

Komparasi Algoritma Naïve Bayes Classifier, K-Nearest Neighbor, dan Decision Tree untuk Menganalisis Sentimen Masyarakat terhadap Kesehatan Mental pada Media Sosial Twitter

Tachiyya Nailal Khusna¹, Bambang Sugiantoro²

^{1,2} Program Studi Magister Informatika, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sunan Kalijaga Yogyakarta
Email : tachiyya@gmail.com¹, bambang.sugiantoro@uin-suka.ac.id²

ABSTRAK

Twitter adalah platform media sosial paling populer di internet dan menawarkan kesempatan kepada pengguna untuk mengekspresikan pendapat mereka tentang berbagai topik, termasuk berita, peristiwa terkini, kabaret, dan topik lainnya. Salah satunya adalah opini Kesehatan mental. Namun, ada banyak opini yang berbeda di Twitter dari pengguna internet yang berbeda, Oleh karena itu, perlu dilakukan analisis dasar mengenai opini publik untuk menjelaskan dan memberikan informasi baru mengenai topik tertentu yang berkaitan dengan kesehatan mental, metode yang digunakan adalah algoritma Naïve Bayes Classifier dan K-Nearest Neighbors (K-NN) dan Decision Tree. Sebanyak 5000 data yang diambil menggunakan API twitter dengan kata kunci "Kesehatan Mental". Dimulai dari klasifikasi opini positif atau negatif, data cleansing, preprocessing, hingga didapatkan hasil akhir. Kemudian dikalkulasikan ke dalam dua algoritma yang berbeda untuk dibandingkan, algoritma yang digunakan yaitu Naïve Bayes Classifier dan K-Nearest Neighbors (K-NN) dan Decision Tree dengan tujuan menemukan akurasi terbaik. Hasil tertinggi dari penelitian ini adalah algoritma Naïve Bayes dengan nilai accuracy 98,3%, precision 79% dan recall 87,17%.

Kata kunci: Analisis Sentimen, Decision Tree, Kesehatan Mental, K-Nearest Neighbors (K-NN), Naïve Bayes Classifier.

ABSTRACT

Twitter is the most popular social media platform on the internet and offers users the opportunity to express their opinions on a wide variety of topics, including news, current events, cabarets and other topics. One of them is mental health opinion. However, there are many different opinions on Twitter from different internet users. Therefore, it is necessary to carry out a basic analysis of public opinion to explain and provide new information on certain topics related to mental health, the method used is the Naïve Bayes Classifier algorithm and K-Nearest Neighbors (K-NN) and Decision Tree. A total of 5000 data were taken using the Twitter API with the keyword "Mental Health". Starting from the classification of positive or negative opinions, data cleansing, pre-processing, until the final results are obtained. Then it is calculated into two different algorithms to be compared, the algorithms used are Naïve Bayes Classifier and K-Nearest Neighbors (K-NN) and Decision Tree with the aim of finding the best accuracy. The highest result of this research is Naïve Bayes algorithm with 98.3% accuracy, 79% precision and 87.17% recall.

Keywords: *Sentiment Analysis, Decision Tree, Mental Health, K-Nearest Neighbors (K-NN), Naïve Bayes Classifier*

1. PENDAHULUAN

Salah satu aspek terpenting dalam mewujudkan kesehatan secara menyeluruh adalah Kesehatan mental. Kesehatan mental juga penting diperhatikan selayaknya Kesehatan fisik. WHO (World Health Organization) mendefinisikan bahwa Kesehatan mental sebagai suatu keadaan sejahtera dimana setiap individu menyadari potensinya sendiri, bisa menghadapi segala tekanan hidup yang normal, bisa melakukan kerja secara produktif dan bermanfaat, juga mampu memberikan kontribusi pada komunitasnya (Gutierrez et al., 2021).

Peningkatan waktu menggunakan gadget dan komputer diteliti dapat meningkatkan masalah kesehatan mental di antara anak-anak muda. Salah satu diantara menggunakan gadget yaitu Ketika bermain sosial media, Hal ini terkait dengan peningkatan screen time. Penelitian cross sectional yang dilakukan oleh (Tang et al., 2021) menunjukkan bahwa peningkatan screen time memiliki hubungan lemah dengan depresi dan kecemasan. Salah satu sosial media yang banyak dipakai anak muda yaitu twitter. Pertumbuhan pengguna Twitter di Indonesia sangat pesat dan menduduki peringkat 5 di dunia (Utami & Marzuki, 2020). Untuk memahami beberapa pengetahuan utama masyarakat tentang kesehatan mental pada periode saat ini, pendapat atau data tentang kesehatan mental dapat digunakan. Disinilah pentingnya penggunaan analisis sentimen masyarakat terhadap Kesehatan mental pada media sosial twitter. Analisis Sentimen merupakan proses pengelompokan teks ke dalam dokumen ataupun kalimat untuk menghasilkan dalam bentuk kelompok positif, negatif dan netral. (Suryono & Taufiq Luthfi, 2021) mengemukakan twitter adalah jejaring sosial yang memungkinkan pengguna mengirim karakter

yang sering disebut tweet atau kicauan. Setiap tweet tidak selalu mengandung makna atau arti yang sama, untuk itulah diperlukan suatu analisis opini masyarakat di twitter untuk mengklasifikasi negatif, positif dan netral dari tweet yang membahas "Kesehatan Mental". Dilakukan crawling data dengan memanfaatkan API twitter yakni dengan menggunakan analisis sentimen dalam proses klasifikasi komentarnya.

Data tweet yang telah diambil dari Twitter dapat dilakukan dengan mengintegrasikan Twitter API dan Rapid miner. Untuk memudahkan dalam mengidentifikasi jenis konten dari data twitter yang berjumlah besar, maka text mining harus diterapkan pada data tweet yang dimaksud dengan memanfaatkan teknik pembobotan kata, (Andryani et al., 2019)

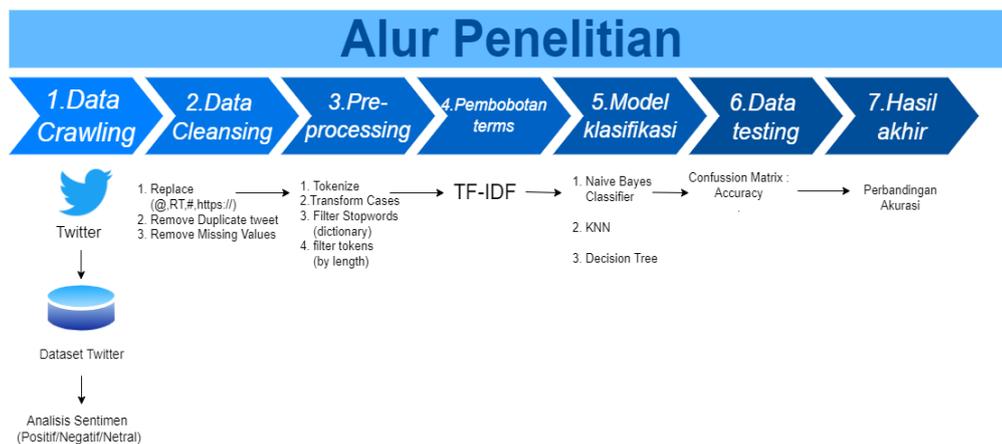
Beberapa penelitian yang telah dilakukan sebelumnya oleh Ivan dkk (Ivan et al., 2019) membuat sistem yang akan mampu mengklasifikasikan tweet pada twitter ke dalam kategori hate speech (HS) ataupun non-hates speech (NONHS). Penggunaan metode Naïve Bayes dengan seleksi fitur information gain dengan perbandingan 80% data latih dan 20% data uji. Hasil akurasi terbaik adalah sebesar 98%, dengan nilai precision 100%, nilai recall 96.15% dan nilai f-means 98.03%. Selanjutnya, berdasarkan studi literatur yang peneliti amati, penelitian mengenai Sentimen Analisis tentang Kesehatan Mental yang dilakukan oleh Karina Aulia dkk tentang: "Analisis Sentimen Twitter Pada Isu Mental Health Dengan Algoritma Klasifikasi Naive Bayes" (Aulia & Amelia, 2020) Dalam penelitian ini dapat diketahui bahwa peneliti menggunakan satu metode saja, yaitu metode Naïve Bayes. Penelitian lainnya yang

dilakukan oleh Rasyah Shafa dkk tentang Analisis Sentimen Terhadap Kesehatan Mental Selama Pandemi Covid-19 Berdasarkan Algoritma Naïve Bayes dan Deep Learning dapat diketahui bahwa peneliti menggunakan metode KNN dijelaskan pada penelitian ini juga hanya terdapat dua metode yang dilakukan oleh peneliti. (Azizah et al., 2023) Dari fenomena ini, peneliti memutuskan untuk melakukan analisis sentimen menggunakan opini masyarakat Indonesia di Twitter, yang nantinya opini tersebut diklasifikasikan menjadi data opini sentimen negatif dan positif. Penelitian ini menggunakan 3 (tiga) algoritma berbeda yaitu Naïve Bayes Classifier, K-Nearest Neighbors (K-NN) dan Decision Tree dikarenakan berdasarkan latar belakang yang dijabarkan penelitian terdahulu hanya menggunakan satu atau dua metode saja, sedangkan pada penelitian ini menggunakan tiga metode untuk mengekstrak pemikiran atau perasaan masyarakat pengguna twitter terhadap Kesehatan mental dan mengelompokkan datanya ke dalam kategori yang berbeda. Adapun tujuan dari penelitian ini yaitu untuk mengkomparasikan hasil dari ketiga algoritma tersebut untuk mengetahui classifier mana yang memberikan hasil terbaik dalam hal accuracy, precision, dan recall. Sehingga penelitian ini dapat menjadi dasar penelitian yang akan datang dengan metode yang paling tepat.

2. METODOLOGI

2.1. Metode yang Digunakan

Dalam penelitian ini metode yang digunakan yaitu metode klasifikasi dengan algoritma Naïve Bayes, KNN, dan Decision Tree. sebenarnya banyak metode klasifikasi seperti SVM, Random Forest, Neural Network, Support Vector Machine, dan Machine Learning (Wibowo & Winarko, 2014). Namun, peneliti hanya akan fokuskan pada 3 (tiga) metode itu saja. Metodologi yang digunakan dalam proses penelitian studi ini dijelaskan secara rinci di bagian metodologi penelitian. Berikut ini merupakan Gambar 1. Alur Penelitian.



Gambar 1. Alur Penelitian

2.2 Algoritma Naïve Bayes Classifier

Naïve Bayes termasuk algoritma Machine Learning yang masuk pada kategori Supervised Learning. Naïve Bayes adalah metode klasifikasi yang menggunakan probabilitas sebagai perhitungannya yang berakar pada Teorema Bayes dan memiliki asumsi ketidaktergantungan

(independen) yang tinggi di masing-masing kondisi atau kejadian (Sari, F. V., & Wibowo, 2019). Bentuk umum Teorema Bayes, dapat dilihat pada persamaan 1 di bawah ini:

$$P(H | X) = \frac{P(X | H) \cdot P(H)}{P(X)}$$

Keterangan :

- X : Data dengan class yang belum diketahui
- H : Hipotesis data X merupakan suatu class spesifik
- P (H | X) : Probabilitas hipotesis H berdasar kondisi X (posteriosi probability)
- P (H) : Probabilitas hipotesis H (prior probability)\
- P (X | H) : Probabilitas berdasarkan kondisi pada hipotesis H
- P (X) : Probabilitas X

2.3 Algoritma K-Nearest Neighbor

Algoritma K-Nearest Neighbor adalah metode klasifikasi terhadap suatu dataset berdasarkan jarak data pembelajaran (neighbor) terdekat. Jauh maupun dekatnya data pembelajaran (neighbor) tersebut dihitung dengan jarak Euclidean (Baharuddin et al., 2019)). Berikut Rumus dari K-Nearest Neighbor :

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2}$$

Keterangan :

- $d(x_i, x_j)$: Jarak Euclidean
- $(x_i)(x_i)$: record ke i
- $(x_j)(x_j)$: record ke j
- $(a_r)(a_r)$: data ke r
- $(i, j)(i, j)$: 1,2,3,...n

2.4 Algoritma Decision Tree

Algoritma Decision Tree merupakan algoritma yang memiliki struktur menyerupai diagram alir (flowchart) yang setiap node merupakan pengujian terhadap variabel atribut. Algoritma ini digunakan dalam analisa klasifikasi dan prediksi dalam bentuk pohon keputusan (Sutoyo, 2018) Beberapa pengembangan yang dilakukan menggunakan algoritma Decision Tree antara lain bisa mengatasi missing value, bisa mengatasi continue data, dan pruning (Azwanti, 2018) Berikut rumus yang tertera dalam persamaan :

$$Gain(S, A) = Entrophy(S) - \sum_{i=1}^n \frac{i_s}{s} * Entrophy(S_i)$$

Keterangan :

- S : Himpunan Khusus
- A : Atribut
- N : Jumlah partisi Atribut A

|S_i| : Jumlah Kasus pada partisi ke-i

|S| : Jumlah Kasus dalam S

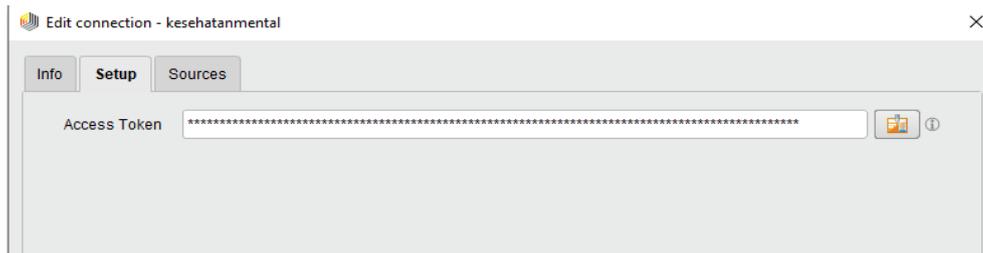
3. HASIL PENELITIAN DAN ANALISIS

3.1 Data Crawling

Pengumpulan data pada penelitian ini dilakukan menggunakan rapidminer. Dataset diperoleh menggunakan stream API Twitter yang disediakan oleh Twitter. Data training yang diujikan diambil di web resmi www.twitter.com dengan keys dan token sebagai authentication. dengan kata kunci "Kesehatan Mental" pada tanggal 29 bulan januari 2023, tweet yang berhasil ditarik disimpan dalam format csv.

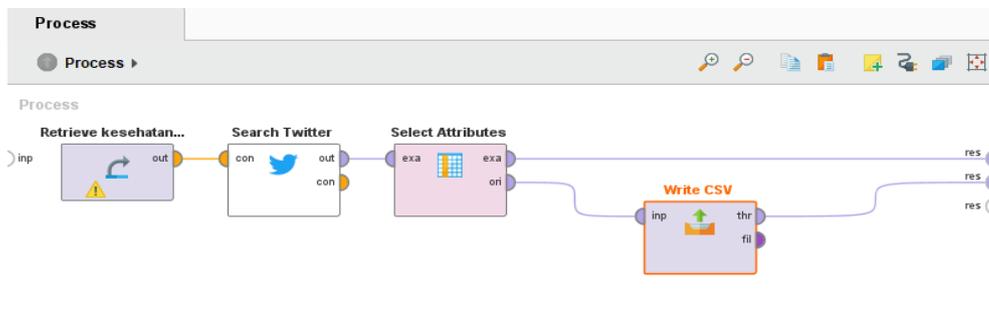


Gambar 2. Authentication



Gambar 3. Access Token

Pada pengambilan data atau yang disebut crawling pada penelitian ini menggunakan rapidminer yang ditunjukkan pada Gambar 4.3



Gambar 4. Crawling Data Twitter

Setelah semua komentar terkumpul, selanjutnya menyimpan hasil crawling data ke dalam file excel dengan format csv. Dalam crawling data, Langkah pertama adalah dengan dengan mengambil data pada twitter dan mencari kata kunci "Kesehatan Mental" yang dapat diambil dari media sosial twitter. Dari 5000 data sebanyak 3334 sampel yang didapatkan yang diambil secara acak. didapatkan hasil sebanyak 3334 data tweet. Setelah data didapatkan maka akan

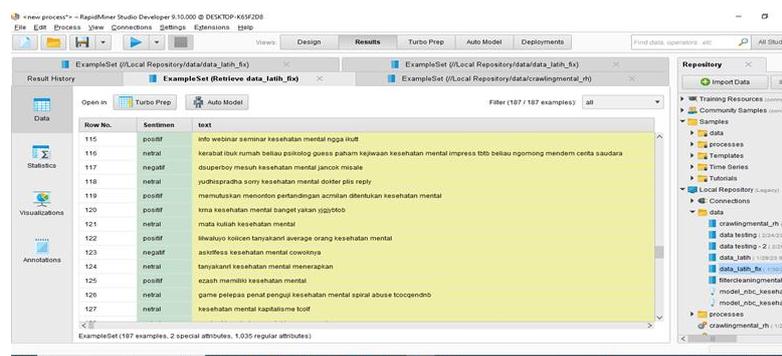
Komparasi Algoritma Naïve Bayes Classifier, K-Nearest Neighbor Dan Decision Tree Untuk Menganalisis Sentimen Masyarakat Terhadap Kesehatan Mental Pada Media Sosial Twitter

tersimpan dan di ekspor dalam bentuk csv. Langkah selanjutnya yaitu Select Attributes berfungsi untuk mengambil kolom mana saja yang ingin diambil atau untuk memfilter data text nya saja yang akan digunakan. Berikut gambar dari hasil data yang diambil text nya saja Pada penelitian ini kolom yang akan diambil hanya kolom Created-at dan Text, kolom Id secara otomatis akan terambil karena sudah bawaan dari RapidMiner. Setelah seluruh komentar berhasil disimpan selanjutnya dilakukan proses pelabelan. Dalam proses pelabelan yang dilakukan seharusnya untuk menentukan komentar yang didapatkan apakah bernilai positif, negatif, ataupun netral. Karena dilakukan secara manual, dalam proses pelabelan ini membutuhkan waktu yang tidak cepat, hal itu menjadi kelemahan apabila data yang harus dilabeli sangatlah banyak. Berikut ini merupakan Tabel sampel data yang digunakan beserta label sentimennya.

Tabel 1. Merupakan contoh sampel data yang digunakan beserta label sentimennya.

No	Text	Sentimen
1	yukeyurike nadachoiriyah tanyarlfees berhak bahagia berani memperjuangkan kebahagiaan mandiri financial	positif
2	detikcom penderita gangguan kesehatan mental berobat gratis bpjs simak syarat penggunaannya	positif
3	kalo ngmngin kehidupan indonesia inferior kemiskinan kesehatan mental tcoryajtjuv	negatif
4	radenrauf kesehatan mental hidup tenang	positif
5	gila kesehatan mental wkwk tcofljpn zayy	positif
6	askrl guys nanya tugas dapetin data kesehatan mental tingkat kota dapet dimana dinkes makasih	positif
7	penderita gangguan kesehatan mental berobat gratis bpjs simak syarat penggunaannya todpg	positif
8	sender neliti kesehatan mental dapetin data kesehatan mental tingkat kota dapet dimana dinkes makasih	positif
9	putellaque kesehatan mental membunuh mentalmu orang enaka	positif
10	askrlfess hadapi takut berani layaknya anak rantau kesehatan mental gimana fokus kerja kesehatan mental terganggu bahagia	positif

Hasil crawling data dan labelling data dapat dilihat pada gambar 5. dibawah ini

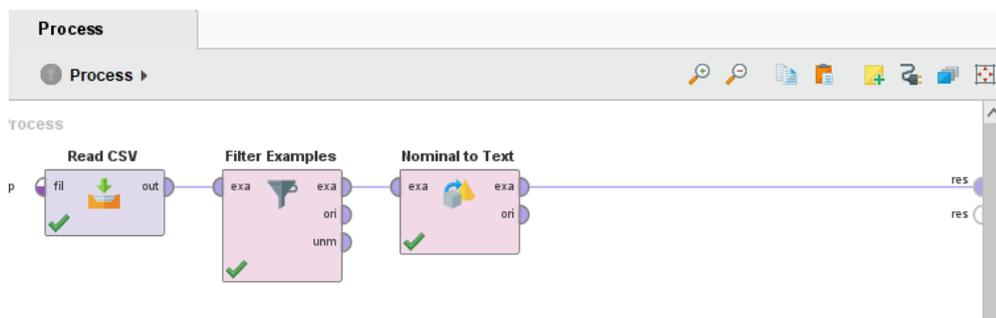


Gambar 5 . Data hasil Crawling dan Labelling

Dari 1053 data sebanyak 187 merupakan data yang telah dilabeli. Setelah data dilabeli, kemudian data yang sudah dilabeli di export lagi kedalam aplikasi rapidminer. Sisanya akan dijadikan sebagai data testing. Data testing nantinya akan kita perbandingkan dengan data yang telah dilabeli menggunakan 3 Algoritma yang berbeda yaitu Naïve Bayes Classifier, K-NN dan Decision Tree.

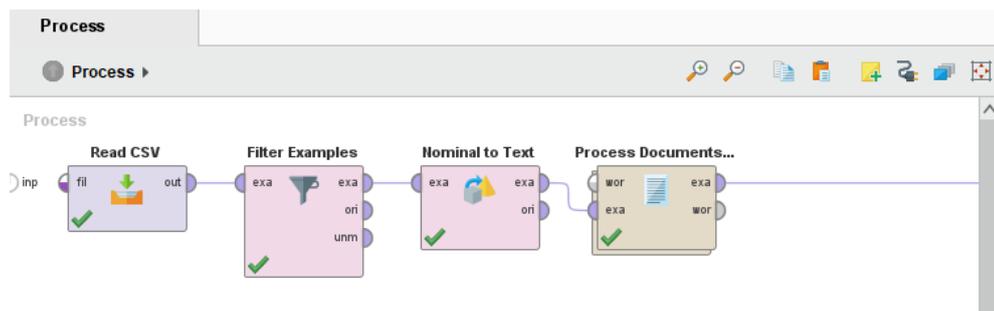
3.2 Pengolahan Data Cleansing

Dalam melakukan proses cleaning tersebut dilakukan 4 tahapan untuk mendapatkan hasil yang maksimal, di antaranya ialah Menghapus tanda baca, Menghapus angka, Menyeragamkan huruf menjadi huruf kecil semua, dan Menghapus kelebihan spasi. Dan hasilnya pada Gambar 6. Pengolahan Data Cleansing :

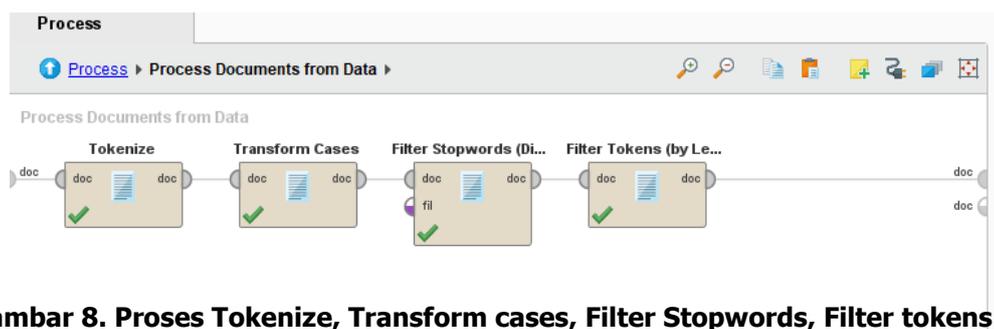


Gambar 6. Pengolahan Data Cleaning

3.3 Hasil Pre Reprocessing



Gambar 7. Proses Preprocessing



Gambar 8. Proses Tokenize, Transform cases, Filter Stopwords, Filter tokens

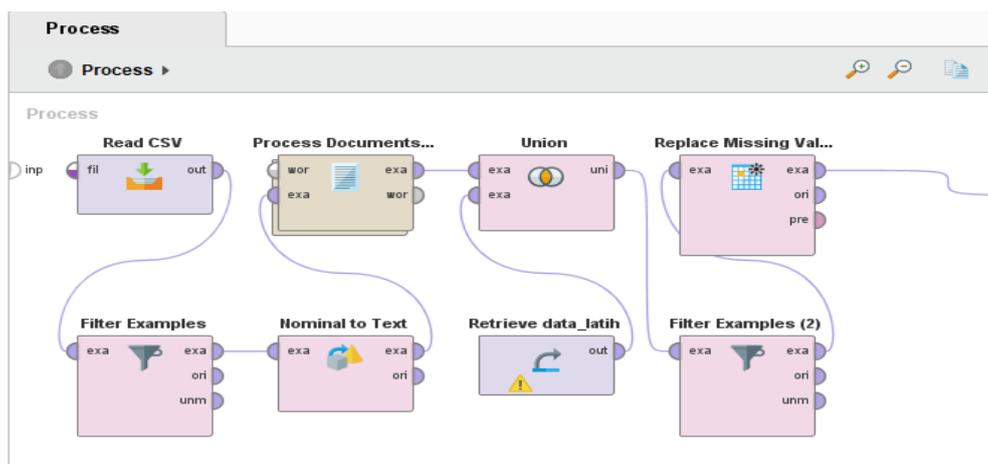
3.4 Pembobotan Kata

Kemudian dihitung menggunakan rumus TF-IDF sehingga menghasilkan nilai yang sudah terbobot. Adapun TF (Term Frequency) adalah frekuensi dari kemunculan sebuah term dalam dokumen yang bersangkutan, sedangkan IDF (Inverse Document Frequency) merupakan sebuah perhitungan dari bagaimana term didistribusikan secara luas pada koleksi dokumen yang bersangkutan. Proses pembobotan kata dilakukan dengan menghitung TF (Term Frequency) terlebih dahulu.

3.5 Model Klasifikasi Menggunakan *Software Rapidminer*

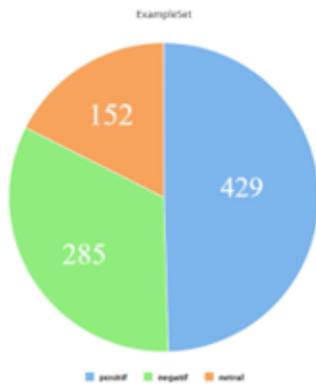
Tahap selanjutnya yaitu memasukkan data training dan data testing yang telah dilakukan preprocessing sebelumnya. Data training dan data testing yang didapatkan melalui twitter terkait Kesehatan mental adalah 187 data. Data ini telah difilter. Setelah dataset training dan testing sudah didapatkan, Langkah selanjutnya yaitu dilakukan proses ekstraksi fitur dan menerapkan metode klasifikasi algoritma naïve bayes classifier, KNN, dan Decision Tree.

Tahap ini merupakan tahap hasil dari uji Analisis Sentimen menggunakan software Rapidminer. Berikut Gambar 8. Proses Analisis Sentimen :

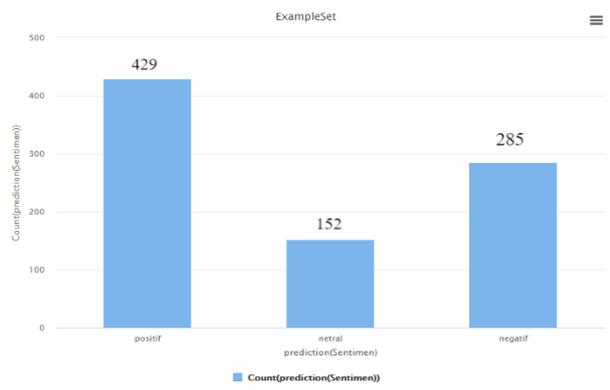


Gambar 8. proses Analisis Sentimen

Gambar 8 merupakan proses analisis sentimen. Data yang telah diambil secara acak sebanyak 5000 data, namun data yang di crawling oleh rapidminer pada 29 Januari 2023 pukul 11.19 Am. Data yang diperoleh selanjutnya dilakukan text preprocessing, agar noise pada text berkurang. Kemudian data dikategorikan menjadi tiga kategori, yaitu opini positif, negatif dan netral. Data yang diperoleh sebanyak 3334 data, setelah dilakukan proses filtering, data yang didapatkan sebanyak 1053 data. 866 merupakan data testing, dan sisanya 187 merupakan data training yang telah dilabeli. Dari 866 tweet data testing, dengan perhitungan opini positif sebanyak 429 sedangkan opini negatif sebanyak 285 opini, dan netral sebanyak 152 opini. Secara keseluruhan, pengklasifikasian data tanggapan yang diperoleh dapat dilihat pada Gambar 9. Pie Chart Jumlah Sentimen dan Gambar 10. Bar Chart Jumlah Sentimen :



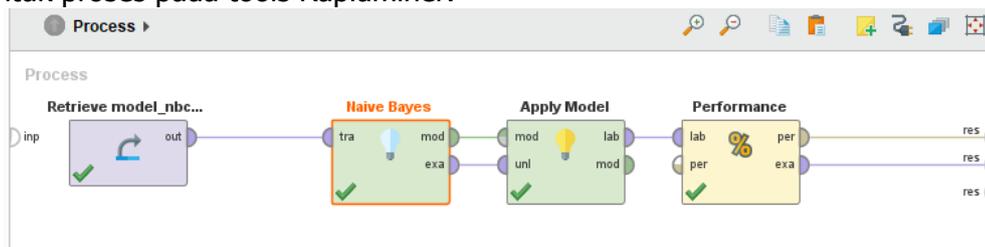
Gambar 9. Pie Chart Jumlah Sentimen



Gambar 10. Bar chart Jumlah Sentimen

3.6 Model Klasifikasi Naïve Bayes Classifier

Setelah dataset training dan testing sudah didapatkan, Langkah selanjutnya yaitu dilakukan proses ekstraksi fitur dan menerapkan metode klasifikasi algoritma naïve bayes classifier. Dengan menggunakan pemodelan klasifikasi Naïve Bayes seperti gambar diatas dengan mengklasifikasi 2 jenis yaitu positif negatif dan netral. Dalam penelitian ini algoritma menghasilkan akurasi 98.40% dengan menggunakan metode Naïve Bayes dari pengujian data yang telah dikumpulkan dan dengan diuji dengan data training dan data testing. Berikut ini gambar untuk proses pada tools Rapidminer.



Gambar 11. Model Klasifikasi Algoritma Naïve Bayes

Dari pengujian hasil data tersebut dengan rapid miner akan mendapatkan class precision dan class recall, yang dimaksud dengan precision yaitu perbandingan dokumen yang relevan terhadap total dokumen hasil query sedangkan recall perbandingan jumlah dokumen relevan terambil terhadap total dokumen relevan. Selain accuracy ada dua hal tersebut yang dihasilkan dalam klasifikasi teks, maka pengujian data dapat dihasilkan pada Gambar 12. :

accuracy: 98.40%

	true positif	true negatif	true netral	class precision
pred. positif	105	0	0	100.00%
pred. negatif	2	51	0	96.23%
pred. netral	0	1	28	96.55%
class recall	98.13%	98.08%	100.00%	

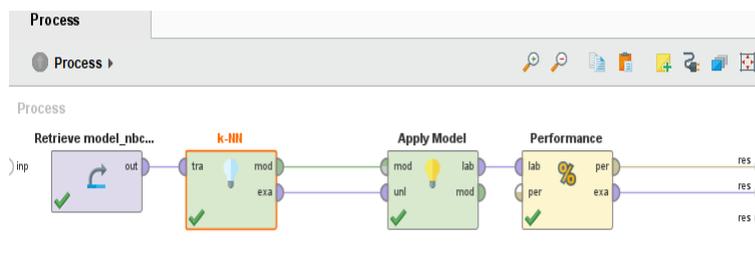
Gambar 12. Hasil Pengujian Rapidminer

Proses desain pada gambar 4.19. menunjukkan hasil pengujian algoritma Naïve Bayes Classifier didapat akurasi sebesar 98.40 %. Dan class precision untuk pred. positif adalah 100.00 %, pred.negatif adalah 96.23 %, pred. netral adalah 96.55 %, selain itu juga terdapat class recall untuk true positif 98.13 %, true negatif 98.08 % dan true netral 100.00 %. Naïve

bayes merupakan salah satu algoritma terbaik untuk melakukan klasifikasi berdasarkan beberapa penelitian sebelumnya.

3.7 Model Klasifikasi Algoritma K-Nearest Neighbors (K-NN)

Pada penelitian ini telah dilakukan pengujian pada Algoritma K-Nearest Neighbors (K-NN) dan mendapatkan hasil sebagai berikut seperti pada gambar dibawah ini :



Gambar 13. Model Klasifikasi Algoritma K-NN

Proses Pengujian Algoritma K-Nearest Neighbors (K-NN) didapatkan hasil akurasi sebesar 77.01 %.

Table View Plot View

accuracy: 77.01%

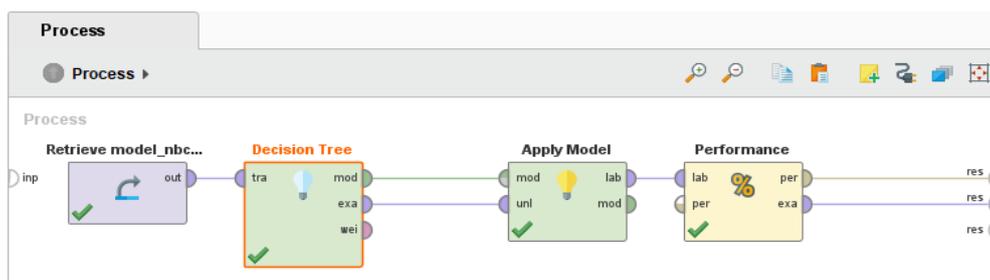
	true positif	true negatif	true netral	class precision
pred. positif	99	23	9	75.57%
pred. negatif	3	28	2	84.85%
pred. netral	5	1	17	73.91%
class recall	92.52%	53.85%	60.71%	

Gambar 14. Hasil Pengujian Algoritma Pada Rapidminer

Proses desain pada Gambar 14. menunjukkan hasil pengujian algoritma K-Nearest Neighbors (K-NN) didapat akurasi sebesar 77.01 %. Dan class precision untuk pred. positif adalah 75.57.00 %, pred.negatif adalah 84.85 %, pred. netral adalah 73.91 %, selain itu juga terdapat class recall untuk true positif 92.52 %, true negatif 53.85 % dan true netral 60.71 %.

3.8 Model Klasifikasi Algoritma Decision Tree

Kemudian algoritma yang ketiga adalah Decision Tree. Decision tree juga merupakan salah satu algoritma terbaik untuk klasifikasi. Berikut Gambar 4.28 merupakan Proses Desain Algoritma Decision Tree menggunakan Rapidminer.



Gambar 15. Model Klasifikasi Algoritma Decision Tree

Proses Pengujian Algoritma K-Nearest Neighbors (K-NN) pada Gambar 16. didapatkan hasil akurasi sebesar 62.03 %.

Table View
 Plot View

accuracy: 62.03%

	true positif	true negatif	true netral	class precision
pred. positif	107	52	19	60.11%
pred. negatif	0	0	0	0.00%
pred. netral	0	0	9	100.00%
class recall	100.00%	0.00%	32.14%	

Gambar 16. Hasil Pengujian Algoritma Decision Tree Pada Rapidminer

Proses desain pada Gambar 4.29 menunjukkan hasil pengujian algoritma Decision Tree didapat akurasi sebesar 60.03 %. Dan class precision untuk pred. positif adalah 60.11 %, pred.negatif adalah 0.00 %, pred. netral adalah 100.00 %, selain itu juga terdapat class recall untuk true positif 100.00 %, true negatif 0.00% dan true netral 32.14 %.

3.9 Hasil Akhir Perbandingan Algoritma

Langkah terakhir yang dilakukan setelah mendapatkan hasil Dari ketiga metode yang digunakan, yaitu Naïve Bayes Classifier, K-Nearest Neighbor (K-NN) dan Decision Tree adalah membandingkan atau mengkomparasikan ketiga algoritma tersebut, akurasi tertinggi didapat menggunakan algoritma Naïve Bayes Classifier. Berikut ini Tabel 4.2, menampilkan hasil Accuracy, Precision, dan Recall dari Setiap Algoritma.

Tabel 4.2 Perbandingan Ketiga Algoritma

Algoritma	Accuracy	Precision	Recall
<i>Naïve Bayes Classifier</i>	98.40 %	97.59 %	98.73 %
K-NN	77.01 %	78.11 %	69.02 %
<i>Decision Tree</i>	62.03 %	53.37 %	44.06 %

Hasil ini menunjukkan accuracy dari algoritma Naïve Bayes Classifier adalah 98.40 %, algoritma K-Nearest Neighbors (K-NN) adalah 77.01 % dan Decision Tree adalah 62.03 %, untuk hasil precision algoritma Naïve Bayes Classifier, K-NN, dan Decision Tree adalah 97.59 %, 78.11 % dan 53.37%, sementara Recall dari algoritma tersebut adalah 98.73 %, 69.02 % dan 44.06%. sehingga dapat ditarik kesimpulan bahwa Algoritma Naïve Bayes Classifier adalah algoritma terbaik untuk digunakan dengan dataset Kesehatan mental karena memberikan prediksi yang lebih tepat dan lebih akurat dibandingkan dengan algoritma K-Nearest Neighbors (K-NN) dan Decision Tree hal ini berdasarkan nilai akurasi, presisi dan recall yang lebih tinggi dibandingkan algoritma K-Nearest Neighbors (K-NN) dan Decision Tree.

3. KESIMPULAN

Berdasarkan analisis dan pembahasan yang telah dilakukan di bab sebelumnya, maka diperoleh kesimpulan sebagai berikut :

1. Hasil analisis sentimen terhadap data twitter mengenai Kesehatan mental adalah Dari 866 tweet data testing, dengan perhitungan opini positif sebanyak 429 sedangkan opini negatif sebanyak 285 opini, dan netral sebanyak 152 opini dapat dilihat bahwa jumlah tanggapan yang memiliki sentimen positif lebih besar daripada tanggapan yang memiliki sentimen negatif, artinya Sebagian besar masyarakat pengguna twitter menanggapi Kesehatan mental dengan positif. . Selain itu, penelitian ini dapat memberikan warning untuk masyarakat yang tidak memiliki gangguan mental agar tidak memberikan judgment yang dapat menjatuhkan dan memperburuk mental orang lain.
2. Algoritma Naïve Bayes Classifier adalah algoritma terbaik untuk digunakan dengan dataset Kesehatan mental karena memberikan prediksi yang lebih tepat dan lebih akurat dibandingkan dengan algoritma K-Nearst Neighbors (K-NN) dan Decision Tree hal ini berdasarkan nilai akurasi, presisi dan recall yang lebih tinggi dibandingkan algoritma K-Nearest Neighbors (K-NN) dan Decision Tree dimana mendapatkan hasil Accuracy sebesar 98.40 %, recall 97.59 %, dan precision 98.73 %, disusul oleh algoritma K-Nearest Neighbors (K-NN) mendapatkan hasil Accuracy sebesar 77.01 %, recall 78.11 %, dan precision 69.02 % dan yang terakhir algoritma Decision Tree mendapatkan hasil Accuracy sebesar 62.03 %, recall 53.37 % dan precision 44.06 % .

5. DAFTAR RUJUKAN

- Andryani, R., Negara, E. S., & Triadi, D. (2019). Social Media Analytics: Data Utilization of Social Media for Research. *Journal of Information Systems and Informatics*, 1(2), 193–205. <https://doi.org/10.33557/journalisi.v1i2.23>
- Aulia, K., & Amelia, L. (2020). Analisis Sentimen Twitter Pada Isu Mental Health Dengan Algoritma Klasifikasi Naive Bayes. *Siliwangi Journal (Seri Sains and Teknologi)*, 6(2), 60–65.
- Azizah, R. S., Kamayani, M., Studi, P., & Informatika, T. (2023). *Analisis Sentimen Terhadap Kesehatan Mental Selama Pandemi Covid-19 Berdasarkan Algoritma Naïve Bayes dan Deep Learning*. 23(6), 38–43.
- Azwanti, N. (2018). Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada Pt.

Capella Dinamik Nusantara Cabang Muka Kuning. *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, 13(1), 33. <https://doi.org/10.30872/jim.v13i1.629>

Baharuddin, M. M., Azis, H., & Hasanuddin, T. (2019). Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca. *ILKOM Jurnal Ilmiah*, 11(3), 269–274. <https://doi.org/10.33096/ilkom.v11i3.489.269-274>

Gutierrez, L. J., Rabbani, K., Ajayi, O. J., Gebresilassie, S. K., Rafferty, J., Castro, L. A., & Banos, O. (2021). Internet of things for mental health: Open issues in data acquisition, self-organization, service level agreement, and identity management. *International Journal of Environmental Research and Public Health*, 18(3), 1–19. <https://doi.org/10.3390/ijerph18031327>

Ivan, Arum Sari, Y., & Adikara, P. P. (2019). *Klasifikasi Hate Speech Berbahasa Indonesia di Twitter Menggunakan Naive Bayes dan Seleksi Fitur Information Gain dengan Normalisasi Kata Eye Movement Detection View project Image Processing View project*. May. <https://www.researchgate.net/publication/334194823>

Sari, F. V., & Wibowo, A. (2019). Analisis Sentimen Pelanggan Toko Online Jd. Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 2(2), 681–686.

Sari, R. (2020). Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (K-Nn). *EVOLUSI: Jurnal Sains Dan Manajemen*, 8(1), 10–17. <https://doi.org/10.31294/evolusi.v8i1.7371>

Suryono, S., & Taufiq Luthfi, E. (2021). Analisis sentimen pada Twitter dengan menggunakan metode Naïve Bayes Classifier. *Jnanaloka*, 81–86. <https://doi.org/10.36802/jnanaloka.2020.v1-no2-81-86>

Sutoyo, I. (2018). Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik. *Jurnal Pilar Nusa Mandiri*, 14(2), 217. <https://doi.org/10.33480/pilar.v14i2.926>

Tang, S., Werner-Seidler, A., Torok, M., Mackinnon, A. J., & Christensen, H. (2021). The

relationship between screen time and mental health in young people: A systematic review of longitudinal studies. *Clinical Psychology Review*, 86, 102021. <https://doi.org/10.1016/j.cpr.2021.102021>

Utami, I., & Marzuki, M. (2020). Analisis Sistem Informasi Banjir Berbasis Media Twitter. *Jurnal Fisika Unand*, 9(1), 67–72. <https://doi.org/10.25077/jfu.9.1.67-72.2020>

Wibowo, A., & Winarko, E. (2014). Paper Review: Data Mining Twitter. *Maintaining Cultural Heritage Through Information Technology for a Smart Future, November 2014*, 1–10. https://www.researchgate.net/publication/329207488_Paper_Review_Data_Mining_Twitter