

KOMPARASI ALGORITMA NAÏVE BAYES CLASSIFIER, K-NEAREST NEIGHBOR (KNN) DAN DECISION TREE UNTUK MENGANALISIS SENTIMEN MASYARAKAT TERHADAP KESEHATAN MENTAL PADA MEDIA SOSIAL TWITTER



Oleh :

TACHIYYA NAILAL KHUSNA

NIM : 21206051002

STATE ISLAMIC UNIVERSITY
PROGRAM STUDI INFORMATIKA
PROGRAM MAGISTER FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA

2023



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI

Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-733/Un.02/DST/PP.00.9/03/2023

Tugas Akhir dengan judul : KOMPARASI ALGORITMA NAIVE BAYES CLASSIFIER, K NEAREST NEIGHBOR (KNN) DAN DECISION TREE UNTUK MENGANALISA SENTIMEN MASYARAKAT TERHADAP KESEHATAN MEMTAL PADA MEDIA SOSIAL TWITTER

yang dipersiapkan dan disusun oleh:

Nama : TACHIYYA NAILAL KHUSNA, S.Kom
Nomor Induk Mahasiswa : 21206051002
Telah diujikan pada : Jumat, 10 Maret 2023
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang

Dr. Ir. Bambang Sugiantoro, S.Si., M.T.
SIGNED

Valid ID: 64118886d2d75



Penguji I

Ir. Muhammad Taufiq Nuruzzaman, S.T.
M.Eng., Ph.D.
SIGNED

Valid ID: 64116d8bc47d5



Penguji II

Dr. Sugiyanto, S.Si., ST., M.Si.
SIGNED

Valid ID: 64118686dac45



Yogyakarta, 10 Maret 2023
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi

Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 6412a80610561

PERNYATAAN KEASLIAN

Yang bertandatangan di bawah ini:

Nama : Tachiyya Nailal Khusna
NIM : 21206051002
Jenjang : Magister
Program Studi : Informatika

Menyatakan bahwa naskah tesis ini dengan judul “Komparasi Algoritma Naive Bayes Classifier, K Nearest Neighbor (KNN) Dan Decision Tree Untuk Menganalisis Sentimen Masyarakat Terhadap Kesehatan Mental Pada Media Sosial Twitter” tidak terdapat pada karya yang pernah diajukan untuk memperoleh gelar Magister di suatu Perguruan Tinggi, dan sepengetahuan saya tidak terdapat karya atau pendapat yang pernah ditulis oleh orang lain, kecuali pada bagian-bagian yang dirujuk sumbernya.

Yogyakarta, 8 Maret 2023

Saya yang menyatakan,



Tachiyya Nailal Khusna

NIM: 21206051002

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

PERNYATAAN BEBAS PLAGIASI

Yang bertandatangan di bawah ini:

Nama : Tachiyya Nailal Khusna
NIM : 21206051002
Jenjang : Magister
ProgramStudi : Informatika

Menyatakan secara keseluruhan bahwa naskah tesis ini secara keseluruhan benar-benar bebas dari plagiasi. Jika dikemudian hari saya terbukti melakukan plagiasi, maka saya siap ditindak sesuai ketentuan hukum yang berlaku.

Yogyakarta, 8 Maret 2023

Saya yang menyatakan,



Tachiyya Nailal Khusna
NIM : 21206051002

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA



Universitas Islam Negeri Sunan Kalijaga

SURAT PERSETUJUAN SKRIPSI/TUGAS
AKHIR

Hal : Persetujuan Tugas
AkhirLamp : -

Kepada
Yth. Dekan Fakultas Sains dan Teknologi
UINSunan Kalijaga di Yogyakarta

Assalamualaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakanperbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa tugas tesis Saudara :

Nama : Tachiyya Nailal Khusna
NIM : 21206051002
Judul Tesis : Komparasi Algoritma Naïve Bayes Classifier, K Nearest Neighbor (KNN) Dan Decision Tree Untuk Menganalisis Sentimen Masyarakat Terhadap Kesehatan Mental Pada Media Sosial Twitter

Saya berpendapat bahwa tesis tersebut sudah dapat diajukan kepada Program Studi Magister Informatika UIN Sunan Kalijaga untuk diajukan dalam rangka memperoleh gelar Magister Informatika.

Wassalamualaikum wr. wb.

Yogyakarta, 8 Maret 2023
Pembimbing,

Dr. Ir. Bambang Sugiantoro, S.Si., M.T.
NIP. 19751024 200912 1 002

ABSTRAK

Twitter merupakan salah satu dari sekian banyaknya media sosial yang sering digunakan orang untuk mengekspresikan pendapat mereka. *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. Isu tentang Kesehatan Satu-satunya masalah yang signifikan di dunia adalah mental. Penurunan Jumlah atau Angka Penderita Penyakit Mental Berhasil Dihilangkan Ketika Masyarakat Berkontribusi Untuk Tidak Mendiskriminasi Penderita. Namun, ada banyak opini yang berbeda di *Twitter* dari pengguna internet yang berbeda, dan salah satu yang paling menonjol adalah opini tentang kesehatan mental. 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 Clasifier* dan *K-Nearest Neighbors* (K-NN). 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 kedalam dua algoritma yang berbeda untuk dibandingkan, algoritma yang digunakan yaitu *Naïve Bayes Classifier* dan *K-Nearest Neighbors* (K-NN) dengan tujuan menemukan akurasi terbaik. Aplikasi Rapidminer Versi 7.1 juga digunakan untuk mempermudah penulis dalam mengolah data. Hasil tertinggi dari penelitian ini adalah algoritma *Naïve Bayes* dengan nilai accuracy 98,3%, precision 79% dan recall 87,17%.

Kata Kunci : Analisis Sentimen, Kesehatan Mental, *K-Nearest Neighbors* (K-NN), *Naïve Bayes Classifier*, *Twitter*.

ABSTRACT

Twitter is one of the many social media platforms that people often use to express their opinions. Twitter is the most popular social media platform on the internet and offers users the opportunity to express their opinions on a variety of topics, including news, current events, cabaret and other topics. One such topic is opinions on mental health. Issues on Health The only significant problem in the world is mental. A decrease in the number or rate of mental illnesses can be successfully eliminated when society contributes to not discriminating against sufferers. However, there are many different opinions on Twitter from different internet users, and one of the most prominent is the opinion on mental health. Therefore, it is necessary to conduct a basic analysis of public opinion to explain and provide new information on certain topics related to mental health, the methods used are Naïve Bayes Classifier and K-Nearest Neighbors (K-NN) algorithms. 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, preprocessing, until the final result is obtained. Then calculated into two different algorithms to compare, the algorithms used are Naïve Bayes Classifier and K-Nearest Neighbors (K-NN) with the aim of finding the best accuracy. The Rapidminer Version 7.1 application is also used to facilitate the author in processing data. The highest result of this research is the Naïve Bayes algorithm with an accuracy value of 98.3%, precision 79% and recall 87.17%.

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

MOTTO

*Karena sesungguhnya sesudah kesulitan itu ada kemudahan,
sesungguhnya sesudah kesulitan itu ada kemudahan*

QS Al Insyirah 5-6



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN PERSEMBAHAN

Tesis ini dipersembahkan penulis untuk :

- Orang tua saya, Ahmad Jazuli, M.Kom dan Nur Hidayah tercinta yang selalu berkerja keras dan membesarkan saya sampai dengan sekarang, tesis ini sebagai tanda bahwa perjuangan orang tua saya untuk memberikan Pendidikan tinggi untuk anaknya tidak sia-sia.
- Untuk Almamater Tercinta Universitas Islam Negeri Sunan Kalijaga Yogyakarta Program Studi Magister Informatika.



KATA PENGANTAR

Alhamdulillah, segala puji syukur kehadiran Allah SWT., karena berkat Rahmat serta Hidayah-Nya, penulis berhasil menyelesaikan tesis berjudul “Komparasi Algoritma *Naïve Bayes Classifier*, Algoritma *K-Nearest Neighbor* (KNN) dan Decision Tree Untuk Menganalisis Sentimen Masyarakat Terhadap Kesehatan Mental Pada Media Sosial Twitter”. Shalawat serta salam penulis haturkan kepada junjungan kami Nabi Muhammad SAW., yang telah membawa kita dari zaman kegelapan kepada zaman yang terang benderang seperti saat ini.

Penyusunan Tesis ini ditujukan untuk memenuhi salah satu syarat diperolehnya gelar Magister Komputer (M.Kom). Selain itu, pengerjaan serta pelaksanaan Tesis ini tak lepas dari bantuan dan dukungan beberapa pihak, untuk itu penulis menyampaikan ucapan terima kasih kepada:

1. Bapak Prof. Dr. Phil. Al Makin, S. Ag., M.A., selaku Rektor UIN Sunan Kalijaga Yogyakarta.
2. Ibu Dr. Dra. Hj. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga..
3. Bapak Dr. Bambang Sugiantoro, S.Si., M.T., selaku ketua program studi Magister Informatika UIN Sunan Kalijaga Yogyakarta sekaligus Dosen Pembimbing Akademik dan Pembimbing Tesis yang sudah banyak membantu penulis mulai dari awal penulis menempuh pendidikan magister hingga sekarang serta memeberikan ilmu, waktu, dan kemudahan dalam bimbingan selama penyusunan tesis ini.
4. Kedua orang tua tercinta, tersayang dan terkasih serta adik saya Muhammad Riqza Zulmi yang senantiasa berdoa untuk keberhasilan dalam menuntut ilmu dan memberikan segala dukungan kepada penulis dalam menyelesaikan Tesis ini.

5. Dan untuk orang-orang baik, Fadilla Amanda, Madiana Nur, Sahabat adikku Riqza yang banyak membantu, memberikan dukungan dalam melakukan penelitian serta penulisan Tesis ini,

Penulis menyadari bahwa masih terdapat kekurangan dan ketidaktepatan dalam skripsi ini, oleh karena itu penulis mengharapkan masukan, saran, dan kritik dari para pembaca agar penulis dapat menjadi lebih baik lagi di masa yang akan datang. Akhir kata, penulis berharap buku ini dapat bermanfaat bagi semua orang yang membacanya, khususnya bagi penulis lain dan pembaca pada umumnya.

Yogyakarta, 18 Maret 2023

Penulis

Tachiyya Nailal Khusna



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR ISI

HALAMAN PENGESAHAN	ii
PERNYATAAN KEASLIAN	iii
PERNYATAAN BEBAS PLAGIASI	iv
SURAT PERSETUJUAN TESIS/TUGAS AKHIR	v
ABSTRAK	vi
ABSTRACT	vii
MOTTO	viii
HALAMAN PERSEMBAHAN	ix
KATA PENGANTAR	x
DAFTAR ISI	xii
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvii
DAFTAR SINGKATAN	xviii
BAB I PENDAHULUAN	1
A. Latar Belakang	1
B. Rumusan Masalah	4
C. Batasan Masalah	5
D. Tujuan Penelitian	5
BAB II KAJIAN PUSTAKA DAN LANDASAN TEORI	7
A. Tinjauan Pustaka	7
B. Landasan Teori	12
1. Pengertian Analisis Sentimen	12
2. Kesehatan Mental	12
3. Pengertian Komparasi	13
4. Pengertian Sosial Media Twitter	13
5. Rapidminer	14

6. <i>Text Mining</i>	15
7. <i>Text Pre-Processing</i>	15
8. Pembobotan Kata.....	16
9. Algoritma <i>Naïve Bayes Classifier</i>	16
10. Algoritma <i>K-Nearest Neighbor</i>	17
11. Algoritma <i>Decision Tree</i>	17
12. Evaluasi.....	18
BAB III METODE PENELITIAN	21
A. Metode yang Digunakan	21
1. <i>Data Crawling</i>	21
2. <i>Search Twitter</i>	22
3. Select Atributte	23
4. Write CSV	24
5. Pengolahan Data <i>Cleansing</i>	25
6. <i>Pre-Processing</i>	26
7. Tokenize	27
8. Transform Cases	27
9. Filter Stopwords.....	28
10. Filter Tokens (by Length).....	29
11. Pembobotan <i>Terms</i>	30
12. Tahap Klasifikasi	30
13. Tahap Testing	30
14. Evaluasi.....	30
BAB IV HASIL DAN PEMBAHASAN	32
A. Hasil Penelitian	32
1 Hasil <i>Crawling Data</i>	32
2 Hasil Pengolahan Data <i>Cleansing</i>	36
3 Hasil <i>Pre Reprocessing</i>	38

4 Hasil Pembobotan Kata.....	38
5. Model Klasifikasi Menggunakan <i>Software</i> Rapidminer	39
a. Model Klasifikasi Naïve Bayes Classifier	53
b. Model Klasifikasi Algoritma <i>K-Nearest Neighbors (K-NN)</i>	55
c. Model Klasifikasi Algoritma <i>Decision Tree</i>	56
6. Hasil Akhir Perbandingan Algoritma.....	58
BAB V KESIMPULAN DAN SARAN	60
A. KESIMPULAN	60
B. SARAN.....	61
DAFTAR PUSTAKA.....	62
LAMPIRAN.....	66



DAFTAR GAMBAR

Gambar 2.1 Rapidminer	13
Gambar 3.1 Alur Penelitian	20
Gambar 3.2 Opini Tentang Kesehatan Mental	21
Gambar 3.3 Opini Tentang Kesehatan Mental	21
Gambar 3.4 Opini Tentang Kesehatan Mental	22
Gambar 3.5 Parameter Search	22
Gambar 3.6. Parameter Select Attributes	22
Gambar 3.7 Parameters Write CSV	23
Gambar 3.8 <i>Pre-Processing</i>	24
Gambar 3.9 Tabel <i>Confussion Matrix</i>	24
Gambar 4.1. Authentication	31
Gambar 4.2. Access Token	31
Gambar 4.3. Data Hasil Crawling	32
Gambar 4.4. Hasil Crawling Data menggunakan Rapidminer	32
Gambar 4.5. Hasil Crawling data bentuk CSV	34
Gambar 4.6. Sebelum Select Attribute	34
Gambar 4.7. Sebelum Select Attribute	34
Gambar 4.8 Data Hasil Crawling dan Labelling	35
Gambar 4.9. Export Data ke Rapidminer	35
Gambar 4.10. Pengolahan Data Cleaning	36
Gambar 4.11. Hasil Proses Cleaning	36
Gambar 4.12. Proses Pre Reprocessing	37
Gambar 4.13. Proses Tokenize, Transform Cases, Filter Stopwords, Filter Tokens	37

Gambar 4.14 Hasil Proses Tokenize, Transform Cases, Filter Stopwords, Filter Tokens	37
Gambar 4.15. Hasil Pembobotan Terms.....	38
Gambar 4.16. Hasil Pembobotan Terms.....	38
Gambar 4.17 Proses Analisis Sentimen	49
Gambar 4.18. Piechart Jumlah Sentimen.....	50
Gambar 4.19. Barchart Jumlah Sentimen.....	50
Gambar 4.20. Perolehan Hasil Jumlah Sentimen pada Rapidminer..	51
Gambar 4.21. Visualisasi Kata Terpopuler dengan WordCloud	51
Gambar 4.22. Model Klasifikasi Algoritma Naïve Bayes Classifier ...	52
Gambar 4.23. Hasil Pengujian Rapidminer	52
Gambar 4.24. Deskripsi Performance Vector Algoritma Naïve Bayes Classifier	53
Gambar 4.25. Model Klasifikasi Algoritma K-NN	54
Gambar 4.26 Hasil Pengujian Algoritma pada Rapidminer	54
Gambar 4.27 Deskripsi Performance Vector Algoritma K-NN	55
Gambar 4.28 Model Klasifikasi Algoritma Decision Tree	55
Gambar 4.29 Hasil Pengujian Algoritma Decision Tree pada Rapidminer	56
Gambar 4.30 Deskripsi Performance Vector Algoritma.....	56

DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian Terkait	8
Tabel 2. 2 Contoh Tabel <i>Confussion Matrix</i>	18
Tabel 3. 1. Contoh <i>Cleaning</i> pada <i>Tweet</i>	27
Tabel 3. 2 Contoh <i>tokenizing</i> pada <i>tweet</i>	29
Tabel 3.3. Contoh <i>Transform Case</i> pada <i>tweet</i>	30
Tabel 3. 4. Contoh <i>Filter Stopwords</i> pada <i>tweet</i>	31
Tabel 4.1. merupakan sampel data yang digunakan	41
Tabel 4.2 Perbandingan	59

DAFTAR SINGKATAN

- CSV : *Comma Separated Value*
SVM : *Support Vector Machine*
API Twitter : *Application Programming Interface Twitter*



BAB I

PENDAHULUAN

A. Latar Belakang

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 (I. 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 sentiment 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, negative dan netral. (Suryono et al., 2018) mengemukakan *twitter* adalah jejaring sosial yang memungkinkan pengguna mengirim karakter

yang sering disebut *twit* atau kicauan. Setiap *twit* tidak selalu mengandung makna atau arti yang sama, untuk itulah diperlukan suatu analisis opini masyarakat di *twitter* untuk mengklasifikasi negative, positif dan netral dari *twit* yang membahas “Kesehatan Mental”. Dilakukan crawling data dengan memanfaatkan API twitter yakni dengan menggunakan analisa sentiment 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 yang dilakukan oleh Denty Welmin dkk tentang: “Anaysis of User Sentiment of Twitter to RUU KUHP”.(Cahyaningrum et al., 2020). Dalam penelitian ini dapat diketahui bahwa peneliti menggunakan R studio untuk bahasa pemrograman. Akan tetapi penelitian ini tidak menjelaskan secara spesifik metode apa yang digunakan dalam penelitian tersebut dan hanya menggunakan satu metode saja. Penelitian lainnya yang dilakukan oleh Retno Sari tentang Analisis Sentimen dapat diketahui bahwa peneliti menggunakan metode KNN. Akan tetapi tidak dijelaskan menggunakan tools apa dalam penelitiannya. Dan

dalam penelitian ini juga hanya terdapat satu metode yang dilakukan oleh peneliti. (Sari, 2020).

Analisis sentimen memiliki kelebihan mengidentifikasi sentimen yang tidak secara eksplisit tertuang di dalam kalimat opini (Firdaus et al., 2021) serta dapat menemukan makna tiap kata pada proses klasifikasi sentimen (Khoo & Johnkhan, 2018). Analisis sentiment lebih baik daripada dilakukan dengan menyebar survey atau quisioner karena tidak memerlukan responden . selain itu pengisian survey atau quisioner terkadang tidak sesuai dengan kenyataan yang sebenarnya dan cenderung melebih-lebihkan jawaban hasil survey ataupun quisioner tersebut, Penelitian survey ataupun quisioner biasanya diperoleh data yang tidak dilakukan secara mendalam mengenai pendapat dan perasaan populasi, terutama jika hanya menggunakan angket. Tidak ada jaminan bahwa angket dijawab oleh seluruh sampel yang diambil. Selain itu Tidak dapat memperoleh informasi yang detail atau lengkap. Jawaban yang diisi terbatas seputar pertanyaan saja. Ada kemungkinan responden salah menafsirkan pertanyaan dan memberikan jawaban. biasanya survey atau quisioner tidak gratis dimana memerlukan dana yang terkadang harus memberikan imbalan seperti hadiah atau honor kepada responden yg mengisi survey ataupun quisioner, selain itu pada penelitian survey ataupun quisioner akan memerlukan waktu yang lama jika data yang di ambil 5000 data. Dengan menggunakan Analisis Sentimen tidak memerlukan responden karena data tweet merupakan bentuk nyata perasaan mereka terhadap suatu topic yang menggambarkan perasaan pengguna twitter sesungguhnya dan ekspresi yang sesungguhnya tanpa tekanan. Analisis sentiment juga dapat meminimalisir hasil quisioner atau survey yang bohong dan tidak sesuai dan analisis sentiment tidak membutuhkan waktu yg lama serta tidak perlu memerlukan biaya, imbalan, dan memberikan hadiah kepada responden kuisioner. Akan tetapi pendekatan machine learning dengan analisis sentimen

memiliki beberapa kekurangan, antara lain membutuhkan proses training dengan data yang cukup besar (Firdaus et al., 2021) dan memerlukan waktu cukup lama untuk proses komputasi (Arabnia et al., 2018). Selain itu, analisis sentimen dengan pendekatan machine learning hanya dapat digunakan pada topik tertentu, hal ini berkaitan dengan corpus yang dibangun pada suatu topik yang dibangun pada saat proses training (Muhammad et al., 2013). Selain itu analisis sentiment pada twitter banyak terdapat kata slang.

Dari fenomena ini, peneliti memutuskan untuk melakukan analisis sentiment menggunakan opini masyarakat Indonesia di Twitter, yang nantinya opini tersebut diklasifikasikan menjadi data opini sentiment negative dan positive. Penelitian ini menggunakan 2 algoritma berbeda yaitu Naïve Bayes *Classifier* dengan K-Nearest Neighbors (K-NN) 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 kedua 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.

B. Rumusan Masalah

Berdasarkan permasalahan yang didapatkan dari latar belakang, dapat dirumuskan masalah yang ada yaitu :

1. Bagaimana analisis sentimen terhadap data Twitter mengenai kesehatan mental pada media sosial twitter?
2. Bagaimana perbandingan hasil algoritma Naïve Bayes *Classifier* dengan K-Nearest Neighbors (K-NN), dan Decision Tree untuk analisis sentiment masyarakat terhadap Kesehatan mental pada media sosial twitter ?

C. Batasan Masalah

Batasan masalah dalam penelitian ini sebagai berikut.

1. Algoritma yang disebutkan dalam penelitian terbatas pada 3 algoritma yaitu algoritma Naïve Bayes *Clasifier*, K-Nearest Neighbors (K-NN), dan Decision Tree
2. Penelitian ini hanya membandingkan hasil dari tingkat akurasi kedua algoritma tersebut.

D. Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Memperoleh hasil analisis sentimen terhadap data Twitter mengenai analisis sentimen opini dari masyarakat terhadap kesehatan mental pada media sosial twitter
2. membandingkan tingkat akurasi dari tiga algoritma yaitu *Naïve Bayes Clasifier* dengan *K-Nearest Neighbors (K-NN)*, dan *Decision Tree* dalam analisis sentimen opini dari masyarakat terhadap kesehatan mental pada media sosial twitter.

E. MANFAAT PENELITIAN

Manfaat penelitian ini adalah sebagai berikut:

1. Menambah wawasan bagaimana mengkombinasikan beberapa metode untuk sebuah pengambilan keputusan
2. Dalam hal pengiriman data secara otomatis, diharapkan adanya integritas. Selain itu, analisis sentimen dapat memberikan informasi mengenai opini publik di platform media sosial Twitter mengenai kesehatan mental.

3. Mengetahui pengolahan dataset menggunakan metode *Naïve Bayes Classifier*, *K-Nearest Neighbors (K-NN)*, dan *Decision Tree* serta mengimplementasikan ketiga algoritma tersebut.
4. Dapat mengetahui mengkategorikan diskusi online tentang penggunaan teknologi yang telah melalui proses analisis sentimen dan membandingkan kinerja ketiga algoritma tersebut.



BAB V KESIMPULAN DAN SARAN

A. 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 tweets 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 sentiment negative, 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-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 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 % .

B. SARAN

Saran pada penelitian selanjutnya yaitu melakukan perbandingan terhadap metode lain seperti SVM yang dikenal juga sebagai algoritma klasifikasi terbaik dan mengkomparasikan algoritma lebih banyak lagi seperti empat atau lima algoritma, Selain itu, disarankan agar data pengujian atau pelatihan berjumlah lebih dari 1000 data agar proses klasifikasi berjalan lebih lancar. Kemudian, untuk meningkatkan kinerja pada metode apa pun yang menghasilkan tingkat kinerja yang tinggi, disarankan untuk menggunakan seleksi fitur untuk meningkatkan kinerja.

DAFTAR PUSTAKA

- 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>
- Asshofie, S., Saladin, A., & Alitopan, M. (2021). *Studi Komparasi Arsitektur Metafora Pada Bangunan Oceanarium Comparative Study of Metaphore Architecture on Oceanarium Building*. 351–356.
- 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.
- Ayuningtyas, D., Misnaniarti, M., & Rayhani, M. (2018). Analisis Situasi Kesehatan Mental Pada Masyarakat Di Indonesia Dan Strategi Penanggulangannya. *Jurnal Ilmu Kesehatan Masyarakat*, 9(1), 1–10. <https://doi.org/10.26553/jikm.2018.9.1.1-10>
- 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/jjm.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>
- Basri, H. (2017). PERANMEDIA SOSIAL TWITTER DALAM INTERAKSI SOSIAL PELAJAR SEKOLAH MENENGAH PERTAMA DI KOTA PEKANBARU (studi kasus pelajar SMPN 1 kota Pekanbaru). *Strategi Bertahan Hidup Petani Penggarap Di Jorong Sarilamak Nagari Sarilamak Kecamatan Harau Kabupaten Lima Puluh Kota*, 4(1), 1–13. <https://media.neliti.com/media/publications/183768-ID-partisipasi-masyarakat-dalam-pelaksanaan.pdf>
- Cahyaningrum, N. I., Yoshida Fatima, D. W., Kusuma, W. A., Ramadhani, S. A., Destanto, M. R., & Nooraeni, R. (2020). Analysis of User Sentiment of Twitter to Draft KUHP. *Jurnal Matematika, Statistika Dan Komputasi*, 16(3), 273. <https://doi.org/10.20956/jmsk.v16i3.8239>
- Ferdinan, A. H., T, C. S. S., Elektro, F. T., Telkom, U., Mining, D., & Neighbor, K. (2018). *Klasifikasi Emosi Pada Lirik Lagu Menggunakan Metode K- Nearest Neighbor Emotion Classification in Song Lyrics Using*. 5(3), 6187–6194.
- 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>

- Ikwan Hi. Sambiu dan Yusdiana Amir. 2018. (2015). Daftar Pustaka Daftar Pustaka. *Pemikiran Islam Di Malaysia: Sejarah Dan Aliran*, 20(5), 40–43. https://books.google.co.id/books?id=D9_YDwAAQBAJ&pg=PA369&lpg=PA369&dq=Prawirohardjo,+Sarwono.+2010.+Buku+Acuan+Nasional+Pelayanan+Kesehatan++Maternal+dan+Neonatal.+Jakarta+:+PT+Bina+Pustaka+Sarwono+Prawirohardjo.&source=bl&ots=riWNmMFyEq&sig=ACfU3UOHyN3I
- 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>
- Jumeilah, F. S. (2017). Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 1(1), 19–25. <https://doi.org/10.29207/resti.v1i1.11>
- Pulakiang, A. R., Wiranatha, I. M. N., & Batmetan, J. R. (2019). ANALISIS KUALITAS LAYANAN E-COMMERCE MENGGUNAKAN TWITTER API (Studi Kasus: Tokopedia, Lazada dan Bukalapak). *IJIS - Indonesian Journal On Information System*, 4(1), 25. <https://doi.org/10.36549/ijis.v4i1.47>
- Reyhana, Z., Pembimbing, D., Magister, P., Statistika, J., Matematika, F., & Data, D. A. N. S. (2018). *Kota Surabaya Melalui Twitter Dengan Menggunakan Support Vector Machine Dan Neural Network*.
- Rodiyana, R., Santoso, E., Puspitasari, W. D., & Miko, H. (2020). The Effect of Multimedia in Increasing the Integers Operation Ability. *Journal of Physics: Conference Series*, 1477(4). <https://doi.org/10.1088/1742-6596/1477/4/042057>
- Ruhyana, N. (2019). Analisis Sentimen terhadap Penerapan Sistem Plat Nomor Gajil/Genap pada Twitter dengan Metode Klasifikasi Naive Bayes. *Jurnal IKRA-ITH Informatika*, 3(1), 94–99. www.situs.com
- 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>
- Septianingrum, F., Jaman, J. H., & Enri, U. (2021a). Analisis Sentimen Pada Isu

- Vaksin Covid-19 di Indonesia dengan Metode Naive Bayes Classifier. *Jurnal Media Informatika Budidarma*, 5(4), 1431.
<https://doi.org/10.30865/mib.v5i4.3260>
- Septianingrum, F., Jaman, J. H., & Enri, U. (2021b). Analisis Sentimen Pada Isu Vaksin Covid-19 di Indonesia dengan Metode Naive Bayes Classifier. *Jurnal Media Informatika Budidarma*, 5, 1431–1437.
<https://doi.org/10.30865/mib.v5i4.3260>
- Sholihin, A., Haviluddin, Puspitasari, N., Wati, M., & Islamiyah. (2019). Sains, Aplikasi, Komputasi dan Teknologi Informasi Analisis Penyakit Difteri Berbasis Twitter Menggunakan Algoritma Naïve Bayes. *Sakti*, 1(1), 7.
- 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>
- Suryono, S., Utami, E., & Luthfi, E. T. (2018). *Analisis Sentiment Pada Twitter Dengan Menggunakan Metode Naive Bayes Classifier*. 9–15.
- 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, D. S., & Erfina, A. (2021). Analisis Sentimen Pinjaman Online di Twitter Menggunakan Algoritma Support Vector Machine (SVM). *SISMATIK (Seminar Nasional Sistem Informasi Dan Manajemen Informatika)*, 1(1), 299–305.
- 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
- Winahyu, J., & Suharjo, I. (2021). Aplikasi Web Analisis Sentimen Dengan Algoritma Multinomial Naïve Bayes. *Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)*, 10(2), 206.
<https://doi.org/10.23887/karmapati.v10i2.36609>

Yan, K., Arisandi, D., Studi, P., Informasi, S., & Tarumanagara, U. (n.d.). Terhadap Kesehatan Mental Masyarakat. *Ilmu Komputer Dan Sistem Informasi*.

Yulita, W., Dwi Nugroho, E., Habib Algifari, M., Studi Teknik Informatika, P., Teknologi Sumatera, I., Terusan Ryacudu, J., Huwi, W., Agung, J., & Selatan, L. (2021). Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier. *Jdmsi*, 2(2), 1–9.

