Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

ANALISIS SENTIMEN KOMENTAR YOUTUBE TERHADAP ISU KESEHATAN MENTAL MENGGUNAKAN ALGORITMA *NAÏVE BAYES* DAN *K-NEAREST NEIGHBOR* (KNN)

Nurhayati[1], Suryandika Tri Nowo[2], Bambang Suhardi[3], Rika Rosnelly[4]

[1,2,3,4] *Magister Ilmu Komputer, Fakultas Teknik dan Ilmu Komputer, Universitas Potensi Utama*
[1,2,3,4] *Jl. K.L. Yos Sudarso KM. 6,5 Tj. Mulia No. 3A, Medan*
[1]*hayatipearce66@gmail.com,* [2]*suryandika@uisu.ac.id,* [3]*bambangsuhardi@uinsu.ac.id,*
[4]*rika@potensi-utama.ac.id*

ABSTRAK

Kesehatan mental merupakan salah satu isu yang semakin mendapat perhatian di seluruh dunia. Masalah kesehatan mental sering kali diabaikan, namun dampaknya dapat merusak kualitas hidup individu dan masyarakat secara keseluruhan. Faktor lain yang mempengaruhi upaya penyuluhan kesehatan mental adalah pemahaman yang kurang baik dan kesadaran yang rendah tentang kesehatan mental. Dari penjelasan permasalahan tersebut maka perlu adanya analisis sentimen untuk mengetahui opini masyarakat terhadap kesehatan mental di media sosial youtube. Analisis sentimen merupakan suatu proses untuk memahami emosi atau sentimen dari suatu teks yang ditulis oleh pengguna baik berupa sentimen positif, netral ataupun negatif. Proses pengambilan dataset dilakukan mengunakan *platform Google Colab* untuk *crawling* data dan terkumpul sekitar 2.703 komentar. Setelah dilakukan proses *cleaning* dan *preprocessing* jumlah data yang tersisa adalah sebanyak 1700. Metode yang digunakan pada penelitian ini menggunakan algoritma *Naïve Bayes* dan *K-Nearest Neighbor (k-NN).* Dalam penelitian ini, dua metode yang digunakan yaitu pelabelan manual dan pelabelan otomatis menggunakan *tools RapidMiner*. Pada tahap pertama, pelabelan manual dilakukan pada 305 data menghasilkan nilai akurasi 95% untuk algoritma *Naïve Bayes* dan nilai akurasi 85.88% untuk algoritma k-NN. Pada tahap kedua, pelabelan otomatis digunakan dengan data latih sebanyak 305 data dan data uji 1.395 data menghasilkan nilai akurasi 68.01% untuk algoritma *Naïve Bayes* dan nilai akurasi 48.97% untuk algoritma k-NN. Hasil penelitian menunjukkan bahwa algoritma *Naïve Bayes* memiliki akurasi yang lebih tinggi dibandingkan dengan algoritma *K-Nearest Neighbor* dalam mengklasifikasikan sentimen dari komentar YouTube terkait isu kesehatan mental.

***Kata Kunci*— Analisis Sentimen, Naive Bayes, K-Nearest Neighbor, Kesehatan Mental, YouTube, RapidMiner.**

ABSTRAKT

Mental health is an issue that is receiving increasing attention around the world. Mental health problems are often overlooked, yet their impact can be detrimental to the quality of life of individuals and society as a whole. Other factors that affect mental health counseling efforts are poor understanding and low awareness of mental health. From the explanation of these problems, it is necessary to analyze sentiment to find out public opinion on mental health on YouTube social media. Sentiment analysis is a process to understand the emotions or sentiments of a text written by users in the form of positive, neutral or negative sentiments. The dataset retrieval process was carried out using the Google Colab platform for crawling data and collected around 2,703 comments. After cleaning and preprocessing the remaining amount of data is 1700. The method used in this research uses the Naïve Bayes and K-Nearest Neighbor (k-NN) algorithms. In this research, two methods are used, namely manual labeling and automatic labeling using RapidMiner tools. In the first stage, manual labeling was performed on 305 data resulting in 95% accuracy value for Naïve Bayes algorithm and 85.88% accuracy value for k-NN algorithm. In the second stage, automatic labeling was used with 305 training data and 1,395 test data resulting in an accuracy value of 68.01% for the Naïve Bayes algorithm and an accuracy value of 48.97% for the k-NN algorithm. The results show that the Naïve Bayes algorithm has higher accuracy than the K-Nearest Neighbor algorithm in classifying sentiment from YouTube comments related to mental health issues.

***Keywords*— Sentiment Analysis, Naive Bayes, K-Nearest Neighbor, Mental Health, YouTube, RapidMiner.**

Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

I. PENDAHULUAN

Kesehatan mental merupakan komponen integral dari kesehatan dan kesejahteraan manusia yang mencakup aspek emosional, psikologis, dan sosial. Dalam beberapa dekade terakhir, perhatian terhadap kesehatan mental telah mengalami peningkatan signifikan seiring dengan meningkatnya kesadaran masyarakat global akan pentingnya aspek ini dalam kehidupan sehari-hari. Kesehatan mental adalah suatu kondisi seseorang yang memungkinkan berkembangnya semua aspek perkembangan, baik fisik, intelektual, dan emosional yang optimal serta selaras dengan perkembangan orang lain, sehingga selanjutnya mampu berinteraksi dengan lingkungan sekitarnya. Gejala jiwa atau fungsi jiwa seperti pikiran, perasaan, kemauan, sikap, persepsi, pandangan dan keyakinan hidup harus saling berkoordinasi satu sama lain, sehingga muncul keharmonisan yang terhindar dari segala perasaan ragu, gundah, gelisah dan konflik batin (pertentangan pada diri individu itu sendiri) [1].

Menurut *World Health Organization* (WHO), sekitar 970 juta orang di seluruh dunia mengalami gangguan kesehatan mental pada tahun 2019. Angka ini meningkat secara drastis selama pandemi COVID-19, dengan peningkatan 25% kasus depresi dan kecemasan global pada tahun pertama pandemi [2]. Di Indonesia sendiri, berdasarkan data Riset Kesehatan Dasar (Riskesdas) tahun 2018, prevalensi gangguan mental emosional pada penduduk usia ≥15 tahun mencapai 9,8% dari total populasi [3].

YouTube merupakan sebuah platform yang menyediakan informasi berupa video, sehingga pengguna juga dapat membagikan video-video berbagai jenis seperti video musik, edukasi, tutorial dan lain-lainnya. YouTube juga sangat berperan penting dalam penyebaran konten berita yang sedang terjadi, karena pengguna YouTube lebih mudah dalam menyampaikan sebuah berita dan penggunanya juga sangat luas [4]. YouTube tidak hanya berfungsi sebagai platform informasi namun juga sebagai wadah bagi individu untuk berbagi pengalaman, wawasan, dan dukungan terkait isu-isu kesehatan mental. Dalam rangka memperingati Hari Kesehatan Jiwa Sedunia, channel YouTube "Menjadi Manusia" membuat sebuah dokumenter tentang kesehatan dan gangguan mental. Tayangan ini berhasil mendapat 1.174.111 penonton dan 2.703 komentar. Meski video tersebut diunggah 5 tahun yang lalu, namun hingga saat ini masih mendapat komentar yang beragam dari penonton YouTube. Selain karena isinya yang menyentuh, tayangan ini juga dianggap memberikan informasi yang menarik terkait kesehatan dan gangguan mental.

Penelitian ini bertujuan untuk menganalisis bagaimana sentimen pengguna YouTube dan menambah pengetahuan dalam memahami pandangan publik, tantangan yang mereka hadapi, dan bagaimana media sosial dapat mempengaruhi persepsi tentang kesehatan mental. Analisis sentimen yaitu suatu bidang pengelolaan data tekstual yang melakukan studi berdasarkan opini, sentimen, evaluasi, prilaku dan emosi seseorang yang dapat digunakan sebagai bahan evaluasi [5]. Analisis sentimen juga dikenal sebagai *Opinion Mining* yang memungkinkan para peneliti untuk menyaring teks yang dikumpulkan melalui berbagai sumber dan memperoleh gambaran tentang perasaan subjek yang sedang dibahas. Bidang ini sangat bergantung pada teknik-teknik dalam *Natural Language Processing* (NLP). NLP memungkinkan mesin untuk memproses bahasa alami manusia dan menerjemahkannya ke dalam format yang dapat dipahami oleh mesin tersebut [6]

Metode yang digunakan dalam penelitian ini adalah algoritma *Naïve Bayes* (NB) dan *K-Nearest Neighbor* (KNN). Algoritma *Naive Bayes* merupakan metode yang berbasis pada prinsip probabilitas dikenal efektif dalam klasifikasi teks. Nilai atribut secara kondisional terpisah satu sama lain apabila diberikan nilai *output,* menurut asumsi penyederhanaan algoritma *Naïve Bayes.* Algoritma *K-Nearest Neighbor* merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut [7].

Penelitian terkait analisis sentimen dengan menggunakan algoritma klasifikasi sudah banyak dilakukan oleh para peneliti, seperti yang dilakukan oleh Mahesworo dan kawan-kawan dengan topik analisis sentimen kesehatan mental menggunakan *K-Nearest Neighbors* pada sosial media *Twitter* tahun 2022. Penelitian tersebut menggunakan algoritma *K-Nearest Neighbors* dengan melakukan perbandingan algoritma klasifikasi *Support Vector Machine* dan *Decision Tree.* Hasil penelitian ini berdasarkan dataset ulasan sentimen positif sebanyak 639 dan ulasan sentimen negatif sebanyak 193, maka hasil pemrosesan modeling dengan menggunakan algoritma *K-Nearest Neighbors* didapatkan hasil terbaik saat menggunakan metode *split data* 70:30 dengan nilai k berada pada angka 5, yaitu menghasilkan *precision* 60.87%, r*ecall* 44.03% dan *accuracy* 58.39% [8].

Penelitian lain juga dilakukan oleh Kenny Yan dan kawan-kawan dengan topik analisis sentimen komentar netizen twitter terhadap kesehatan mental masyarakat Indonesia tahun 2022. Pada penelitian ini menggunakan algoritma *Naïve Bayes*. Hasil analisis sentimen yang telah dilakukan diketahui bahwa komentar pengguna media sosial Twitter terhadap kesehatan mental adalah negatif dengan klasifikasi negatif sebesar 50.8% dan akurasi dari algoritma *Naïve Bayes* sebesar 79% [9].

Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

Penelitian yang dilakukan oleh Muhammad Daffa Al Fahreza dan kawan-kawan dengan topik analisis sentimen: pengaruh jam kerja terhadap kesehatan mental generasi Z menggunakan algoritma *Naïve Bayes, Support Vector Machine* dan *Stemming Sastrawi*. Pada penelitian ini menggunakan dataset twitter sebanyak 2956 tweet. Hasil penelitian menunjukkan bahwa algoritma *Stemming Sastrawi* dengan model SVM lebih unggul dengan akurasi sebesar 91% dan algoritma *Stemming Sastrawi* dengan model *Naïve Bayes* dengan akurasi sebesar 84% [10].

Penelitian yang dilakukan oleh Syahril dan kawan-kawan dengan topik analisis sentimen relokasi Ibukota Nusantara menggunakan algoritma *Naïve Bayes* dan *KNN* tahun 2023. Pada penelitian ini menggunakan data *tweet* hasil *crawling* menggunakan *RapidMiner* dengan jumlah 800 data yang sudah melalui proses pembersihan data. Hasil penelitian menyajikan algoritma KNN lebih unggul dengan tingkat akurasi 88.12%, *precision* 93.98% dan *recall* 81.53% sedangkan pada algoritma *Naïve Bayes* tingkat akurasi sebesar 82.27%, *precision* sebesar 86.36% dan *recall* sebesar 76.93% [11].

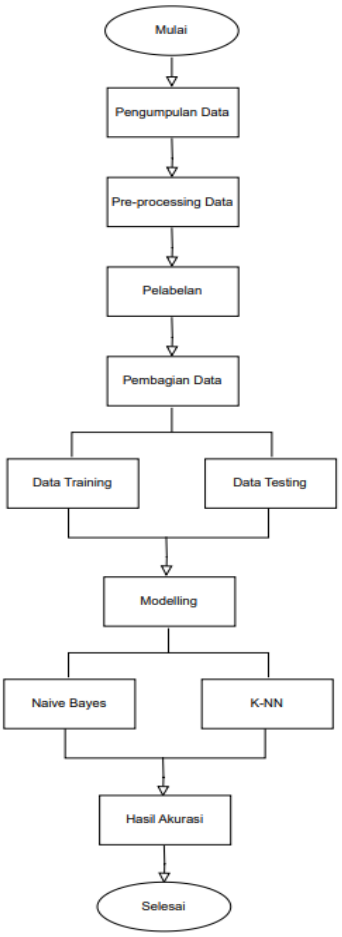
Sentimen Analisis dengan algoritma *Naïve Bayes* dan *KNN* juga digunakan pada penelitian Fatmanisa dan kawan-kawan tahun 2019 dengan topik perbandingan metode klasifikasi sentimen analisis penggunaan *e-wallet*. Hasil *accuracy* dari masing-masing model klasifikasi yaitu NB sebesar 73.03% dan *k*-NN sebesar 89.44%, *precision* NB sebesar 21.40% dan *k*-NN sebesar 65.45%, dan *recall* NB sebesar 48.32% dan *k*-NN sebesar 22.25%. Dari hasil perbandingan metode membuktikan bahwa algoritma *k*-NN dengan *accuracy* terbaik yaitu sebesar 89.44% [12].

Penelitian yang dilakukan oleh Muhamad Taufik dan kawan-kawan dengan topik analisis komentar pengguna Youtube terhadap kebijakan baru BPJS menggunakan *Naïve Bayes* tahun 2024. Hasil penelitian menunjukan tingkat akurasi tertinggi model pada data uji mencapai 96% dengan rasio 80:20. Hal ini menunjukan bahwa model mampu dengan baik dalam mengklasifikasikan sentimen pada komentar. Penelitian ini didominasi oleh sentimen komentar positif sebesar 45.9% atau sebanyak 1.354 data dari total 2.948 data komentar [13].

Berdasarkan penjelasan di atas maka penelitian ini bertujuan untuk melakukan analisis sentimen terhadap komentar-komentar di YouTube yang pada penelitian sebelumnya analisis sentimen hanya berfokus pada data twiter, mengungkap bagaimana masyarakat bereaksi terhadap konten yang berkaitan dengan kesehatan mental menggunakan perbandingan algoritma *Naïve Bayes* dan *k-NN* dengan *tools RapidMiner*. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi terhadap pemahaman isu kesehatan mental dalam konteks media sosial serta memberikan rekomendasi bagi pembuat konten dan penggiat kesehatan mental untuk lebih memahami audiens dan merespons dengan cara yang lebih efektif.

II. METODE PENELITIAN

Penelitian ini mengunakan pendekatan *Machine Learning* lebih tepatnya pada bagian *text mining*. Pendekatan *Machine Learning* dalam analisis sentimen dimulai dengan menstandarisasi data teks melalui preprocesing dan menghapus informasi yang tidak relevan [14].



Gambar 1. Tahapan Alur Penelitian

Adapun yang menjadi metode penelitian ini dengan menggunakan *Cross-Industry Standard Process for Data Mining (CRISP-DM).* Metode tersebut memiliki beberapa tahapan seperti yang terlihat pada gambar 1 [15]

Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

Gambar 2. Metode CRISP-DM

A. *Business Understanding*

Pada tahapan ini, dilakukan pemahaman tentang objek penelitian yang akan dilakukan yaitu analisis sentimen terhadap isu kesehatan mental pada media sosial YouTube.

B. *Data Understanding*

Pada tahap *data understanding*, proses pengambilan data mentah dilakukan sesuai dengan variabel yang dibutuhkan. Data dikumpulkan dari komentar channel YouTube "Menjadi Manusia", yaitu video dokumenter tentang kesehatan dan gangguan mental pada tanggal 12 Desember 2024. Jumlah data yang diperoleh sebanyak 2.703 komentar. Proses pengambilan data dilakukan mengunakan *platform Google Colab* untuk *crawling* data. Setelah data terkumpul maka proses *cleaning* dilakukan dan didapatkan komentar sebanyak 1.700 data.

C. *Data Preparation*

Pada tahapan ini yaitu proses persiapan data yang bertujuan untuk mendapatkan data bersih dan siap digunakan dalam penelitian. Proses ini meliputi *Transform Cases, Tokenize, Filter Tokens, Filter Stopwords* dan *Stemming*. Data yang telah dikumpulkan, selanjutnya dilakukan proses *preprocessing* menggunakan *Microsoft Excel* dan juga bantuan *tools RapidMiner*. Berikut tahapan yang terdapat dalam *data preparation [16]*

1. *Transform Cases*

Proses *Transform Cases* adalah proses mengubah semua huruf pada teks menjadi huruf kecil atau *lowercase*. Hal ini dilakukan supaya kata yang sebenarnya sama namun ditulis dengan huruf besar atau kecil dianggap sama dalam proses analisis data.

2. *Tokenize*

Proses dimana kalimat yang akan dipisah menjadi pecahan kata tunggal atau sebagai token.

3. *Filter Tokens*

Filter Tokens adalah proses untuk menyaring hasil token berdasarkan panjang karakter atau jumlah minimal huruf yang terdapat dalam satu kata. Pada proses ini menggunakan operator *filter tokens (by length)* dan mengubah minimal *chars* menjadi 3.

4. *Filter Stopword*

Filter Stopword adalah penghapusan kata-kata yang umumnya tidak memiliki makna atau tidak memiliki informasi dan dapat diabaikan pada analisis teks. Pada umumnya kata-kata seperti "dan", "atau", "dari" termasuk dalam daftar kata *Stopword*. Proses *Stopword* ini menggunakan modul *'nltk'* (*Natural Language Toolkit*) dalam bahasa Indonesia.

5. *Stemming*

Stemming adalah proses mengubah kata-kata imbuhan dalam teks menjadi kata dasar, sehingga kata yang berbeda tetapi memiliki makna dasar yang sama akan diubah menjadi bentuk kata yang seragam.

D. *Modeling*

Merupakan tahap pemilihan teknik penambangan dengan menentukan algoritma yang akan digunakan. Penelitian ini menggunakan *tools* yang digunakan untuk melakukan pemodelan sesuai dengan teknik yang telah ditentukan, *tools* tersebut adalah *RapidMiner* versi 9.10. Penelitian ini menggunakan 2 (dua) algoritma klasifikasi sebagai modelnya. Algoritma klasifikasi yang digunakan yaitu, *Naïve Bayes* (NB) dan *K-Nearest Neighbor (KNN)*. Untuk mendapatkan nilai akurasi terbaik pada setiap algoritma, pengetesan setiap model menghasilkan klasifikasi komentar berupa komentar positif, netral atau negatif.

E. *Evaluation*

Setelah menyelesaikan tahapan *modeling*, dilakukan validasi guna menguji model yang diajukan dan melakukan evaluasi pada *dataset* model menggunakan *Confusion Matrix* menggunakan *tools* Rapidminer. *Confusion Matrix* merupakan metode untuk mengevaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah [17]. *Accuracy, precission,* dan *recall* dihitung dengan menggunakan teknik *confusion matrix*. Teknik *confusion matrix* terdiri dari : *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Nilai-nilai tersebut dapat dihitung dengan memperhatikan *confusion matrix* yang terdapat pada tabel 1 di bawah ini [18].

Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

Tabel 1. *Confusion Matrix*

Actual	Predicted	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Keterangan:
TP = *True Positive* terprediksi positif benar
FP = *False Positive* terprediksi salah positif
TN = *True Negative* terprediksi benar negatif
FN = *False Negative* terprediksi salah negatif

F. *Deployment*

Pada tahap ini, informasi dari hasil analisis perbandingan metode klasifikasi yang diperoleh pada tahap sebelumnya dibuat ke dalam penulisan laporan sederhana dan artikel jurnal tentang hasil penelitian.

III. HASIL DAN PEMBAHASAN

A. *Dataset*

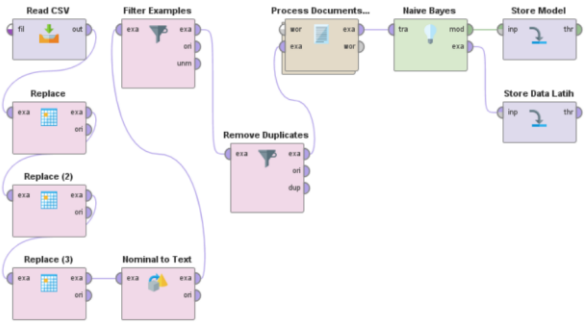
Penelitian ini menggunakan data yang diperoleh dari komentar channel YouTube "Menjadi Manusia", yaitu video dokumenter tentang kesehatan dan gangguan mental pada tanggal 12 Desember 2024. Jumlah data yang diperoleh sebanyak 2.703 komentar. Proses pengambilan data dilakukan mengunakan *platform Google Colab* untuk *crawling data.* Setelah data terkumpul maka proses *cleaning* dilakukan dan didapatkan komentar sebanyak 1.700 data yang telah berlabel dengan label positif, netral dan negatif.

Dengan membagi data menjadi dua kelompok, *labeling* bertujuan untuk menentukan apakah sentimen tersebut mengarah ke arah positif, netral atau negatif. Peneliti akan membandingkan pelabelan data dalam konteks ini. Tahap pertama, dataset sebanyak 305 data akan dilabeli secara manual. Setelah itu, peneliti menggunakan *tools RapidMiner* untuk menerapkan pelabelan secara otomatis pada dataset yang sama. Namun, untuk tahap pelabelan otomatis, algoritma *Naïve Bayes* dan *k-NN* harus dilatih dengan 305 data latih yang akan dilabeli secara manual. Tujuan pelabelan data dengan dua metode ini adalah untuk membandingkan dan menilai hasil ketepatan pelabelan dari dari kedua tahap tersebut. Tabel 2 menunjukkan proses pelabelan manual terhadap teks dengan sentimen positif, netral dan negatif.

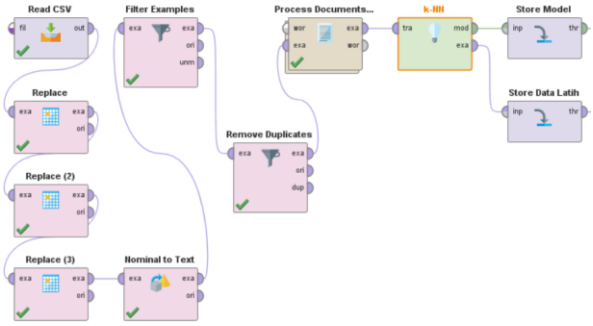
Tabel 2. *Labeling* Manual Data Latih

Text	Sentimen
Sy merasa kosong dan tak berguna otak tak mampu berfikir normal, tak mampu berinteraksi sosial dg baik lebih suka menyendiri, dan sering berfikir untuk mengahiri hidup, sy tak mampu berbicara untuk mengungkapkan perasaan.. Sy ingin hidup normal ya allah, doakan saya untuk mampu bertahan kawan	Negatif
Kami juga manusia....	Netral
Untuk seluruh orang yang mengalami, mengidap dan merasakan mental ill baik yang minor ataupun major. Semangat yaa kalian tetap manusia doa terbaik ku untuk kalian kalian pasti bisa dan semoga kalian bisa mendapatkan kehidupan yang tenang dan bahagia. Bagi yang memiliki trauma semoga semua trauma dan luka batin kalian sembuh total dan bisa mendapatkan kebahagiaan dalam hidup. Untuk kalian semua jika ada orang yang ga ngerti tentang kalian dan menjudge, menghina dan menyingkirkan kalian. Biarkan dan gausah di hiraukan karena yang paling mengerti diri kalian adalah tuhan dan diri kalian sendiri.	Positif

Setelah menyelesaikan langkah pelabelan secara manual, langkah selanjutnya adalah melakukan analisis sentimen secara otomatis menggunakan *tools* RapidMiner pada dataset yang belum mempunyai label sebanyak 1.395 data. Tahapan proses pelabelan otomatis dapat dilihat pada gambar berikut.



Gambar 2. Model pelabelan dengan algoritma *Naïve Bayes*



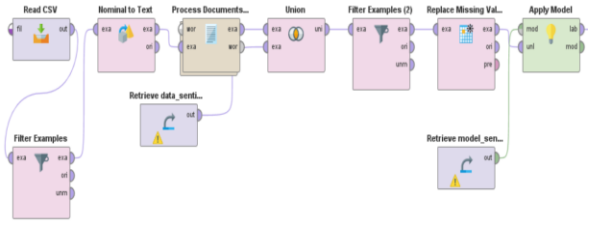
Gambar 3. Model pelabelan dengan algoritma *k-NN*

Model klasifikasi digunakan untuk menentukan apakah sentimen pada data yang belum diberi label secara manual memiliki nilai positif, netral atau negatif, seperti yang ditunjukkan pada Gambar 2 dan Gambar 3. Setelah model *Naïve Bayes* dan *k-NN* berhasil

Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

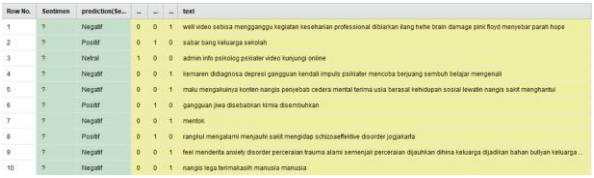
menentukan polaritas sentimen, peneliti akan menyimpan data tersebut ke dalam *operator store*. Selanjutnya, *store* data latih dan *operator store* model akan digunakan untuk proses berikutnya.

Model sentimen yang dibuat dengan algoritma *Naïve Bayes* dan *k-NN* sebelumnya akan diterapkan pada data yang belum mempunyai label di tahapan berikutnya. Selanjutnya, pelabelan akan diproses secara otomatis dengan *RapidMiner*, seperti pada gambar 4.



Gambar 4. Proses pelabelan otomatis

Berdasarkan Gambar 4 di atas, dataset terdiri dari dua dataset, masing-masing dengan label manual dan otomatis. Untuk menggabungkan kedua dataset, operator *union* dan *filter examples* digunakan untuk menghilangkan data latih yang tidak sesuai. Untuk mengintegrasikan data yang tidak lengkap atau kosong, digunakan operator untuk menggantikan nilai yang hilang *(missing values)*. Pelabelan secara otomatis dengan RapidMiner telah selesai dilakukan dan hasilnya dapat dilihat pada Gambar 5.



Gambar 5. Hasil pelabelan otomatis

B. *Data Pre-paration*

Setelah tahap pengumpulan dan pelabelan selesai dilakukan, langkah berikutnya adalah data *pre-paration* atau data *preprocessing*. Tahapan ini terdiri dari *Transform Cases, Tokenize, Filter Stopword, Filter Tokens* dan *Stemming* yang ditunjukkan pada Gambar 6.

Gambar 6. Tahap data *preprocessing*

Hasil dari setiap proses *preprocessing* dapat dilihat pada tabel 3.

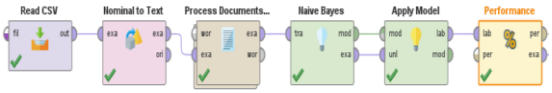
Tabel 3. Tahapan *Preprocessing*

Preprocessing	Sebelum	Sesudah
Transform Cases	Saya Bipolar Gejala nya memang seperti apa yang di bagikan di atas sesuai dengan yang mereka alami Namun juga rasanya setiap orang yang mengalami nya punya rasa dan cara yang berbeda	saya bipolar gejala nya memang seperti apa yang di bagikan di atas sesuai dengan yang mereka alami namun juga rasanya setiap orang yang mengalami nya punya rasa dan cara yang berbeda
Tokenize	saya bipolar gejala nya memang seperti apa yang di bagikan di atas sesuai dengan yang mereka alami namun juga rasanya setiap orang yang mengalami nya punya rasa dan cara yang berbeda	saya, bipolar, gejala, nya, memang, seperti, apa, yang, di, bagikan, di, atas, sesuai, dengan, yang, mereka, alami, namun, juga, rasanya, setiap, orang, yang, mengalami, nya, punya, rasa, dan, cara, yang, berbeda
Filter Tokens	saya, bipolar, gejala, nya, memang, seperti, apa, yang, di, bagikan, di, atas, sesuai, dengan, yang, mereka, alami, namun, juga, rasanya, setiap, orang, yang, mengalami, nya, punya, rasa, dan, cara, yang, berbeda	saya, bipolar, gejala, memang, seperti, apa, yang, bagikan, atas, sesuai, dengan, mereka, alami, namun, juga, rasanya, setiap, orang, mengalami, punya, rasa, dan, cara berbeda
Filter Stopwords	saya, bipolar, gejala, memang, seperti, apa, yang, bagikan, atas, sesuai, dengan, mereka, alami, namun, juga, rasanya, setiap, orang, mengalami, punya, rasa, dan, cara berbeda	bipolar, gejala, alami, rasa, cara, berbeda
Stemming	bipolar, gejala, alami, rasa, cara, berbeda	bipolar, gejala, alami, rasa, cara, beda

C. *Modeling*

Pada tahapan ini, pendekatan algoritma yang dibahas menggunakan algoritma *Naïve Bayes* dan *k-NN*. Pada Gambar 7 dan Gambar 8 menunjukkan tampilan aplikasi RapidMiner yang menggunakan algoritma tersebut yaitu pemrosesan data yang telah diberi label secara manual untuk melatih model klasifikasi.

Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

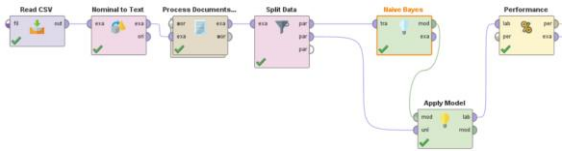


Gambar 7. Klasifikasi *labeling* manual algoritma *Naïve Bayes*

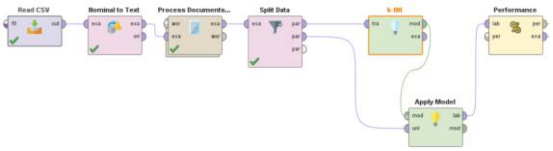
Gambar 8. Klasifikasi *labeling* manual algoritma *k-NN*

Adapun tujuan dari tahapan modeling tersebut adalah untuk membuat model klasifikasi yang mampu menerapkan pola dari data latih ke data yang tidak pernah dilihat sebelumnya, meningkatkan ketepatan prediksi dalam proses pengambilan keputusan, dan menggunakan kemampuan analisis data untuk mengelola data dengan baik.

Selain itu, peneliti melakukan tahapan klasifikasi algoritma *Naïve Bayes* dan *k-NN* pada data latih yang telah diberi label. Selanjutnya, pengujian data uji dilakukan dengan algoritma tersebut untuk menemukan dan memprediksi label atau sentimen yang tepat untuk setiap data uji.



Gambar 9. Klasifikasi *labeling* otomatis algoritma *Naïve Bayes*



Gambar 10. Klasifikasi *labeling* otomatis algoritma *k-NN*

Pada Gambar 9 dan Gambar 10, proses yang dilakukan pada tahapan ini menggunakan *Operator Split Data* untuk memisahkan dataset menjadi dua bagian penting untuk proses *machine learning.* Pada bagian parameters, membaginya menjadi 20% untuk data latih *(training)* dan 80% untuk data data uji *(testing),* menggunakan rasio 80:20.

Selanjutnya, melakukan evaluasi penerapan algoritma menggunakan operator Apply Model. Proses ini digunakan sebagai metode pembelajaran mesin untuk menguji kinerja model serta memprediksi hasil berdasarkan data yang belum memiliki label. Untuk mengevaluasi kinerja model yang telat dibuat operator *performance* digunakan juga untuk membantu dalam memahami sejauh apa model tersebut dapat memprediksi data dengan baik.

Visualisasi Wordcloud

Wordcloud bertujuan menampilkan data *text* secara visual. Semakin besar kata tersebut, maka jumlah frekuensi kata tersebut semakin banyak. Pada Gambar 11, menunjukkan visualisasi wordcloud yang menampilkan sentimen dari kalimat yang paling banyak muncul pada penelitian ini.

Gambar 11. Hasil *Wordcloud*

D. Hasil Pengujian dan Evaluasi

Selanjutnya pada proses evaluasi, yang bertujuan untuk memastikan bahwa hasil klasifikasi adalah valid dengan menggunakan perhitungan *confusion matrix* yang didasarkan pada penelitian tentang analisis sentimen kesehatan mental. Untuk melakukan perhitungan masing-masing dari dua tahap klasifikasi pada pelabelan data yang telah dilakukan di atas. Pada Gambar 12 dan Gambar 13 menunjukkan perhitungan *confusion matrix* untuk hasil pelabelan secara manual.

accuracy: 95.00%

	true Netral	true Positif	true Negatif	class precision
pred. Netral	431	2	22	94.73%
pred. Positif	21	452	40	88.11%
pred. Negatif	0	0	732	100.00%
class recall	95.35%	99.56%	92.19%	

Gambar 12. Hasil akurasi algoritma *Naïve Bayes*

accuracy: 85.88%

	true Netral	true Positif	true Negatif	class precision
pred. Netral	274	5	3	97.16%
pred. Positif	26	396	1	93.62%
pred. Negatif	152	53	790	79.40%
class recall	60.62%	87.22%	99.50%	

Gambar 13. Hasil akurasi algoritma *k-NN*

Selanjutnya pada Gambar 14 dan Gambar 15 menunjukkan perhitungan *confusion matrix* untuk hasil pelabelan secara otomatis.

Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

accuracy: 68.01%

	true Netral	true Positif	true Negatif	class precision
pred. Netral	187	64	37	64.93%
pred. Positif	58	202	62	62.73%
pred. Negatif	117	97	536	71.47%
class recall	51.66%	55.65%	84.41%	

Gambar 14. Akurasi Algoritma Naïve Bayes

accuracy: 48.97%

	true Netral	true Positif	true Negatif	class precision
pred. Netral	71	27	57	45.81%
pred. Positif	203	296	279	38.05%
pred. Negatif	88	40	299	70.02%
class recall	19.61%	81.54%	47.09%	

Gambar 15. Akurasi algoritma k-NN

KESIMPULAN

Hasil dari penelitian terkait sentimen terhadap isu kesehatan mental di *platform* YouTube yaitu pengumpulan data dilakukan pada tanggal 12 Desember 2024, dengan mengumpulkan sekitar 2.703 komentar. Jumlah data yang tersisa adalah 1.700 data setelah dilakukan proses *cleaning* dan *preprocessing*. Penelitian ini menggunakan dua pendekatan yaitu pelabelan secara manual pada 305 data dan pelabelan secara otomatis menggunakan *tools* RapidMiner untuk diklasifikasikan menggunakan algoritma *Naïve Bayes* dan *k-NN*. Pada tahap pelabelan pertama dilakukan secara manual pada 305 data menghasilkan nilai *accuracy* 95% untuk algoritma *Naïve Bayes* dan nilai *accuracy* 85.88% untuk algoritma k-NN. Selanjutnya, pada tahap kedua dilakukan pelabelan secara otomatis dengan data latih sebanyak 305 data dan data uji 1.395 data menghasilkan nilai *accuracy* 68.01% untuk algoritma *Naïve Bayes* dan nilai *accuracy* 48.97% untuk algoritma k-NN. Disimpulkan bahwa proses pelabelan manual pada algoritma *Naïve Bayes* dan *k-NN* untuk melakukan klasifikasi yang sangat tepat pada dataset yang telah diberi label secara manual. Kemudian pada tahap pelabelan otomatis memiliki tingkat keakuratan yang rendah. Pada penelitian ini memungkinkan faktor-faktor yang mempengaruhi kinerja model, diantaranya kualitas data latih dan data uji. Hal ini memberikan pemahaman bahwa pelabelan yang dilakukan secara otomatis dapat memberikan hasil yang baik, namun harus memperhatikan dan menigkatkan kualitas data latih dan data uji untuk mendapatkan tingkat akurasi yang lebih tinggi. Semakin banyak data yang sudah dilabeli manual, semakin tinggi tingkat akurasi *Naïve Bayes* dan *k-NN* dalam memberikan label secara otomatis. Tujuan dari proses ini adalah untuk melatih algoritma *Naïve Bayes* dan *k-NN* agar algoritma tersebut dapat melakukan pelabelan secara otomatis dan mendapat tingkat akurasi yang tinggi. Saran untuk penelitian lanjutan agar meningkatkan performa model, disarankan untuk menangani ketidakseimbangan data dengan metode seperti oversampling kelas minoritas, undersampling kelas mayoritas, atau menggunakan algoritma yang lebih robust seperti SMOTE. Kemudian menggunakan algoritma klasifikasi yang lain yang di optimasikan menggunakan *Particle Swarm Optimization (PSO)* untuk mendapatkan nilai akurasi yang lebih tinggi.

UCAPAN TERIMA KASIH

Penelitian ini tidak dapat terselesaikan tanpa dukungan dan bantuan dari berbagai pihak. Maka dari itu, kami ingin mengucapkan terima kasih yang sebesar-besarnya kepada; teman-teman mahasiswa dan Ibu Prof. Dr. Rika Rosnelly, M.Kom selaku dosen yang telah membantu menyelesaikan penelitian ini.

REFERENSI

[1] D. V. Fakhriyani, “Kesehatan Mental,” 2019. [Online]. Available: https://www.researchgate.net/publication/348819060

[2] World Health Organization (WHO), “Transforming mental health for all,” Geneva, 2022.

[3] Kementerian Kesehatan RI, “Laporan Riskesdas 2018 Nasional,” Jakarta, 2018.

[4] H. Hidayat, F. Santoso, and L. F. Lidimillah, “Analisis Sentimen Pengguna YouTube Tentang Rohingya Menggunakan Algoritma SVM (Support Vector Machine),” *G-Tech: Jurnal Teknologi Terapan*, vol. 8, no. 3, pp. 1729–1738, Jul. 2024, doi: 10.33379/gtech.v8i3.4497.

[5] N. Satya Marga, A. Rahman Isnain, and D. Alita, “Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA),” *Abstrak*, vol. 453, no. 4, pp. 453–463, 2021, [Online]. Available: http://jim.teknokrat.ac.id/index.php/informatika

[6] A. Rajput, “Natural language processing, sentiment analysis, and clinical analytics,” in *Innovation in Health Informatics: A Smart Healthcare Primer*, Elsevier, 2019, pp. 79–97. doi: 10.1016/B978-0-12-819043-2.00003-4.

[7] M. Faid, M. Jasri, and T. Rahmawati, “Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi,” *Teknika*, vol. 8, no. 1, pp. 11–16, Jun. 2019, doi: 10.34148/teknika.v8i1.95.

[8] M. Langgeng Wicaksono and D. Apriana, “ANALISIS SENTIMEN KESEHATAN MENTAL MENGGUNAKAN K-NEAREST

NEIGHBORS PADA SOSIAL MEDIA TWITTER," 2022.
[9] K. Yan, D. Arisandi, and) Tony, "Jurnal Ilmu Komputer dan Sistem Informasi ANALISIS SENTIMEN KOMENTAR NETIZEN TWITTER TERHADAP KESEHATAN MENTAL MASYARAKAT INDONESIA."
[10] M. Daffa, A. Fahreza, A. Luthfiarta, M. Rafid, M. Indrawan, and A. Nugraha, "Analisis Sentimen: Pengaruh Jam Kerja Terhadap Kesehatan Mental Generasi Z," *JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST)*, vol. 5, no. 1, pp. 2723–1453, 2024, doi: 10.52158/jacost.715.
[11] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *Jurnal KomtekInfo*, pp. 1–7, Jan. 2023, doi: 10.35134/komtekinfo.v10i1.330.
[12] F. M. Delta Maharani, A. Lia Hananto, S. Shofia Hilabi, F. Nur Apriani, A. Hananto, and B. Huda, "Perbandingan Metode Klasifikasi Sentimen Analisis Penggunaan E-Wallet Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor," *METIK JURNAL*, vol. 6, no. 2, pp. 97–103, Dec. 2022, doi: 10.47002/metik.v6i2.372.
[13] K. Baru Badan Penyelenggara Jaminan Kesehatan Sosial Menggunakan Naïve Bayes Muhamad Taufik Sugandi and U. Hayati, "Jurnal Informatika dan Rekayasa Perangkat Lunak Analisis Sentimen Komentar Pengguna Youtube terhadap".
[14] K. L. Tan, C. P. Lee, and K. M. Lim, "A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research," Apr. 01, 2023, *MDPI*. doi: 10.3390/app13074550.
[15] S. Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti *et al.*, "Terakreditasi SINTA Peringkat 2 Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada Komentar Media Berita Online," *masa berlaku mulai*, vol. 1, no. 3, pp. 176–183, 2017.
[16] J. W. Iskandar and Y. Nataliani, "Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 6, pp. 1120–1126, Dec. 2021, doi: 10.29207/resti.v5i6.3588.
[17] M. Windarti and A. Suradi, "PERBANDINGAN KINERJA 6 ALGORITME KLASIFIKASI DATA MINING UNTUK PREDIKSI MASA STUDI MAHASISWA," 2019.
[18] A. A. Pratiwi and M. Kamayani, "Perbandingan Pelabelan Data dalam Analisis Sentimen Kurikulum Proyek di platform TikTok: Pendekatan Naïve Bayes," *Jurnal Eksplora Informatika*, vol. 14, no. 1, pp. 96–107, Sep. 2024, doi: 10.30864/eksplora.v14i1.1093.