

Analisis Sentimen *Tweet* Pengguna *Twitter* Terkait Diabetes Menggunakan Metode *Naïve Bayes*

Afiyatar Asyer^{1*}, Magdalena A. Ineke Pakereng²

Teknik Informatika, Universitas Kristen Satya Wacana, Salatiga, Indonesia

*e-mail *Corresponding Author*: 672019061@student.uksw.edu

Abstract

Diabetes is a disease that has serious implications for the health and overall well-being of individuals, and there are diverse perspectives within society regarding this condition, leading to prolonged debates. Twitter is a popular platform where public opinions are often shared, and sentiment analysis on Twitter can provide insights into users' perceptions of diabetes, whether positive, negative, or neutral, to understand the health literacy regarding obesity among Twitter users. Sentiment analysis was conducted using text mining methodology, with a total of 26,038 crawled tweets, resulting in a sample of 4,635 data points. Through the Naïve Bayes algorithm, the sentiment analysis yielded 22% positive sentiment, 14% negative sentiment, and 64% neutral sentiment. The accuracy rate achieved was 87%. The sentiment analysis of tweets related to diabetes using text mining leaned more towards neutral sentiment than negative or positive sentiment. The accuracy of the Naïve Bayes algorithm falls within the category of "Good Classification".

Keywords: *Sentiment Analysis; Text Mining; Twitter; Diabetes; Naïve Bayes.*

Abstrak

Diabetes merupakan suatu penyakit yang berdampak serius terhadap kesehatan dan seluruh bagian tubuh penderitanya, banyak pandangan dari masyarakat terhadap penyakit diabetes yang menimbulkan perdebatan berkepanjangan. *Twitter* merupakan pilihan *platform* yang sering digunakan untuk memberikan opini publik, analisis sentimen di *Twitter* dapat memberikan gambaran persepsi pengguna terhadap diabetes baik secara positif, negatif ataupun netral untuk mengetahui literasi kesehatan terhadap persepsi obesitas pengguna *platform Twitter*. Analisis sentimen dilakukan dengan metode *text maining* dengan jumlah *crawling* sebanyak 26,038 dan menghasilkan sampel 4,635 data. Melalui algoritme *Naïve Bayes* hasil analisis sentimen didapatkan sentimen positif sebesar 22%, sentimen negatif sebesar 14%, dan sentimen netral sebesar 64%. Nilai akurasi dihasilkan sebesar 87%. Analisis sentimen *tweet* terkait diabetes menggunakan *text maining* lebih mengarah pada sentimen netral dibandingkan sentimen negatif dan sentimen positif. Nilai dari akurasi algoritme *Naïve Bayes* masuk dalam kategori "Good Classification".

Kata Kunci: *Analisis Sentimen; Text Maining; Twitter; Diabetes; Naïve Bayes.*

1. Pendahuluan

Diabetes merupakan penyakit bertahun yang ditandai dengan kadar gula darah tinggi yang melebihi normal, tanpa disadari terjadinya komplikasi terhadap penyandanganya sehingga dapat menyerang seluruh bagian tubuh. Diabetes sendiri memiliki tiga tipe yang bergejala berbeda, melalui catatan Organisasi *International Diabetes Federation* (IDF) setidaknya 463 juta penyandang diabetes di antara usia 20-79 tahun, pada penelitian yang dilakukan tahun 2019 [1]. Diabetes menjadi masalah kesehatan global dengan penderita yang terus meningkat, sehingga persepsi dan pandangan masyarakat terhadap diabetes dapat berperan dalam pencegahan penyakit diabetes.

Permasalahan kesehatan seperti diabetes tidak lepas dari pengguna aplikasi twitter sebagai *platform* penyedia layanan dalam memberikan tanggapan dan komentar. *Twitter* sendiri merupakan media sosial tempat membagikan *opini public* sebagai interaksi dengan pengguna lain melalui *tweet* (ciutan) sebagai informasi maupun pendapat pribadi terhadap isu dari diabetes [2]. *Twitter* juga sebagai media data *open source* dalam melakukan penelitian berupa analisis

sentimen (*text mining*) yang didapat dari komentar pengguna twitter untuk melihat sejauh mana masyarakat memberikan pendapat dan persepsi terhadap diabetes.

Analisis sentimen (*text mining*) merupakan sebuah teknik menambang opini data yang disampaikan berupa kalimat sentimen positif, negatif maupun netral, analisis sentimen menjadi pengolahan bahasa alami komputasi dari *linguistic* yang digunakan dalam menganalisis isi dari suatu ungkapan terhadap topik dapat berupa sikap pertimbangan dan evaluasi [3]. Analisis sentimen dapat digabungkan dengan algoritme *naïve bayes*, prediksi dari komentar pengguna dapat dikelompokkan menjadi hasil berupa kategori sentimen positif, negatif ataupun netral. Algoritme *naïve bayes* adalah metode dari pembelajaran mesin yang akan mengelompokkan komentar sentimen klasifikasi, dalam proses klasifikasi terhadap asumsi atribut yang saling tidak berkaitan dari *dataset* akan diberikan nilai terhadap variabel kelas [4]. *Dataset* yang digunakan pada penelitian diambil dari *website* resmi aplikasi twitter berupa *tweet* yang telah dipublikasi oleh *user* melalui proses pengumpulan data dari API (Antarmuka pemrograman aplikasi), data yang akan digunakan dimuat dalam bentuk tabel *spreadsheet* dan akan dilakukan *preprocessing* data, labeling sentimen dilakukan otomatis menggunakan algoritme *naïve bayes* sehingga akan mempermudah proses prediksi analisis sentimen.

Penelitian analisis sentimen *tweet* pengguna twitter terkait diabetes menggunakan metode *naïve bayes*, memberikan hasil akhir sentimen positif, negatif ataupun netral. Penelitian diharapkan memberikan *output* sejauh mana penerapan metode dalam penyelesaian masalah serta memberikan persepsi sentimen pengguna twitter terhadap diabetes untuk informasi yang berguna pada kampanye edukasi kesehatan dan berbagai pihak dalam bidang kesehatan masyarakat.

2. Tinjauan Pustaka

Penelitian sebelumnya oleh Felix Fridom Mailo dengan judul “Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia” membahas terkait *text mining* dengan menggunakan algoritme *naïve bayes* terhadap kata kunci obesitas dengan *webscaping* data sebanyak 67.942 dan diproses *cleaning* yang menghasilkan 43.435 data *tweet*, penelitian yang dilakukan menghasilkan sentimen positif sebanyak 22.246 data, sentimen negatif sebanyak 12.015 data dan sentimen netral 9.174 data, menghasilkan akurasi sebesar 94% yang menunjukkan jika metode *naïve bayes* mendapat hasil yang akurat [5].

Penelitian oleh Rita Apriani dengan judul “Analisis Sentimen Dengan *Naïve Bayes* Terhadap Komentar Aplikasi Tokopedia” membahas tentang *text mining* dengan algoritme *naïve bayes* dalam menganalisis sentiment secara otomatis *realtime* melalui *rapidminer*, hasil klasifikasi sentimen adalah positif dan negatif. Pengujian dilakukan dengan menggunakan 1.500 data dan menghasilkan sentimen positif sebesar 36.47% dan sentimen negatif sebesar 63.53%, dari hasil perhitungan didapatkan jika sentimen negatif dari pengguna lebih mendominasi, sehingga dapat memberikan saran terhadap aplikasi untuk memberikan kepercayaan kepada pengguna [6].

Penelitian oleh Tanthy Tawaqalia Widowati dengan judul “Analisis Sentimen Twitter Terhadap Tokoh Publik Dengan Algoritme *Naïve Bayes* dan *Support Vector Machine*”, penelitian analisis sentimen terhadap tokoh publik dengan menggunakan 2.500 data yang dibagi menjadi 2.200 data training dan 300 data testing, dari data tersebut dilakukan analisis sentimen dengan metode *naïve bayes* dan menghasilkan klasifikasi sentimen 1.743 data positif dan 757 data negatif, didapat nilai akurasi 91.48%, nilai prediksi 89.28% serta nilai *recall* 91.58% menandakan dalam penelitian yang dilakukan metode *naïve bayes* memiliki keakuratan dan lebih tepat dibanding *support vector machine* [7].

Penelitian oleh Winda Yulita dengan judul “Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritme *Naïve Bayes Classifier*”, analisis sentimen terhadap vaksinasi Covid-19 dengan data twitter sebanyak 3.780 *tweet*. Proses analisis sentimen dengan metode *naïve bayes* menghasilkan 2.278 (60.3%) sentimen positif, 1.299 (34.4%) sentimen netral dan 203 (5.4%) sentimen negatif yang menghasilkan nilai akurasi sebesar 93%, melalui hasil sentiment lebih mengarah kepada klasifikasi sentimen positif yang menunjukkan jika vaksin Covid-19 lebih diterima secara positif oleh masyarakat [8].

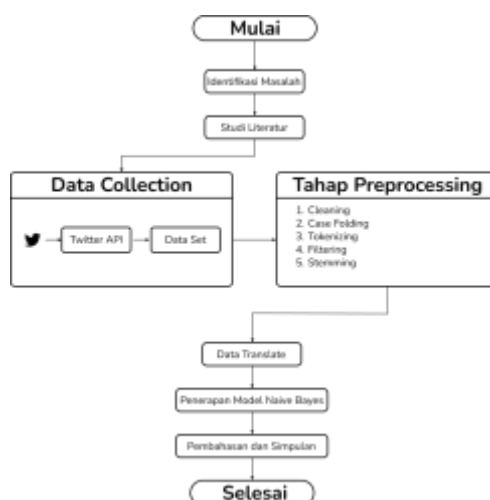
Penelitian oleh Lingga Aji Andika dengan judul “Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode *Naïve Bayes Classifier*” tentang analisis sentimen terhadap hitung cepat pemilihan presiden Indonesia pada tahun 2019 menggunakan metode *naïve bayes* dengan data sebanyak 1.366 menghasilkan 471 (34.5%) sentimen positif dan 895 (65.5%) sentimen negatif dan

menghasilkan akurasi terbaik sebesar 82.90%, dari hasil didapat jika klasifikasi sentimen negatif mendominasi sehingga untuk kedepannya dapat menjadi masukan untuk hitung cepat pemilihan umum [9].

Penelitian oleh Neng Resti Wardani dengan judul “Analisis Sentimen Masyarakat Terhadap Layanan Konsultasi Dokter Menggunakan Algoritme Naive Bayes” tentang analisis sentimen terhadap layanan konsultasi dokter menggunakan metode naïve bayes dengan total data yang diambil 300 data tweet dari tiga aplikasi berbeda, penelitian dilakukan dengan membandingkan tiga aplikasi yang membandingkan nilai akurasi dari metode yang digunakan. Aplikasi pertama menghasilkan nilai akurasi sebesar 98.75%, aplikasi kedua menghasilkan akurasi sebesar 82.86% dan aplikasi ketiga menghasilkan 62.86%, melalui hasil akurasi yang dilakukan aplikasi pertama dan kedua memiliki akurasi yang baik sedangkan aplikasi ketiga memiliki akurasi dibawah dari aplikasi pertama dan kedua dikarenakan dari data masih banyak makna ganda [10].

3. Metodologi

Metode penelitian yang digunakan pada penelitian yang dilakukan, ditunjukkan seperti pada Gambar 1.



Gambar 1 Flow Metode Penelitian

Flow metode penelitian pada Gambar 1, dijelaskan sebagai berikut. Tahapan *flow* metode penelitian dilakukan secara bertahap, sehingga nantinya menghasilkan keluaran yang dapat dipahami dalam bentuk informasi akurasi maupun diagram. Tahapan awal dilakukan

3.1. Identifikasi Masalah

Dilakukan identifikasi permasalahan terkait penyakit diabetes yang hadir dalam masyarakat terutama pada media sosial twitter. Tanggapan atau ciutan yang diunggah oleh pengguna twitter dapat bersifat positif, negatif, ataupun netral terhadap pengidap diabetes. Argumentasi yang beragam dari pengguna twitter dapat diteliti dengan sebuah metode kecerdasan buatan untuk memberikan citra pandangan masyarakat terhadap diabetes.

3.2. Studi Literatur,

Pada tahap studi literatur dilakukan dengan mencari teori-teori dan metode yang relevan terhadap penelitian yang sudah pernah dilakukan sebelumnya, penelitian yang dilakukan untuk menghasilkan solusi terhadap permasalahan yang ada. Teori yang digunakan yaitu Opini, *Data Mining*, *Machine Learning*, dan *Naïve Bayes*.

3.3. Data collection

Pada tahap ini dilakukan dengan cara *crawling* yaitu melakukan permintaan langsung kepada pihak twitter untuk mendapatkan data yang diperlukan melalui pemanfaatan *Application Program Interface* (API) [11]. *API key* dan *website netlytic* digunakan untuk mendapatkan *dataset* dari pihak twitter. *Dataset* yang digunakan dari hasil *crawling* dengan *hashtag* pencarian #Diabetes

#DiabetesMilitus #PenyakitDiabetes #Diabet. Proses *crawling* data dilakukan dengan bahas pemrograman *python* dan akan disimpan dengan format *.csv*.

3.4. Pre-processing

Pada tahap ini dilakukan untuk mempersiapkan data mentah menjadi data yang dapat diolah sesuai dengan format untuk menganalisa setiap frase yang ada dalam ciutan twitter, data akan melewati proses pembersihan, penyesuaian alfabet, normalisasi, pemisahan frasa, penghapusan kata yang tidak bermakna, dan mengubah kata menjadi bentuk dasar [12]. Tahapan yang akan dilakukan dalam *pre-processing*: *Cleaning* atau pembersihan data dilakukan untuk menghasilkan data yang bersih dan bertujuan untuk menghapus setiap ciutan yang memiliki karakter yang tidak diperlukan seperti tanda baca, karakter serta simbol khusus, ciutan yang dikirim berulang, dan emoji dari tiap kalimat. *Case folding* dilakukan untuk menyamakan setiap kata dari ciutan menjadi huruf kecil, sehingga tidak ada huruf dari kata yang berawalan kapital. *Tokenizing* atau tokenisasi dilakukan dengan membagi tiap kalimat menjadi kata yang terpisah, proses tokenisasi akan menghasilkan kata yang akan dipisah dengan spasi. *Filtering* dilakukan dengan cara *stopword* yaitu proses untuk menghilangkan kata yang tidak memiliki makna dan tidak memberikan kontribusi dalam tahapan pemrosesan data. *Stemming* atau pemrosesan bahasa alami dilakukan dengan mengubah setiap kata yang berimbunan menjadi kata paling dasar "*root word*" [13], tujuan dari *stemming* adalah mencari kata kunci dari setiap kalimat untuk memudahkan analisis dari setiap kata yang ada dalam ciutan.

3.5. Data translate

Pada tahap ini menjadikan ciutan yang sudah bersih diterjemahkan ke dalam bahasa Inggris karena memiliki banyak memiliki arti kosa kata [14]. Data yang sudah dilakukan proses *translate* serta tidak ada baris data kosong, akan masuk pada proses labeling data menggunakan tools google colabs python dengan mengitung polaritas dari setiap komentar, sentimen masuk dalam komentar positif ketika polaritas lebih dari nol, sentimen masuk dalam komentar negatif ketika polaritas kurang dari nol serta sentimen masuk dalam komentar netral ketika polaritas sama dengan nol. Pada tahapan ini sentimen setiap komentar sudah ada klasifikasi dan akan dilakukan pengecekan dengan menggunakan algoritme naïve bayes.

3.6. Penerapan Model Naïve bayes

Proses algoritme naïve bayes dengan perhitungan statistika berdasar pada probabilitas kemunculan tiap kata menggunakan pemrosesan bahasa alami yang akan dikategorikan ke dalam sentimen positif, negatif atau netral [15]. Data yang menjadi *input* kedalam algoritme naïve bayes berupa komentar dari pengguna twitter yang telah melalui proses *pre-processing* dan proses *translate*, algoritme naïve bayes akan menghitung setiap bobot kata dan menghasilkan luaran sentimen pada setiap komentar akan mengarah pada sentimen positif, negatif ataupun netral. Rumus persamaan dari perhitungan probabilitas naïve bayes:

$$P(X|Y) = \frac{P(Y|X) \times P(X)}{P(Y)} \dots\dots\dots(1)$$

Keterangan:

- X = Hipotesis data dari *class* spesifik
- Y = Data *class* yang tidak diketahui
- P(X|Y) = Probabilitas X berdasar kondisi Y
- P(Y|X) = Probabilitas Y berdasar kondisi X
- P(X) = Probabilitas X
- P(Y) = Probabilitas Y

Setelah perhitungan algoritme naïve bayes, akan dilakukan visualisasi dalam bentuk diagram dengan presentase sentimen serta menampilkan wordcloud yaitu kata yang sering muncul dalam sentimen. Perhitungan data latih dan uji dilakukan dengan menggunakan *random sample*, pada tahap akhir akan dilakukan evaluasi pada kinerja dari proses klasifikasi yang telah diuji menggunakan *confusion matrix* dan *classification report* dengan parameter *accuracy*, *precision*, dan *recall* sebagai berikut:

1) *Precision*

Precision merupakan visualisasi presentase keakuratan dari metode, rumus dari *precision*:

$$Precision = \frac{TP}{TP + FP} \dots\dots\dots(2)$$

2) *Recall*

Recall merupakan visualisasi penyesuaian dalam pencarian dari sebuah metode, rumus dari *recall*:

$$Recall = \frac{TP}{TP + FN} \dots\dots\dots(3)$$

3) *Accuracy*

Accuracy merupakan visualisasi dari kinerja keakuratan model, rumus dari *accuracy*:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \dots\dots\dots(4)$$

Berdasarkan data yang telah didapat akan disimpulkan dengan menghitung jumlah sentimen positif, negatif ataupun netral.

4. Hasil dan Pembahasan

Pada penelitian analisis sentimen tweet pengguna twitter terkait diabetes menggunakan metode naive bayes akan dilakukan analisis sentiment pada data yang akan digunakan sebagai *dataset*, tujuan dari penelitian adalah untuk mengetahui analisis sentimen terhadap diabetes terhadap penelitian sebelumnya dengan memperbanyak library *filtering* pada tahapan *pre-processing* dan menambahkan data *translate* untuk menghitung dari probabilitas kosa kata untuk sentimen positif, negatif ataupun netral. Pada penelitian yang dilakukan, data didapat melalui *crawling* dari twitter dengan mendaftarkan akses token API dan melalui *website netlytic*. Hasil dari pengumpulan data sebanyak 26,038 berupa *username* dan *text* (komentar) dapat dilihat seperti pada Tabel 1 yang merupakan hasil *crawling*, data yang terkumpul masih banyak duplikat dan belum dapat diolah menggunakan algoritme naive bayes harus melewati tahapan *pre-processing* untuk menghasilkan data yang bersih. Inputan yang akan dilakukan pada tahapan *pre-processing* adalah *text* (komentar), untuk menghilangkan data duplikasi dari komentar yang di *retweet*.

Tabel 1 Data Crawling

Username	text
Followajedah	@convomf Takut ah sugar sugaran, lg banyak kena diabetes
CupeiPei	#news #berita #fakta #indonesia #viral #beritaterkini makanan untuk penderita diabetes agar tidak lemas https://t.co/mAW8b89LTC

Proses *pre-processing* dilakukan setelah mendapatkan data dari proses *crawling*, Adapun tahapan dari *pre-processing* yang akan dilakukan:

1) Proses *cleaning* dilakukan penghapusan *dataset* yang merupakan *retweet*, penghapusan simbol, *link* dan *taggar*. Sehingga *dataset* yang awalnya sebanyak 26,038 menjadi 4,635 *dataset* seperti ditunjukkan pada Tabel 2.

Tabel 2 Proses Pembersihan Ciutan

Ciutan	Pembersihan Simbol Khusus	Pembersihan Link	Pembersihan Taggar
@convomf Takut ah sugar sugaran, lg banyak kena diabetes	Takut ah sugar sugaran, lg banyak kena diabetes	Takut ah sugar sugaran, lg banyak kena diabetes	Takut ah sugar sugaran, lg banyak kena diabetes
#news #berita #fakta	#news #berita #fakta #indonesia #viral	#news #berita #fakta	news berita fakta indonesia viral

Ciutan	Pembersihan Simbol Khusus	Pembersihan Link	Pembersihan Taggar
#indonesia #viral #beritaterkini makanan untuk penderita diabetes agar tidak lemas https://t.co/mAW8b89LTC	#beritaterkini makanan untuk penderita diabetes agar tidak lemas https://t.co/mAW8b89LTC	#indonesia #viral #beritaterkini makanan untuk penderita diabetes agar tidak lemas	beritaterkini makanan untuk penderita diabetes agar tidak lemas

2) *Case folding* dilakukan agar kata yang sama dapat memiliki makna yang sama juga seperti terlihat pada Tabel 2.

Tabel 3 Proses *Case Folding*

Pembersihan Taggar	<i>Case Folding</i>
Kanker Pankreas Disebabkan Karena Terlalu Banyak Mengonsumsi Makanan Berlemak Ketimbang Sayuran. Bagi Anda Yang Menderita Diabetes,Hati-Hati	kanker pankreas disebabkan karena terlalu banyak mengonsumsi makanan berlemak ketimbang sayuran. bagi anda yang menderita diabetes,hati-hati
Kebiasaan orang tua kasih makanan ternyata bisa memicu diabetes buat anak lho, terutama diabetes tipe . Yang biasanya dipicu sama gaya hidup gak sehat. Misalnya kayak nonton tv kelamaan, kurang olahraga, dan makan junk food. Yuk baca postingannya!	kebiasaan orang tua kasih makanan ternyata bisa memicu diabetes buat anak lho, terutama diabetes tipe . yang biasanya dipicu sama gaya hidup gak sehat. misalnya kayak nonton tv kelamaan, kurang olahraga, dan makan junk food. yuk baca postingannya!

3) *Tokenizing* dilakukan dengan membagi tiap kalimat menjadi kata yang terpisah, proses tokenisasi akan menghasilkan kata yang akan dipisah dengan spasi. Pembagian kalimat ini memudahkan dalam menentukan kata positif, negatif atau netral. Misalkan kalimat “seseorang juga mungkin lebih rentan terkena diabetes, obesitas, dan kondisi kesehatan mental akibat terganggunya irama sirkadian” setelah melalui proses tokenisasi menjadi “seseorang”, “juga”, “mungkin”, “lebih”, “rentan”, “terkena”, “diabetes”, “obesitas”, “dan”, “kondisi”, “kesehatan”, “mental”, “akibat”, “terganggunya”, “irama”, “sirkandi”.

4) *Filtering stopwords* dapat menggunakan pustaka kata yang sudah tersedia maupun membuat daftar kata. Contoh pustaka kata seperti ditunjukkan pada Tabel 4.

Tabel 4 *Filtering*

Stopword				
a	ada	adalah	Adanya	adapun
agak	agaknya	agar	Akan	akankah
akhir	akhirnya	akhirnya	Aku	akulah
amat	amatlah	anda	Andalah	antar
antara	antaranya	apa	Apaan	apabila
apakah	apalagi	apatah	Arti	artinya
asal	asalkan	atas	Atau	ataukah
ataupun	awal	awalnya	B	bagai
bagaikan	bagaimana	bagaimanakah	Bagaimanapun	bagainamakah
bagi	bagian	bahkan	Bahwa	bahwasannya
bahwasanya	baik	baiklah	Bakal	bakalan
balik	banyak	bapak	Baru	bawah
beberapa	begini	beginian	Beginikah	beginilah
begitu	begitukah	begitulah	Begitupun	bekerja

Stopword				
belakang	belakangan	belum	Belumlah	benar
benarkah	benarlah	berada	berakhir	berakhirlah

5) *Stemming* dilakukan untuk membuat setiap kata menjadi bentuk dasar dimana setelah melakukan tahapan *case folding* akan dilakukan tahapan *stemming*, seperti ditunjukkan pada Tabel 4.

Tabel 5 Proses *Stemming*

<i>Case folding</i>	<i>Stemming</i>
kanker pankreas disebabkan karena terlalu banyak mengonsumsi makanan berlemak ketimbang sayuran. bagi anda yang menderita diabetes,hati-hati	['kanker', 'pankreas', 'sebab', 'konsumsi', 'makan', 'lemak', 'ketimbang', 'sayur', 'derita', 'diabetes']
kebiasaan orang tua kasih makanan ternyata bisa memicu diabetes buat anak lho, terutama diabetes tipe . yang biasanya dipicu sama gaya hidup gak sehat. misalnya kayak nonton tv kelamaan, kurang olahraga, dan makan junk food. yuk baca postingannya!	['makan', 'picu', 'diabetes', 'lho', 'diabetes', 'tipe', 'picu', 'tv', 'olahraga', 'makan', 'junk', 'food', 'postingannya']

Data *Translate* digunakan untuk membuat data yang sudah bersih menjadi suatu bahasa yang memiliki banyak kosa kata yang dapat diprediksi masuk sentimen positif, negatif atau netral seperti Tabel 5. Setelah melakukan *data translate*, akan dilakukan perhitungan polaritas melalui python untuk menghitung sentimen. Komentar masuk dalam sentimen positif ketika polaritas lebih dari nol, sentimen masuk dalam komentar negatif ketika polaritas kurang dari nol serta sentimen masuk dalam komentar netral ketika polaritas sama dengan nol, dapat dilihat pada Tabel 6.

Tabel 6 Proses Data *Translate*

Data Translate	Sentimen
pancreatic cancer due to consumption of eating fat rather than vegetables suffer from diabetes	Negatif
eating triggers diabetes, you know type diabetes triggers TV sports eating junk food posts	Netral

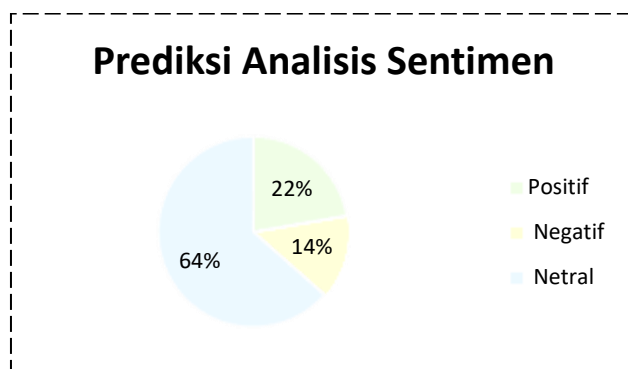
Setelah didapatkan sentimen pada tahapan data translate, akan dilakukan perhitungan *algoritme naïve bayes* untuk menghasilkan *output* klasifikasi sentiment, dapat dilihat pada Tabel 7.

Tabel 7 Hasil Sentimen Algoritme Naïve Bayes

Data Translate	Sentimen	Sentiment algoritme naïve bayes
pancreatic cancer due to consumption of eating fat rather than vegetables suffer from diabetes	Negatif	Negatif
eating triggers diabetes, you know type diabetes triggers TV sports eating junk food posts	Netral	Netral

Sehingga didapatkan melalui analisis sentimen yang dilakukan terhadap algoritme naïve bayes dapat diklasifikasikan, dari data pada Tabel 7 komentar pertama masuk ke dalam klasifikasi

sentimen negatif dan untuk komentar kedua masuk kedalam klasifikasi sentimen netral. Hasil dari analisis sentimen berupa visualisasi ciutan yang telah dilakukan proses prediksi, hasil analisis sentimen juga bertujuan untuk memberikan gambaran persentase dari data yang telah diprediksi masuk kedalam klasifikasi sentimen positif, negatif, atau netral. Berdasarkan diagram analisis sentimen pada Gambar 2, didapatkan persentase positif sebesar 22% dengan banyak data 1,033, negatif sebesar 14% dengan banyak data 659 dan netral sebesar 64% dengan banyak data 2,943 dari 4,635 dataset.



Gambar 2 Diagram Analisis Sentimen



Gambar 3 Kata Positif Analisis Sentimen

Gambar 3 memvisualisasikan dari banyak kata positif yang muncul dalam ciutan, kata positif yang sering muncul akan semakin besar ukuran kata dan sebaliknya semakin kecil ukuran kata, kemunculan dari ciutan semakin sedikit. Contoh untuk kata positif yaitu “*healthy*” yang artinya sehat dan “*good*” yang artinya baik, dalam ciutan ada kata positif yang akan masuk dalam klasifikasi positif.



Gambar 4 Kata Negatif Analisis Sentimen

Gambar 4 memvisualisasikan dari banyak kata negatif yang muncul dalam ciutan, contoh untuk kata negatif yaitu “bad” yang artinya buruk dan “danger” yang artinya berbahaya, dalam ciutan ada kata negatif yang akan masuk dalam klasifikasi negatif.



Gambar 5 Kata Analisis Sentimen Netral

Gambar 5 memvisualisasikan dari banyak kata netral yang muncul dalam ciutan, contoh untuk kata netral yaitu “medication” yang artinya pengobatan, dalam ciutan ada kata netral yang akan masuk dalam klasifikasi netral.

Tahapan akhir adalah dengan melakukan evaluasi terhadap algoritme *naïve bayes* untuk menilai dari kinerja metode, evaluasi dilakukan dengan melihat nilai dari *confusion matrix* dalam bentuk numerik, dapat dilihat pada Tabel 8.

Tabel 8 Multiclass Confusion Matrix

	TP (True Positive)	FP (False Positive)	FN (False Negative)
Positif	152	8	2
Negatif	56	609	73
Netral	8	7	244

Dalam mengukur evaluasi dari metode klasifikasi *naïve bayes* menggunakan *Classification Report* terhadap kinerja model klasifikasi, didapatkan hasil seperti ditunjukkan pada Tabel 9.

Tabel 9 Hasil Classification Report

	Precision	Recall		
Negatif	0.92	0.72		
Netral	0.84	0.97		
Positif	0.94	0.78		
Accuracy			0.87	4635
Macro Avg	0.90	0.82	0.85	4635

Berdasarkan hasil *classification report* pada Tabel 9, menggunakan metode analisis sentimen *naïve bayes* didapatkan hasil sebagai berikut: nilai akurasi *naïve bayes* diperoleh sebesar 87%, diartikan bahwa model dapat memprediksi sentimen dengan benar terhadap 87% *dataset*, nilai *precision naïve bayes* diperoleh sebesar 90%, nilai *recall naïve bayes* diperoleh sebesar 82%. Dapat disimpulkan bahwa algoritme *naïve bayes* berhasil dalam memprediksi kategori dari sentimen dengan akurasi yang masuk ke dalam “*Good Classification*”.

5. Simpulan

Penelitian yang dilakukan untuk menganalisis sentimen terhadap pendapat pengguna media sosial twitter terhadap diabetes dengan menggunakan metode *naïve bayes*, berdasarkan hasil penelitian dan pembahasan dalam bentuk visualisasi dan evaluasi didapatkan hasil prediksi klasifikasi sentimen positif sebesar 1,033 data, sentimen negatif sebesar 659 data dan sentimen netral sebesar 2,943 data. Berdasarkan hasil yang diteliti, dapat disimpulkan jika pengguna *twitter* lebih banyak memberikan sentimen netral terhadap penyakit diabetes kemudian disusul dengan sentimen positif serta negatif, sehingga bisa dikatakan jika masyarakat masih belum memiliki literasi yang cukup terhadap penyakit diabetes, agar kedepannya melalui penelitian ini literasi terhadap edukasi kesehatan masyarakat dapat lebih ditingkatkan terutama pada penyakit diabetes. Berdasarkan akurasi yang didapat yaitu sebesar 87% membuktikan jika algoritme *naïve bayes* dapat menyelesaikan permasalahan sentimen masyarakat terhadap diabetes. Melalui penelitian juga didapatkan jika *library* yang banyak pada *stopword* dalam proses *pre-processing* dapat memberikan kinerja yang cukup baik terhadap akurasi algoritme *naïve bayes*.

Daftar Referensi

- [1] Kementerian Kesehatan RI., "Infodatin tetap produktif, cegah, dan atasi Diabetes Melitus 2020," *Pusat Data dan Informasi Kementerian Kesehatan RI*. pp. 1–10, 2020.
- [2] Z. Liu and I. Weber, "Is twitter a public sphere for online conflicts? a cross-ideological and cross-hierarchical look," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8851, pp. 336–347, 2014, doi: 10.1007/978-3-319-13734-6_25.
- [3] T. Luo, S. Chen, G. Xu, and J. Zhou, "Trust-based Collective View Prediction," *Trust. Collect. View Predict.*, no. June, 2013, doi: 10.1007/978-1-4614-7202-5.
- [4] A. Yasar and M. M. Saritas, "Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification," *Int. J. Intell. Syst. Appl. Eng.*, vol. 7, no. 2, pp. 88–91, 2019, doi: 10.18201/ijisae.2019252786.
- [5] F. F. Mailo and L. Lazuardi, "Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia," *J. Inf. Syst. Public Heal.*, vol. 4, no. 1, pp. 28–36, 2019.
- [6] R. Apriani *et al.*, "Analisis Sentimen dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia," *J. Rekayasa Teknol. Nusa Putra*, vol. 6, no. 1, pp. 54–62, 2019, [Online]. Available: <https://rekayasa.nusaputra.ac.id/article/view/86>
- [7] T. T. Widowati and M. Sadikin, "Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritme Naive Bayes dan Support Vector Machine," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 2, pp. 626–636, 2021, doi: 10.24176/simet.v11i2.4568.
- [8] W. Yulita *et al.*, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritme Naïve Bayes Classifier," *Jdmsi*, vol. 2, no. 2, pp. 1–9, 2021.
- [9] L. A. Andika, P. A. N. Azizah, and R. Respatiwan, "Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier," *Indones. J. Appl. Stat.*, vol. 2, no. 1, p. 34, 2019, doi: 10.13057/ijas.v2i1.29998.
- [10] N. R. Wardani and A. Erfina, "Analisis Sentimen Masyarakat Terhadap Layanan Konsultasi dokter Menggunakan Algoritme Naive Bayes," *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.)*, pp. 12–18, 2021.
- [11] J. Eka Sembodo, E. Budi Setiawan, and Z. Abdurahman Baizal, "Data Crawling Otomatis pada Twitter," no. October 2018, pp. 11–16, 2016, doi: 10.21108/indosc.2016.111.
- [12] S. Khairunnisa and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi)," vol. 5, no. April, pp. 406–414, 2021, doi: 10.30865/mib.v5i2.2835.
- [13] I. M. A. Agastya, "Pengaruh Stemmer Bahasa Indonesia Terhadap Peforma Analisis Sentimen Terjemahan Ulasan Film," *J. Tekno Kompak*, vol. 12, no. 1, p. 18, 2018, doi: 10.33365/jtk.v12i1.70.
- [14] G. Shalunts, G. Backfried, and N. Commeignes, "The Impact of Machine Translation on Sentiment Analysis," *Fifth Int. Conf. Data Anal.*, no. c, pp. 51–56, 2016.
- [15] M. Taboada, "Sentiment Analysis: An Overview from Linguistics," *Annu. Rev. Linguist.*, vol. 2, no. February, pp. 325–347, 2016, doi: 10.1146/annurev-linguistics-011415-040518.