

ANALISIS SENTIMEN PENGGUNA APLIKASI X TERHADAP PROGRAM MAKAN SIANG GRATIS DENGAN METODE NAÏVE BAYES CLASSIFIER

Linggar Nursinggah^{1*}, Ruuhwan², Teuku Mufizar³

^{1,2,3} Universitas Perjuangan, Jln. Peta No. 177, Kahuripan, Kec. Tawang, Kota Tasikmalaya (0265) 326058

Received: 23 April 2024

Accepted: 31 Juli 2024

Published: 7 Agustus 2024

Keywords:

Analisis Sentimen, *Naïve Bayes Classifier*, X, program makan siang gratis, SEMMA.

Correspondent Email:

ruuhwan@unper.ac.id

Abstrak. Saat ini perkembangan teknologi berkembang dengan pesat. Informasi lebih mudah didapatkan melalui portal atau media sosial. Informasi yang disajikan biasanya berupa teks pada media sosial berdasarkan pengelompokan isi informasi. Program makan siang gratis merupakan gagasan salah satu pasangan calon pada pemilu presiden Indonesia tahun 2024. Target dari program ini adalah ibu hamil, balita, anak-anak TK, SD, SMP dan sederajat. Penelitian ini berfokus pada analisis sentimen pengguna aplikasi X terhadap program makan siang gratis. Data yang diambil merupakan cuitan pengguna aplikasi X terhitung mulai tanggal 14 Februari – 20 Maret 2024. Penelitian analisis sentimen ini dapat mengetahui seberapa banyak sentimen positif dan negatif masyarakat. Metode yang digunakan adalah SEMMA data mining proses menggunakan algoritma Naïve Bayes. Hasil yang didapatkan berupa confusion matrix yang terdiri dari akurasi, presisi dan recall. Set K-fold validation dengan nilai K10 digunakan untuk menentukan nilai confusion matrix. Diketahui hasil cuitan sentimen positif sebesar 1543 dan sentimen negatif 1366 pada aplikasi X terhadap program makan siang gratis. Hasil sentimen positif 53,04% dan sentimen negatif 44,95%. Didapatkan hasil akurasi 80,31% data yang benar dan $\pm 5,27$ margin error, hasil presisi sentimen positif menghasilkan 80,69% dan sentimen negatif sebesar 79,95%, hasil recall sentimen positif sebesar 79,71% dan sentimen negatif 80,93%.

Abstract. Currently, the development of technology is growing rapidly. Information is more easily obtained through portals or social media. The information presented is usually in the form of text on social media based on the grouping of information content. The free lunch program is the idea of one of the candidate pairs in the Indonesian presidential election in 2024. The targets of this program are pregnant women, toddlers, children in kindergarten, elementary school, junior high school and equivalent. This research focuses on analyzing the sentiment of X application users towards the free lunch program. The data taken is the tweets of X application users starting from February 14 - March 20, 2024. Sentiment analysis research can find out how much positive and negative sentiment of the community. Method used is SEMMA data mining process using Naïve Bayes algorithm. The results obtained in the form of confusion matrix consisting of accuracy, precision and recall. The K-fold validation set with a value of K10 is used to determine the value of the confusion matrix. It is known that the results of positive sentiment tweets are 1543 and 1366 negative sentiments in application X towards the free lunch program. The result of positive sentiment is 53.04% and negative sentiment is 44.95%. The accuracy result is 80.31% correct data and ± 5.27 margin of error, the precision result of

positive sentiment results in 80.69% and negative sentiment of 79.95%, the recall result of positive sentiment is 79.71% and negative sentiment 80,93%.

1. PENDAHULUAN

Saat ini, teknologi berkembang dengan cepat. Seiring berjalannya waktu, lebih mudah untuk mendapatkan informasi melalui portal atau media sosial. Informasi yang disajikan di media sosial biasanya berupa teks dan disusun berdasarkan bagaimana isi disusun. Media sosial adalah tempat di mana orang dapat menyampaikan pendapat mereka, yang dapat berupa pujian, ujaran kebencian, atau hoax, yang dapat menyebabkan perdebatan. Twitter adalah salah satu media di mana orang dapat berbicara.[1]

Sebuah survei yang dilakukan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) menunjukkan bahwa jumlah orang Indonesia yang menggunakan internet akan mencapai 221.563.479 pada tahun 2024, meningkat dari 278.696.200 pada tahun sebelumnya. Ini mencakup 79,5 persen dari populasi keseluruhan. Sebaliknya, laporan We Are Social menyatakan bahwa per oktober 2023, Indonesia memiliki sekitar 27,5 juta pengguna twitter dengan alias X, menempatkan negara itu di peringkat keempat dari semua negara yang memiliki pengguna twitter. Istilah warganet atau netizen mudah dijumpai dan ditemukan pada masa sekarang ini. Menurut wikipedia, Warganet berasal dari kata "warga" dan "internet." Kata ini mengacu pada seseorang yang aktif terlibat dalam komunitas maya atau internet secara keseluruhan. Istilah ini juga sering digunakan untuk menunjukkan minat dan aktivitas yang terjadi di internet, yang membuatnya wadah sosial dan intelektual, atau struktur politik di sekitarnya. Khususnya, ini berkaitan dengan masalah seperti akses terbuka, netralitas internet, dan kebebasan berbicara.

Twitter alias X merupakan aplikasi media sosial yang di dalamnya para pengguna aplikasi tersebut dapat melakukan interaksi yang disebut dengan tweet. Warganet dapat saling bertukar informasi dan opini tentang segala hal yang sedang dibicarakan. X mempunyai karakteristik dapat mengirim opini tanpa batas mencari berita terbaru, membagikan *tweet* orang lain dan memberi komentar. Di aplikasi X sendiri dapat dengan

mudah dan cepat sebuah informasi tersebar, hal inilah yang menjadi media untuk mengetahui sentimen opini seseorang baik itu positif maupun negatif.

Program makan siang gratis merupakan gagasan salah satu pasangan calon pada pemilu presiden Indonesia tahun 2024. Target dari program ini adalah ibu hamil, balita, anak-anak TK, SD, SMP dan sederajat. Hal tersebut menjadi kontroversi sehingga banyak muncul pendapat dan opini yang dilontarkan masyarakat terutama pada media sosial X. kata kunci makan siang gratis banyak memunculkan opini masyarakat.

Analisis sentimen merupakan teknik untuk mengevaluasi nada emosional teks digital untuk menentukan apakah nada emosionalnya positif, negatif, atau netral. Penerapan analisis sentimen yaitu menganalisa pendapat, sentimen, evaluasi, emosi, penilaian, atau sikap pada suatu produk, tokoh, organisasi, isu, produk, layanan dan peristiwa. Selain itu, analisis sentimen akan selalu berhubungan dengan masyarakat karena sumber informasi yang didapat di media sosial dimana masyarakat sebagai penggunanya[1]

Penelitian ini menggunakan metode SEMMA (*sample, explore, modify, model, assess*) yang didalamnya terdapat algoritma *Naïve Bayes Classifier*. Metode ini dipilih karena memiliki algoritma sederhana dengan akurasi yang tinggi. Penelitian ini akan mengkaji tentang sentimen masyarakat pengguna aplikasi X terhadap program makan siang gratis.

2. TINJAUAN PUSTAKA

2.1 Analisis Sentimen

Analisis sentimen, juga dikenal sebagai *sentiment analysis*, adalah proses menggunakan *text analytics* untuk mendapatkan data dari berbagai sumber di internet. Terdapat beberapa tipe analisis sentimen, seperti *fine-grained sentiment analysis* yang fokus pada tingkat polaritas pendapat, *intent sentiment analysis* untuk mengidentifikasi motivasi dibalik pesan pengguna, dan *aspect-Based Sentiment Analysis* untuk menghubungkan sentimen spesifik dengan aspek produk atau layanan.

Analisis sentimen bertujuan untuk menilai emosi, sikap, dan pendapat serta pendapat yang disampaikan oleh orang-orang di berbagai media tentang barang, merek, layanan, politik, atau organisasi. Metode ini termasuk machine learning dan lexicon-based, yang memungkinkan data dikategorikan ke dalam kategori seperti sangat positif, positif, netral, negatif, dan sangat negatif.[2]

Analisis sentimen terdiri dari beberapa jenis yaitu *emotion detection*, *aspect-based sentiment analysis*, dan *fine grand sentiment analysis*. *Fine sentiment analysis* adalah jenis analisis yang memiliki penilaian spesifik dan biasa digunakan pada bidang *e-commerce*. *Emoticon detection* adalah jenis analisis yang bertujuan untuk mengetahui emosi yang ada pada pesan misalnya emosi bahagia, sedih, marah, dan lain-lain. *Aspect-based sentiment analysis* merupakan jenis analisis untuk mengetahui aspek yang berpengaruh dan penilaian dari pelanggan[3]

2.2 Media Sosisal

Media sosial adalah platform digital yang memainkan peran penting dalam kehidupan manusia modern. Fungsi utamanya adalah untuk memfasilitasi komunikasi antar individu, berbagi informasi, dan menciptakan interaksi sosial tanpa batasan ruang dan waktu. Media sosial adalah wadah media online yang bekerja dengan teknologi berbasis *web* dan mengubah cara komunikasi yang dahulunya satu arah menjadi dua arah atau dialog interaktif.[4]

2.3 Data Mining

Data mining merupakan proses untuk pengumpulan dan pengolahan data ini memiliki tujuan mengekstrak informasi penting dari data. Proses ini dapat dicapai melalui perangkat lunak yang menggunakan penghitungan matematika atau kecerdasan buatan. Pada data mining terdapat berbagai metode, salah satunya yaitu SEMMA (*sample, Explore, modify, Model, dan Assess*).

2.4 Google colab

Pengguna dapat menulis, menjalankan, dan berbagi kode *Python* secara gratis di platform berbasis *cloud Google Colab*. *Google Colab* adalah platform berbasis *cloud* yang memungkinkan pengguna menulis, menjalankan, dan berbagi kode *Python* secara gratis. Dengan GPU dan TPU, platform memudahkan pengguna untuk mengolah data

dan melakukan eksperimen dalam bidang statistik, matematika, dan pembelajaran mesin. *Google Colab* telah menyiapkan hampir semua pustaka perangkat lunak yang diperlukan. Dari segi perangkat keras, *Google Colab* menyediakan fasilitas seperti media penyimpanan yang terintegrasi dengan *Google Drive*, prosesor CPU, GPU, TPU, dan RAM, serta pendukung lainnya, seperti *Matplotlib* untuk pembuatan grafik. *TensorFlow* versi 1.x dan 2.x dan *Python* versi 2.x dan 3.x juga tersedia.[5]

2.5 Text preprocessing

Preprocessing teks adalah proses mengubah data teks yang tidak terstruktur menjadi terstruktur sehingga proses komuterisasi menjadi lebih mudah.[6] *text preprocessing* memiliki beberapa tahap dalam prosesnya yaitu *cleaning, tokenize, transform case, stopword removal, dan filter*.

2.6 Naïve Bayes Classifier

Teorema Bayes digunakan oleh algoritma pembelajaran mesin *naive Bayes* untuk melakukan klasifikasi. Dalam pengolahan teks, algoritma ini menggunakan pengalaman masa lalu untuk memprediksi apa yang akan terjadi di masa depan, yang disebut teorema Bayes..[7]

Metode *Naive Bayes* banyak digunakan dalam teknik klasifikasi, terutama dalam konteks aplikasi X. Data yang diambil dari aplikasi X akan diklasifikasikan menjadi kelas positif dan kelas negatif dengan menggunakan metode ini. Salah satu kelebihan metode adalah mudah digunakan dan sangat akurat. Bentuk *Naive Bayes Classifier* biasanya seperti berikut:[8]

$$P(H|X) = (P(X | H)P(H))/P(X)$$

di mana X adalah data yang tidak dikenali dan H adalah hipotesis bahwa data X adalah kelas khusus.

$P(H|X)$ adalah kemungkinan hipotesis H yang didasarkan pada kondisi X;

$P(H)$ adalah kemungkinan hipotesis H; $P(X|H)$ adalah kemungkinan hipotesis X yang didasarkan pada kondisi H; dan

$P(X)$ adalah kemungkinan hipotesis X yang didasarkan pada kondisi H.

2.7 Confusion Matrix

Confusion matrix adalah teknik yang umum digunakan untuk mengukur tingkat akurasi *data mining*. Ini mengukur berbagai kualitas seperti akurasi, presisi, *recall*, spesifikasi, dan skor F1 untuk mengetahui seberapa baik kinerja pemodelan sebelumnya.[9]

Matriks ini membandingkan nilai aktual dari data dengan nilai prediksi yang dihasilkan oleh model. Manfaat dari matrix ini yaitu mengevaluasi akurasi model, mengidentifikasi kesalahan yang dibuat oleh model, membandingkan kinerja model yang berbeda. Matriks ini mudah dipahami dan dapat memberikan informasi yang berharga tentang bagaimana model klasifikasi bekerja. Berikut contoh *confusion matrix*[10]

Tabel 2.1 confusion matrix

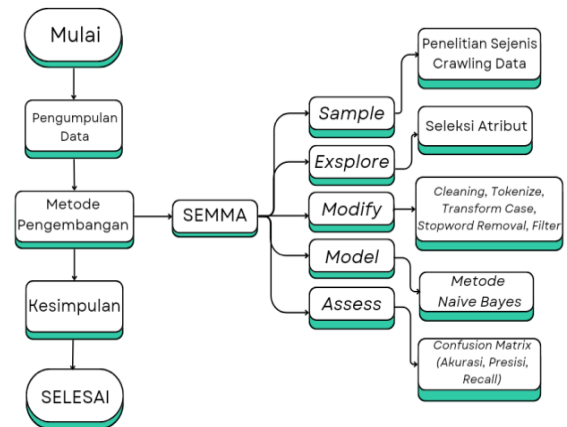
NILAI PREDIKSI	NILAI AKTUAL	
	<i>positive</i>	<i>Negative</i>
<i>Positive</i>	<i>True positive</i>	<i>False positive</i>
<i>Negative</i>	<i>False negative</i>	<i>True Negative</i>

2.8 Rapid Miner

Aplikasi ini merupakan perangkat lunak dari Blanchardstown Polytechnic Dibuat oleh Dr. Mark Hoffman. Aplikasi ini bersifat *open source* menggunakan bahasa pemrograman Java. *Rapidminer* tersedia untuk sistem operasi apa pun. *Rapidminer* juga menyediakan fasilitas yang dibutuhkan untuk menganalisis data, sehingga pengguna tidak harus bisa coding[11] *Rapidminer* biasanya digunakan untuk Menganalisis data dalam jumlah besar. Antarmuka pengguna *Rapidminer* bersifat user friendly adalah salah satu kelebihan aplikasi ini. Algoritma yang digunakan Disimpan dalam bentuk operator.[12]

3. METODE PENELITIAN

Gambar berikut adalah langkah-langkah dalam proses penelitian dari awal hingga akhir:



Gambar 3.1 metodologi penelitian

3.1. Pengumpulan Data.

3.1.1. Studi literatur

Pada penelitian ini terdapat data primer dan sekunder. Data primer pada penelitian ini didapatkan pada aplikasi X sedangkan data sekunder didapat dari jurnal, buku, internet dan penelitian sejenis.

3.1.2. Crawling Data

Proses otomatis untuk mengumpulkan dan mengindeks data dari berbagai sumber, seperti dokumen, database, dan situs web, disebut crawling data. Pada penelitian ini pengumpulan data dilakukan pada aplikasi X yang dihubungkan dengan Rapidminer

3.2 SEMMA

Penelitian ini menggunakan metode SEMMA *data mining proses* yang didalamnya terdapat algoritma *Naive Bayes*. Metode SEMMA berfokus pada memodifikasi penambahan data dan pemodelan. Sedangkan untuk klasifikasi menggunakan algoritma *Naive Bayes Classifier*. Dengan dilakukannya tahapan tersebut maka didapatkan hasil analisis sentimen terhadap program makan siang gratis.

4. HASIL DAN PEMBAHASAN

4.1 Sample

4.1.1 Penelitian Sejenis

Dalam penelitian sejenis penulis telah melakukan pencarian terhadap landasan teori dan penelitian sejenis tentang analisis sentimen. Didapatkan beberapa teori dan landasan mengenai metode SEMMA dan *Naive Bayes Classifier*

4.1.2 Crawling Data

Crawling data dilakukan pada aplikasi X, pengambilan data dengan kata kunci “program

makan siang gratis” berupa teks yang dikirim dalam cuitan oleh pengguna aplikasi X. Data yang diambil merupakan cuitan pengguna aplikasi X dari tanggal 14 Februari – 22 Maret 2024 menggunakan *google colab*. Data yang telah diambil berjumlah 4038 cuitan.

Gambar 4.1 Data Hasil Crawling

4.2 Explore

Data yang didapatkan berupa file csv yang didalamnya terdapat beberapa tabel diantaranya, *created_at*, *id_str*, *full_text*, *quote_count*, *reply_count*, *retweet_count*, *favorite_count*, *lang*, *user_id_str*, *conversation_id_str*, *username*, *tweet_url*. Data yang diambil hanya atribut *text* saja untuk memudahkan dalam menganalisis data. Aplikasi *rapidminer* digunakan untuk melakukan pemfilteran. Untuk memudahkan analisis, data yang telah diambil akan melalui proses *select atribut*. Proses ini menghilangkan atribut yang tidak diperlukan dan hanya mengambil atribut teks untuk analisis sentimen.

Gambar 4.2 Hasil Pemisahan Atribut

4.3 Modify

Aplikasi *rapidminer* digunakan untuk melakukan modifikasi menggunakan *set text preprocessing*. Tujuan dari teknik ini adalah untuk membantu sistem komputer memahami berbagai jenis data set. Selain itu, fase ini memiliki kemampuan untuk mengubah data yang tidak tersusun menjadi data yang tersusun

dan rapi. *Text preprocessing* terdiri dari proses *cleaning*, tokenisasi, *transform case*, dan *filtering*.

4.2.1 Cleaning

Proses *cleaning* dilakukan untuk menghilangkan kata yang tidak dibutuhkan. Data hasil *crawl* dari aplikasi X masih mengandung *url*, *retweet*, *hashtag*, *mention* dan tanda baca. Tahap ini dilakukan beberapa kali menggunakan *set replace*, setiap *replace* akan membersihkan kata dan data yang berbeda.

Tabel 4.1 Contoh *Cleaning*

Contoh data kotor	Contoh data hasil <i>cleaning</i>
Kenapa pulak asing ikut kwatir program makan siang gratis? Sampai Tarik Dana dari Surat Utang RI, Khawatir Makan Siang Gratis.	Kenapa pulak asing ikut kwatir program makan siang gratis? Sampai Tarik Dana dari Surat Utang RI Khawatir Makan Siang Gratis

Gambar 4.3 Hasil *Cleaning*

Dapat dilihat beberapa perbedaan pada gambar 4.3 seperti *url* telah dihilangkan, tanda baca telah dihilangkan, *retweet* telah dihilangkan, *hashtag* telah dihilangkan dan *mention* telah dihilangkan. Maka dengan itu telah didapatkan data yang bersih untuk proses analisis sentimen selanjutnya.

4.2.2 Tokenize

Dalam proses *tokenize*, data yang telah bersih akan kembali melalui tahap tokenisasi. Suatu kalimat akan dipecahkan menjadi beberapa suku kata untuk mempermudah tahap

selanjutnya dalam pengolahan kata. Setiap kata akan disusun dari mulai huruf A sampai Z

Row No.	A	AATA	ADU	ADPM	ADRI	ADPMCA	ADA	ADLAN	ADU
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0

Gambar 4.4 Hasil *Tokenize*

Data yang sebelumnya masih berupa kalimat telah difilter berdasarkan setiap suku kata. Pada hasil tersebut dapat diketahui bobot kata di setiap kalimat atau cuitan.

4.2.3 Transform Cases

Pada data diatas setiap kalimat memiliki huruf yang berbeda. Terdapat beberapa kalimat yang memiliki huruf kapital dan beberapa kalimat yang memiliki huruf kecil. Pada proses ini kalimat atau cuitan tersebut akan menggunakan huruf kecil semua untuk meminimalisir *rapidminer* membaca data yang ganda.

Row No.	test	AATA	ADU	ADPM	ADRI	ADPMCA	ADA	ADLAN	ADU
1	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
2	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
3	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
4	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
5	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
6	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
7	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
8	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
9	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
10	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
11	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
12	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
13	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
14	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0

Gambar 4.5 Hasil *Transform Cases*

Setelah dilakukan *transform cases*, data yang dihasilkan akan memiliki huruf kecil semua.

4.2.4 Stopword Removal

Dalam tahap ini kata yang tidak penting akan dihilangkan untuk mengetahui inti dari sebuah teks yang akan di analisis. Proses ini menggunakan operasi *filter stopwords dictionary* yang didalamnya sudah dimasukkan *stopword* bahasa Indonesia. Kata – kata imbuhan yang tidak diperlukan akan dihilangkan untuk mengetahui inti dari sebuah opini.

Tabel 4.2 Contoh *Stopword Removal*

Sebelum	Sesudah
pendidikan adalah kunci keberhasilan masa	Pendidikan kunci keberhasilan masa

masa depan terima kasih kepada pak jusef hamka dan para pengusaha yang turut mendukung program makan siang gratis bagi siswa pelajar bersama kita membuka pintu menuju masa depan yang lebih baik	depan terima kasih pak jusef hamka pengusaha turut mendukung program makan siang gratis siswa pelajar membuka pintu masa depan lebih baik
---	---

4.2.5 Filtering

Filtering dilakukan dengan beberapa tahap menggunakan set token by length dan filter example. Dalam tahap ini data akan difilter berdasarkan panjang kata, kata yang kurang dari 4 huruf dan lebih dari 25 huruf akan dihilangkan. Tujuan dari proses ini untuk menghilangkan kata terlalu pendek dan panjang.

Row No.	test	AATA	ADU	ADPM	ADRI	ADPMCA	ADA	ADLAN	ADU
1	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
2	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
3	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
4	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
5	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
6	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
7	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
8	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
9	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
10	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
11	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
12	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
13	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
14	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0

Gambar 4.6 Hasil *Filtering Token by Length*

Row No.	test	AATA	ADU	ADPM	ADRI	ADPMCA	ADA	ADLAN	ADU
1	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
2	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
3	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
4	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
5	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
6	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
7	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
8	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
9	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
10	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
11	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
12	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
13	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0
14	pendidikan adalah kunci keberhasilan masa	0	0	0	0	0	0	0	0

Gambar 4.7 Hasil *Text Preprocessing*

setelah dilakukan *text preprocessing*, data yang awalnya berjumlah 4038 yang masih kotor telah dibersihkan sehingga didapatkan data bersih 2909

4.2.6 Wordcloud

Tahap selanjutnya akan dilakukan visualisasi berdasarkan kata yang sering muncul dalam sebuah cuitan. Kata yang sering muncul akan diurutkan berdasarkan jumlah dari yang terbesar.

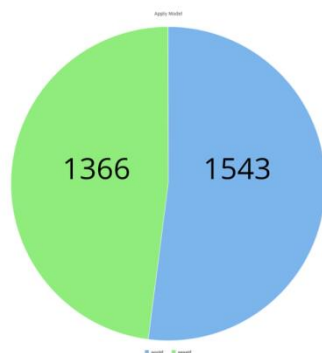
Pada visualisasi kali ini, data hanya diambil dari 20 terbesar saja.



Gambar 4.8 Wordcloud

4.4 Model

Tahap ini dilakukan dengan memberi label sentimen positif dan negatif pada setiap cuitan. Pemberian label secara manual sebanyak 730 atau 25% dari keseluruhan data yang berjumlah 2909. Terdapat 409 sentimen positif dan 321 sentimen negatif dengan jumlah total keseluruhan data train 730. Pelabelan yang dilakukan oleh algoritma *naïve bayes*. Data train yang sebelumnya telah melalui proses untuk melatih *naïve bayes* memahami data akan disimpan pada model data latih. Data set yang telah melewati proses *naïve bayes* akan keluar sebagai data uji. Didapatkan hasil untuk sentimen positif 1134 dan sentimen negatif 1045 dengan jumlah data test 2179. Jumlah seluruh sentimen positif sebesar 1543 atau 53,04% dan sentimen negatif sebesar 1366 atau 44,95%.



Gambar 4.9 Hasil Analisis

4.5 Assess

Tahap ini dilakukan untuk mengevaluasi hasil yang telah dilakukan oleh metode *naïve bayes*. Dalam tahap ini hasil yang didapatkan berupa *confusion matrix* yang terdiri dari akurasi, presisi dan *recall*. Set *K-fold validation* dengan nilai $K=10$ digunakan untuk menentukan nilai

confusion matrix. Proses *crossvalidation* digunakan untuk melihat seberapa akurasi yang dihasilkan algoritma *naïve bayes* untuk analisis sentimen sebuah data.

	true negatif	true positif	class precision
pred. negatif	331	83	79.95%
pred. positif	78	326	80.69%
class recall	80.93%	79.71%	

accuracy: 80.31% \pm 5.27% (micro average: 80.32%)

Gambar 4.10 Hasil Crossvalidation

Nilai akurasi algoritma *Naive Bayes* 80,31%, yang menunjukkan bahwa model *Naive Bayes* dapat mengklasifikasikan 80,31% data yang benar dan $\pm 5,27\%$ margin error. Untuk hasil presisi pada sentimen positif menghasilkan 80,69% dan sentimen negatif sebesar 79,95%. Sementara itu hasil *recall* yang didapatkan pada proses *crossvalidation* untuk sentimen positif sebesar 79,71% dan sentimen negatif 80,93%.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis sentimen pengguna aplikasi X terhadap program makan siang gratis dengan *Naïve bayes Classifier* dan menggunakan metode SEMMA *data mining process* diperoleh kesimpulan sebagai berikut:

- Telah melakukan analisis sentimen media sosial X terhadap program makan siang gratis dengan algoritma *Naïve Bayes Classifier* dengan jumlah data 2909
- Telah diketahui hasil cuitan sentimen positif sebesar 1543 dan sentimen negatif 1366 pada aplikasi X terhadap program makan siang gratis. Hasil sentimen positif 53,04% dan sentimen negatif 44,95%.
- Didapatkan hasil akurasi 80,31% data yang benar dan $\pm 5,27\%$ margin error, hasil presisi sentimen positif menghasilkan 80,69% dan sentimen negatif sebesar 79,95%, hasil *recall* sentimen positif sebesar 79,71% dan sentimen negatif 80,93%.

5.2 SARAN

Penelitian ini masih belum sempurna sehingga dapat lebih dikembangkan untuk penelitian selanjutnya. Penulis memberikan saran untuk meningkatkan *dataset* yang akan di analisis, menambahkan sentimen netral dan menggunakan operasi *text preprocessing* lainnya yang belum dilakukan pada penelitian ini.

DAFTAR PUSTAKA

- [1] Rusdiaman, D., & Rosiyadi, D. (2019). Analisa Sentimen Terhadap Tokoh Publik Menggunakan Metode Naïve Bayes Classifier Dan Support Vector Machine. *J. Comput. Eng. Syst. Sci*, 4(2), 2502-7131.
- [2] R. R. Cahyani and R. Cahyani, "Analisis Sentimen Pada Media Sosial Twitter Terhadap Tokoh Publik Peserta Pilpres 2019," *Matics*, vol. 12, no. 1, p. 79, 2020, doi: 10.18860/mat.v12i1.8356.
- [3] khoerul ummah, "ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP TOKOH GUS DUR MENGGUNAKAN METODE NAÏVE BAYES DAN SUPPORT VECTOR MACHINE (SVM)," *γ787*, no. 8.5.2017, pp. 2003–2005, 2022.
- [4] N. N. Aini and R. Sugeng Basuki, "PENGARUH ELECTRONIC WORD OF MOUTH MEDIA SOSIAL INSTAGRAM @gartenhaus_co DAN STORE ATMOSPHERE TERHADAP MINAT BELI DI CAFE GARTENHAUS MALANG JAWA TIMUR," *J. Apl. Bisnis*, vol. 6, no. 1, pp. 25–28, 2020.
- [5] R. T. Handayanto and H. Herlawati, "Prediksi Kelas Jamak dengan Deep Learning Berbasis Graphics Processing Units," *J. Kaji. Ilm.*, vol. 20, no. 1, pp. 67–76, 2020, doi: 10.31599/jki.v20i1.71.
- [6] I. Vewawati and B. S. Audit, "Algoritma Naïve Bayes Classifier Untuk Analisis Sentiment Pengguna Twitter Terhadap Provider By.u," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1411, 2022, doi: 10.30865/mib.v6i3.4132.
- [7] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.
- [8] B. Mas Pintoko and K. Muslim, "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier," *e-Proceeding Eng.*, vol. 5, no. 3, pp. 8121–8130, 2018.
- [9] Saputra, I., & Kristiyanti, D. A. (2022). Machine learning untuk pemula. *Bandung: Informatika*.
- [10] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform.*, vol. 5, no. 2, pp. 697–711, 2021.
- [11] Fatmawati, "Perbandingan Algoritma Klasifikasi Data Mining Model C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Diabetes," *J. Techno Nusa Mandiri*, vol. XIII, no. 1, p. 50, 2016.
- [12] A. P. Natasuwarna, "Tantangan Menghadapi Era Revolusi 4.0 - Big Data dan Data Mining," *Sindimas*, vol. 1, no. 1, pp. 23–27, 2019, [Online]. Available: <https://www.sisfotenika.stmikpontianak.ac.id/index.php/sindimas/article/view/530>
- [13] Azzahra, S. N., Ruuhwan, R., & Hikmatyar, M. (2023). Analisis Sentimen Taman Wisata Karang Resik Dengan Metode Naïve Bayes. *E-JURNAL JUSITI: Jurnal Sistem Informasi dan Teknologi Informasi*, 12(2), 200-209.
- [14] Mulyani, E. D. S., Mufizar, T., SM, N. N. F., Pramana, H. J., & Hartiwan, I. (2023). Analisis Sentimen Pengguna Twitter Terhadap Vaksin Sinovac (Covid-19) Dengan Menggunakan Metode Naïve Bayes. *Computer Science Research and Its Development Journal*, 15(1), 12-21.