di antaranya mengalami kendala teknis seperti error atau crash ketika membuka aplikasi, sedangkan yang lain mengalami kesulitan saat mendaftar dan login ke dalam aplikasi, Lambatnya respon dari petugas kesehatan: Beberapa



pengguna melaporkan bahwa mereka mengalami kesulitan dalam mendapatkan respons dari petugas kesehatan setelah mengirimkan laporan melalui aplikasi SATUSEHAT. Hal ini membuat banyak pengguna merasa tidak nyaman dan khawatir tentang kesehatan mereka dan Kurangnya informasi dan edukasi Beberapa pengguna menganggap bahwa informasi dan edukasi yang disediakan oleh aplikasi SATUSEHAT masih kurang. Mereka mengharapkan adanya penjelasan yang lebih detail tentang gejala-gejala penyakit, cara penanganan, dan tindakan pencegahan yang tepat.

Analisis sentimen merupakan kegiatan untuk mencari Isu masyarakat tentang sebuah objek yang ingin di ketahui. Penelitian ini dilakukan untuk mengetahui pendapat masyarakat terhadap Isu Aplikasi SATUSEHAT pada media sosial *Twitter*. Salah satu algoritma yang dapat digunakan untuk melakukan analisis sentimen adalah *Naive Bayes* dan *Support Vector Machine*, Pada penelitian Penulis Bertujuan menggunakan Algoritma *Naive Bayes* dan *Support Vector Machine* dalam menganalisa label sentimen positif dan negatif.

Dengan aplikasi peduli lindungi dapat membantu pemerintah untuk memantau masyarakat yang sudah vaksin atau belum Tujuan penelitian ini yaitu untuk mengujinya factor – factor yang mempengaruhi penerimaan aplikasi peduli lindungi untuk anak mudah (17 tahun – 25 tahun) terhadap penggunaan aplikasi peduli lindungi. Selanjutnya untuk untuk menguji tingkat tingkat penerimaan aplikasi PeduliLindungi untuk anak muda (17 tahun – 25 tahun) 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 kegunaanya 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. Penglompokan ulasan dapat menggunakan teknik *text mining* atau penambahan *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. AppStore

AppStore merupakan aplikasi yang di kembangkan oleh *Apple Inc*. Aplikasi ini merupakan *marketplace* perangkat lunak yang di khususkan untuk pengguna *Apple*, *AppStore* memungkinkan pengguna untuk dapat membeli dan mengunduh aplikasi yang ada di *AppStore*[1].

C. 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].

D. Support Vector Machine (SVM)

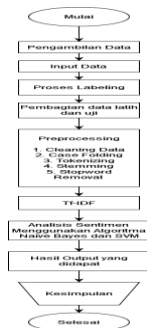
Support Vector Machine (SVM) adalah metode pembelajaran supervised yang menganalisis data dan mengenali pola untuk klasifikasi dan regensi, SVM bekerja menggunakan cara mencari *hyperplane* terbaik dengan memaksimalkan jarak antar kelas, *hyperplane* merupakan sebuah fungsi yang digunakan untuk pemisah 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].

E. Support Vector Machine (SVM)

Text Mining dimanfaatkan untuk pengolahan 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]

F. Alur Penelitian

Data latih yang digunakan sebanyak 1081 opini maupun komentar, sedangkan untuk data uji yang digunakan sebanyak 1081 opini maupun komentar dengan asumsi perbandingan data latih yang digunakan sebanyak 85% dan data uji sebanyak 15% dari total data yang digunakan dalam penelitian sebanyak 1081 data opini Twitter, setelah data dibagi menjadi dua data diolah menggunakan metode klasifikasi *Naive Bayes* dan *Support Vector Machine*, pengolahan data menggunakan metode klasifikasi *Naive Bayes* dan *Support Vector Machine* dilakukan untuk mendapatkan hasil atau output yang diinginkan. 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

G. Metode Pengumpulan Data

Pengambilan data dilakukan lewat media sosial Twitter dengan menggunakan API Twitter untuk pengambilan crawling data untuk pengambilan data Twitter dengan kata kunci yang digunakan yaitu " SatuSehat" Total jumlah sampel yang digunakan dalam penelitian ini minimal 1046 data komentar atau lebih. Sampel data yang akan diambil memiliki rentang waktu setahun yaitu pada tahun 2023.

H. Kelas Penelitian

Pengambilan data dilakukan lewat media sosial *Twitter* dengan menggunakan API Twitter untuk pengambilan *crawling* data untuk pengambilan data *Twitter* dengan kata kunci yang digunakan yaitu " SatuSehat" Total jumlah sampel yang digunakan dalam penelitian ini minimal 1046 data komentar atau lebih. Sampel data yang akan diambil memiliki rentang waktu setahun yaitu pada tahun 2023.

Tabel 1. Kelas Penelitian

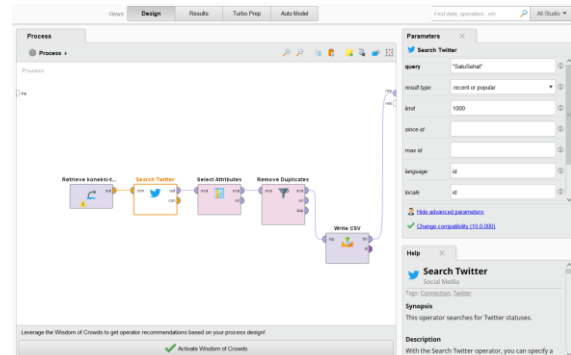
Kelas	Tweet
Sentimen Positif	
Sentimen Negatif	

I. 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.

J. Proses Pengumpulan Data

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 "SatuSehat" sebanyak 2000 data yang selanjutnya di filter untuk membuang tweet duplikasi, lalu data di ubah kedalam bentuk CSV.



Gambar 2. Alur Penelitian

- 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.

K. Pelabelan Dataset

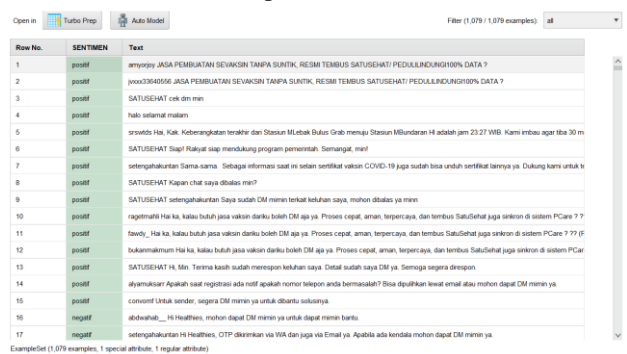
Sebanyak 940 data Pada Gambar 3 akan dilabeli secara manual yang nantinya akan digunakan sebagai data latih. Untuk mengurangi penilaian secara subyektif, proses pelabelan dilakukan dengan bantuan istri penulis. Pelabelan data dibagi menjadi 2 label, yaitu label "Positif" dan label "Negatif".

	A	B	C	D
	@SATUSEHAT Hi, Min. Terima kasih sudah merespon keluhan saya. Detail sudah saya DM ya. Semoga segera direspon.	positif		
14	@alyamuksarr Apakah saat registrasi ada notif apakah nomor telepon anda bermasalah? Bisa dipulihkan lewat email atau mohon dapat DM mimin ya.	positif		
15	@convomf Untuk sender, segera DM mimin ya untuk dibantu solusinya.	positif		
16	@abdwahab__ Hi Healthies, mohon dapat DM mimin ya untuk dapat mimin bantu.	negatif		
17	@setengahakuntan Hi Healthies, OTP dikirimkan via WA dan juga via Email ya. Apabila ada kendala mohon dapat DM mimin ya.	negatif		
18	@rizkyhanfh @ardeliana_ Hai ka, kalau butuh jasa vaksin dariku boleh DM aja	negatif		

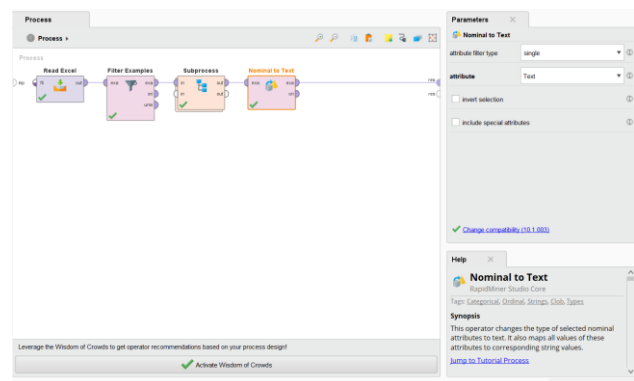
Gambar 3. Tahap Pelabelan

L. 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.



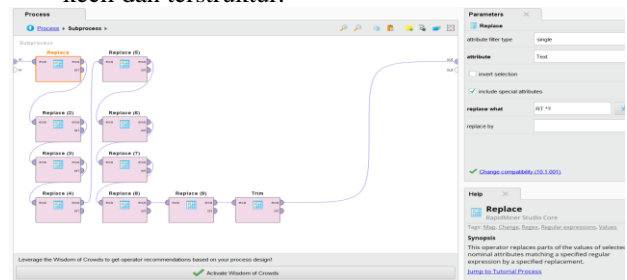
Gambar 4. Tahap Sebelum Cleaning Data



Gambar 5. Operator Cleaning Data

Keterangan Operator pada rapidminer studio 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 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

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:

- Replace pertama berfungsi untuk menghapus ReTweet yang berada pada depan atau tengah kalimat, masukkan “RT.*?” pada Parameters.
- Replace kedua berfungsi untuk menghapus Mention yang berada pada depan atau tengah kalimat, masukkan “@.*?” pada Parameters.
- Replace ketiga berfungsi untuk menghapus Hastag yang berada pada depan atau tengah kalimat, masukkan “#.*?” pada Parameters.
- Replace keempat berfungsi untuk menghapus Url yang berada pada depan atau tengah kalimat, masukkan “http.*?” pada Parameters.
- Replace kelima berfungsi untuk menghapus ReTweet yang berada pada belakang kalimat, masukkan “RT.*” pada Parameters.
- Replace keenam berfungsi untuk menghapus Mention yang berada pada belakang kalimat, masukkan “@.*” pada Parameters.
- Replace ketujuh berfungsi untuk menghapus Hastag yang berada pada belakang kalimat, masukkan “#.*” pada Parameters.
- Replace kedelapan berfungsi untuk menghapus Url yang berada pada belakang kalimat, masukkan “http.*” pada Parameters.
- Replace kesembilan berfungsi untuk menghapus Simbol yang tidak dibutuhkan dalam kalimat, masukkan “[!\"#\$%&'()*+/,.;:[]\`_{}~]” pada Parameters.
- 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.

Row No.	SENTIMEN	Text
1	positif	amnyoyy JASA PEMBUATAN SEWAKSI TANPA SUNTIK, RESMI TEMBUS SATUSEHAT! PEDULILINDUNG100% DATA ?
2	positif	juwa33640556 JASA PEMBUATAN SEWAKSI TANPA SUNTIK, RESMI TEMBUS SATUSEHAT! PEDULILINDUNG100% DATA ?
3	positif	SATUSEHAT cek drn min
4	positif	halo selamat malam
5	positif	stradid Hia, Kak. Keberangkatan terakhir dari Stasiun Mutebak Buluk menuju Stasiun MBundaran H adalah jam 23:27 WIB. Kamiimbau agar bisa 30 m
6	positif	SATUSEHAT Siap! Rakyat siap mendukung program pemerintah. Semangat, min!
7	positif	setengahakuntan Sama-sama. Sebagai informasi saat ini seban sertifikat vaksin COVID-19 juga sudah bisa unduh sertifikat ternya ya. Dukung kami untuk ti
8	positif	SATUSEHAT Kapan chat saya dibalas min?
9	positif	SATUSEHAT setengahakuntan Saya sudah DM min terkait keluhan saya, mohon dibalas ya min
10	positif	ragemah Hia ka, kalau butuh jasa vaksin dariku boleh DM aja ya. Proses cepat, aman, terpercaya, dan tembus SatuSehat juga serikon di sistem PCare ? ?
11	positif	faedy... Hia ka, kalau butuh jasa vaksin dariku boleh DM aja ya. Proses cepat, aman, terpercaya, dan tembus SatuSehat juga serikon di sistem PCare ? ? ? (f
12	positif	bukanmaksud Hia ka, kalau butuh jasa vaksin dariku boleh DM aja ya. Proses cepat, aman, terpercaya, dan tembus SatuSehat juga serikon di sistem PCar
13	positif	SATUSEHAT Hi, Min. Terima kasih sudah merespon keluhan saya. Detail sudah saya DM ya. Semoga segera direpon.
14	positif	alyamukun Apakah saat registrasi ada notif apakah nomor telepon anda bermasalah? Bisa dipanggil lewat email atau mohon dapat DM minin ya.
15	positif	conorah Untuk sender, segera DM minin ya untuk dibantu solusinya.
16	negatif	abdelrah... H Healthies, mohon dapat DM minin ya untuk dapat minin bantu.
17	negatif	setengahakuntan H Healthies, OTP dikirimkan via WA dan juga via Email ya. Apabila ada kendala mohon dapat DM minin ya.

Gambar 7. Setelah Proses Cleaning Data

M. Tahap Pre-Processing Data

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

- 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).
- Case Folding, yaitu tahapan untuk mengubah data tweet menjadi lower case (huruf kecil).
- Stemming, yaitu tahapan untuk membersihkan kata-kata imbuhan awalan dan akhiran yang terdapat dalam teks seperti “mengerjakan” menjadi “kerja”.
- Filter Tokens (by Length), yaitu tahapan untuk membuang kata-kata yang kurang dari 2 huruf dan kata-kata yang melebihi 25 huruf.
- Filter Stopwords, yaitu tahapan untuk menghapus kata bantu seperti “saya”, “dia”, “aku”, dan “mereka”.

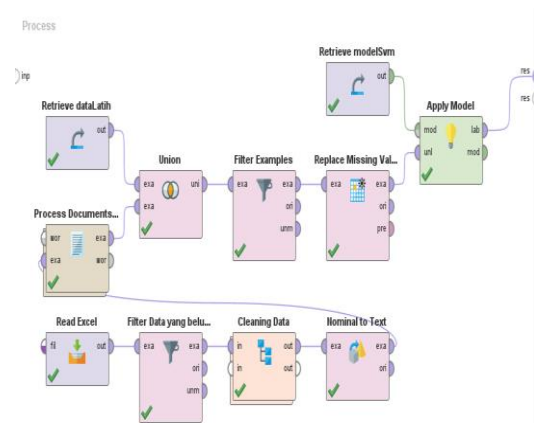


Gambar 8. Tahap Pre-Processing Data

III. HASIL DAN PEMBAHASAN

A. Tahap Pengujian

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



Gambar 9. 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 87.95%, nilai Recall Positif sebesar 1.53%, nilai Recall Negatif sebesar 100.00%, nilai Precision Positif sebesar 100.00%, nilai Precision Negatif sebesar 87.93%. Dari 1080 data latih, penulis melabeli sebanyak 132 data sebagai Sentimen Positif dan 947 data sebagai Sentimen Negatif. Untuk hasil prediksi dari Sentimen Negatif, terdapat 1079 data terprediksi Negatif dan 1 data yang terprediksi Positif.

PerformanceVector

```
PerformanceVector:
accuracy: 87.95%
ConfusionMatrix:
True:  positif negatif
positif:      2      0
negatif:     130    947
precision: 87.93% (positive class: negatif)
ConfusionMatrix:
True:  positif negatif
positif:      2      0
negatif:     130    947
recall: 100.00% (positive class: negatif)
ConfusionMatrix:
True:  positif negatif
positif:      2      0
negatif:     130    947
AUC (optimistic): 0.691 (positive class: negatif)
AUC: 0.690 (positive class: negatif)
AUC (pessimistic): 0.690 (positive class: negatif)
```

Gambar 10. Hasil Pengujian *Support Vector Machine*

Hasil perhitungan akurasi data latih dengan menggunakan metode *Naive Bayes*, didapatkan nilai Accuracy sebesar 81.65% nilai *Recall* Positif sebesar 100.00%, nilai *Recall* Negatif sebesar 79.09%, nilai *Precision* Positif sebesar 40.00%, nilai *Precision* Negatif sebesar 100.00%. Dari 1080 data latih, penulis melabeli sebanyak 132 data sebagai Sentimen Positif dan 947 data sebagai Sentimen Negatif. Untuk hasil prediksi dari Sentimen Negatif, terdapat 947 data terprediksi Negatif dan 1 data yang terprediksi Positif.

PerformanceVector

```
PerformanceVector:
accuracy: 81.65%
ConfusionMatrix:
True:  positif negatif
positif:     132    198
negatif:      0    749
precision: 100.00% (positive class: negatif)
ConfusionMatrix:
True:  positif negatif
positif:     132    198
negatif:      0    749
recall: 79.09% (positive class: negatif)
ConfusionMatrix:
True:  positif negatif
positif:     132    198
negatif:      0    749
AUC (optimistic): 1.000 (positive class: negatif)
AUC: 0.500 (positive class: negatif)
AUC (pessimistic): 0.791 (positive class: negatif)
```

Gambar 11. Hasil Pengujian *Naive Bayes*

C. Perbandingan Hasil Akurasi Algoritma 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>	81.65%
<i>Support Vector Machine</i>	87.95%

D. Hasil Sentimen Analisis

Hasil akhir yang diharapkan dari hasil Perbandingan dengan dua metode pengujian ini, yaitu hasil prediksi Sentimen Masyarakat Terhadap Isu Aplikasi SatuSehat berdasarkan data yang didapat dari Twitter dan diimplementasikan dengan metode SVM (*Support Vector Machine*) menunjukkan nilai akurasi sebesar 87.95%. Dari 1080 data uji, terprediksi 132 data sebagai Sentimen Positif dan 947 data sebagai Sentimen Negatif. Untuk hasil prediksi dari Sentimen Negatif, terdapat 1080 data terprediksi Negatif dan 1 data yang terprediksi Positif. dan Metode *Naive Bayes* menunjukkan nilai akurasi sebesar 81.65%. Dari 1081 data uji, terprediksi sebesar 947 data sebagai Sentimen Negatif dan 132 data sebagai Sentimen Positif Untuk hasil prediksi dari Sentimen Negatif, terdapat 947 data terprediksi Negatif dan 1 data yang terprediksi Positif.

IV. KESIMPULAN

Berdasarkan pembahasan, hasil, dan pengujian yang telah dilakukan, dalam penerapan dan perbandingan algoritma *Support Vector Machine* dan *naive Bayes* untuk melakukan Sentimen Analisis, maka dapat diambil kesimpulan bahwa Sentimen analisis dapat dilakukan menggunakan *Software Rapid Miner Studio* menggunakan Metode *Support Vector Machine* dan *naive Bayes* dengan mengolah data yang didapat dari Media Sosial Twitter melalui proses *Crawling* data, *Labeling* data, *Cleaning* data, *Preprocessing* dan *Klasifikasi*. Dari proses Sentimen analisis didapatkan hasil perbandingan nilai akurasi metode *Support Vector Machine* sebesar 87.95% dan *naive Bayes* 81.65% dari 585 data uji, terprediksi sebesar 1078 data sebagai Sentimen Negatif dan 2 data sebagai Sentimen Positif Untuk hasil prediksi dari Sentimen Negatif, terdapat 1078 data terprediksi Negatif dan 2 data yang terprediksi Positif. Sehingga dapat disimpulkan bahwa sebagian besar masyarakat tidak mempercayai berita hoax Aplikasi SatuSehat.

UCAPAN TERIMA KASIH

Saya ingin mengucapkan terima kasih yang sebesar-besarnya atas bimbingan, dukungan, dan dedikasi Bapak Francis Matheos Sarimole 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.

V. 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. Widiastwi, 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/psnmu.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.