Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 1033-1042
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5826

Analisis Sentimen Program Makan Gratis Pada Media Sosial X Menggunakan Metode NLP

Wisda Anggriyani[*], Muhammad Fakhriza

Sains dan Teknologi, Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Meda, Indonesia
Email: [1,*]wisdaangraini@gmail.com, [2]fakhriza@uinsu.ac.id
Email Penulis Korespondensi: wisdaangraini@gmail.com
Submitted: 20/08/2024; Accepted: 26/08/2024; Published: 27/08/2024

Abstrak−Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap program makan gratis yang diinisiasi di Media Sosial X. Dengan memanfaatkan metode *Natural Language Processing* (NLP) dengan klasifikasi navie bayes, penelitian ini mengolah data teks yang dikumpulkan dari berbagai komentar dan postingan pengguna di platform tersebut. Data yang terkumpul kemudian diklasifikasikan ke dalam kategori sentimen positif, negatif, dan netral. Proses analisis melibatkan teknik *preprocessing* teks, termasuk tokenisasi, stemming, dan penghapusan stop words untuk meningkatkan akurasi model sentimen. Hasil analisis menunjukkan bahwa sebagian besar pengguna memberikan respons positif terhadap program ini, terutama terkait dengan manfaat sosial yang ditawarkan. Namun, sejumlah sentimen negatif juga terdeteksi, terutama terkait dengan pelaksanaan program dan kualitas makanan yang disediakan. Temuan ini memberikan wawasan yang berharga bagi penyelenggara program untuk memahami persepsi publik secara lebih komprehensif dan untuk melakukan perbaikan di masa mendatang. Penelitian ini juga menyoroti pentingnya penggunaan NLP dalam analisis data media sosial sebagai alat untuk mengidentifikasi dan memahami opini publik dalam skala besar.

Kata Kunci: Analisis Sentimen; Program Makan Gratis; Media Sosial; X; *Natural Language Processing* (NLP)

Abstract−This study aims to analyze public sentiment toward the free meal program initiated on Social Media X. Utilizing Natural Language Processing (NLP) methods klasification navie bayes, this research processes text data collected from various user comments and posts on the platform. The collected data is then classified into positive, negative, and neutral sentiment categories. The analysis process involves text preprocessing techniques, including tokenization, stemming, and stop words removal, to enhance the accuracy of the sentiment model. The analysis results show that most users responded positively to the program, particularly regarding the social benefits it offers. However, some negative sentiments were also detected, primarily related to the program's implementation and the quality of the provided meals. These findings offer valuable insights for program organizers to comprehensively understand public perception and make improvements in the future. This study also highlights the importance of using NLP in social media data analysis as a tool to identify and understand public opinion on a large scale.

Keywords: Sentiment Analysis; Free Meal Program; Social Media; X; Natural Language Processing (NLP)

1. PENDAHULUAN

Analisis sentimen adalah teknik dalam bidang pemrosesan bahasa alami (NLP) yang digunakan untuk mengidentifikasi dan mengkategorikan opini yang diekspresikan dalam sebuah teks [1][2][3]. Analisis sentimen melibatkan beberapa langkah utama, termasuk *preprocessing* teks (membersihkan data dari elemen yang tidak relevan seperti tanda baca dan stop words), tokenisasi (memecah teks menjadi kata-kata atau frasa yang lebih kecil), dan pelabelan sentimen (mengklasifikasikan teks berdasarkan polaritas emosionalnya) [3][4]. Metode *machine learning* seperti Naive Bayes, *Support Vector Machines* (SVM), dan teknik deep learning seperti *Recurrent Neural Networks* (RNN) dan Transformers (misalnya BERT) sering digunakan untuk membangun model analisis sentimen [5][6] . Dipenelitian ini saya menggunakan klasifikasi navie bayes yang dimana salah satu algoritma yang paling sederhana dan paling efektif untuk tugas klasifikasi terutama pada pemrosesan teks.

Pada penelitian ini saya menggunakan apliaksi media sosial X yang mana media X ini adalah platform yang ideal untuk analisis sentimen karena karakteristiknya yang real-time dan luasnya jangkauan pengguna. Pengguna seringkali memposting opini mereka tentang berbagai isu, termasuk program makan gratis, yang membuatnya menjadi sumber data yang kaya untuk analisis sentimen. Dengan menggunakan metode NLP, kita dapat mengumpulkan dan menganalisis data dari platform ini untuk mendapatkan wawasan tentang persepsi publik terhadap program tersebut [7][8].

Topik yang yang saya ambil dari penelitian ini yakni program makan gratis. Makan gratis ini merupakan misi dari presiden terpilih Prabowo/Gibran. Dimana banyak sekali isu yang beredar dari program makan gratis ini baik yang program ini tidak layak, anggaran yang di potong, isu bahwa anggaran makan gratis ini menggunakan dana bos dan lainnya. Uji coba program makan gratis ini sudah terlaksana di sekolah dasar. Menu yang disajikan juga bervariasi dari nasi kotak yang berisi lauk, ayam, sayur, buah pisang dan susu [9].

Metode Naive Bayes adalah salah satu metode klasifikasi yang berakar dari teorema Bayes [10]. Metode ini sering digunakan dalam berbagai aplikasi seperti klasifikasi teks, pengenalan pola, dan klasifikasi gambar [11]. Keunggulan utama dari metode Naive Bayes adalah sederhana, cepat, dan efisien dalam penggunaannya. Meskipun memiliki asumsi yang sangat sederhana (naif) tentang independensi antara fitur-fitur, namun dalam banyak kasus metode ini memberikan hasil yang cukup baik [12][13].

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 1033-1042
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5826

Penelitian yang sudah dilakukan [9] berjudul " Sistem Analisis Sentiment Bakal Calon Presiden 2024 Menggunakan Metode NLP Berbasis Web". penelitian yang telah dilakukan untuk menghasilkan pola sentimen terhadap 3 calon kandidat Presiden Indonesia 2024 dengan keyword Ganjar Pranowo, Prabowo Subianto dan Anies Baswedan terhadap 2 sumber digital media, kompas.com dan detik.com, terlihat calon kandidat Ganjar Pranowo unggul untuk sementara ini dari sisi popularitas sebanyak 907 artikel (37.31%), jumlah sentimen positif sebanyak 769 artikel (31.63%) dan spike jumlah pemberitaan terbanyak pada tanggal 27 Mei 2023 sebanyak 55 artikel terhadap 2 sumber digital media tersebut.

Pada penelitian ini data yang akan di analisis adalah opini masyarakat terkait berbagai statement atau komentar tentang Program makan gratis dan susu gratis.

2. METODOLOGI PENELITIAN

2.1 Pengumpulan Data

Data yang digunakan untuk penelitian ini berasal dari 400 ulasan aplikasi X yang dikumpulkan dari aplikasi X. Data ini dikumpulkan menggunakan teknik scrapping web, yang merupakan teknik untuk mengekstrak data dalam jumlah besar dari web dan kemudian disimpan dalam bentuk file CSV [14].

2.2 Preprocessing

Pelabelan dilakukan dengan menandai ulasan yang bersifat komentar positif atau negatif secara manual dengan mengacu pada Kamus Besar Bahasa Indonesia (KBBI). Ini merupakan tahapan awal yang dilakukan dalam memproses teks, dimana terdapat beberapa proses didalamnya yaitu data cleaning untuk membersihkan data dengan menghilangkan simbol-simbol tertentu, text processing untuk menganalisis dan menyortir data teks yang tidak terstruktur guna mendapatkan wawasan berharga, case folding merupakan proses mengganti semua huruf besar menjadi huruf kecil pada data, stopword removal yaitu menghilangkan kata yang dianggap tidak berpengaruh dalam kalimat, tokenization memisahkan kalimat menjadi kata per kata, dan stemming menghapus awalan dan akhiran kata untuk menghasilkan kata dasar. Kemudian dilakukan pembagian dataset menjadi 2 jenis yaitu data latih dan data uji [15][16]. Data latih (training) adalah data yang diolah dan hasilnya sebagai prediksi untuk data uji (testing). Penelitian ini menggunakan rasio perbandingan data 80:20.

2.3 TF-IDF

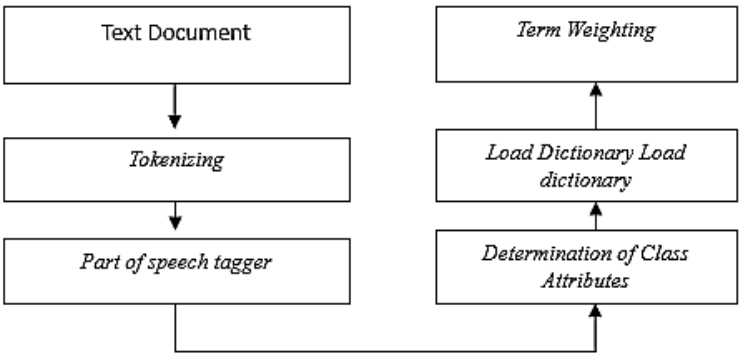
Dalam fungsi ini untuk menghitung nilai setiap kata dengan TF (Term Frequency) dan IDF (Inverse Document Frequency) adalah bagian dari metode TFIDF. Metode ini digunakan dalam berbagai aplikasi pemrosesan teks dan analisis sentimen dan pembobotan dari sampel data yang didapatkan [17].

2.4 Modeling Naïve Bayes

Metode Naive Bayes digunakan untuk mengklasifikasikan ulasan bersifat komentar positif atau negatif. Pada penelitian ini menggunakan metode Naive Bayes untuk analisis sentimen, data ulasan dibagi menjadi dua kelas, yaitu kelas positif dan negatif, dan dilakukan klasifikasi untuk mengidentifikasi ulasan yang termasuk dalam masing-masing kelas [5][18]. Metode Naive Bayes telah terbukti efektif dalam mengklasifikasikan sentimen berdasarkan ulasan pengguna, seperti ulasan film, ulasan aplikasi, dan komentar di media sosial seperti X dan YouTube [19][20].

2.5 Evaluasi

Dalam tahapan evaluasi model melakukan pengujian pada sampel yang dilakukan bertujuan untuk mengetahui performa terbaik dari model tersebut. Hasil performa model disajikan dalam bentuk heatmap, tabel, dan langkah perhitungan matematis terkait berdasarkan pengujian model yang diimplementasikan dan dilatih. Hasil yang dihasilkan jadi bahan evaluasi untuk kedepannya.



Gambar 1. NLP *(Natural Language Processing)*

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 1033-1042
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5826

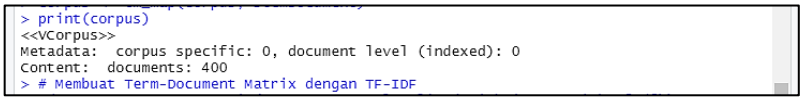
Pada gambar 1, setelah di sediakan text dokumen yang akan dijadikan sebagai file yang digunakan dalam sebuah NLP maka berikut tahapan selanjutnya yang akan dilakukan:

1. *Tokenizing* Proses NLP (*Natural Language Processing*) yang selanjutnya adalah Tokenizing, yaitu proses memotong setiap kata didalam teks dan mengkonversi huruf didalam dokumen menjadi huruf kecil. Didalam proses ini, hanya huruf yang dapat diterima, karakter khusus atau tanda baca dihilangkan.
2. *Part of speech tagger* adalah sebuah proses yang dapat memberikan kelas pada sebuah kata. Dalam proses POS tagger dilakukan dengan cara parsing, setelah itu ditentukan kelas setiap kata dengan menggunakan bantuan kamus yang dibuat sendiri berdasarkan Kamus Besar Bahasa Indonesia (KBBI).
3. *Determination of Class Attributes* Setelah dilakukan Preprocessing, dari data artikel akan ditentukan class attribute. Class attribute yang dibahas pada penelitian ini, diantaranya positif, negatif, dan netral. Dengan adanya class attribute ini diharapkan dapat memberikan penilaian kepada masyarakat secara akurat terhadap suatu objek tertentu.
4. *Load Dictionary Load dictionary* merupakan proses pemanggilan kamus, dengan menggunakan kata kunci yang dapat menunjukkan sentimen positif, negatif, netral, atau mentelusuri bahasa gaul pada frasa. Proses load dictionary dilakukan agar dapat melakukan klasifikasi kata sesuai dengan makna aslinya dari kata yang dimaksud.
5. *Term Weighting* Setelah semua kata diklasifikasikan ke dalam kategori kata positif, negatif, dan netral. Kemudian akan dilakukan proses penghitungan dengan melakukan akumulasi bobot setiap kata dari sebuah kalimat opini.

3. HASIL DAN PEMBAHASAN

3.1 Pra-pemrosesan Teks

Pada penelitian ini menggunakan 400 data text komen dari social media x dengan keyword MAKAN GRATIS ANGGARAN 7500 , berserta label statement dari komen tersebut. Tahapan pra-pemrosesan dilakukan dengan Case Folding yaitu mengubah semua teks menjadi huruf kecil. Selain itu ada dapat menghapus Tanda Baca dan Angka dengan membersihkan teks dari tanda baca dan angka yang tidak relevan. Kemudian tahapan stopword Removal yaitu menghilangkan kata-kata umum yang tidak signifikan. Dan terakhir itu ada stemming dengan mengubah kata menjadi bentuk dasar.



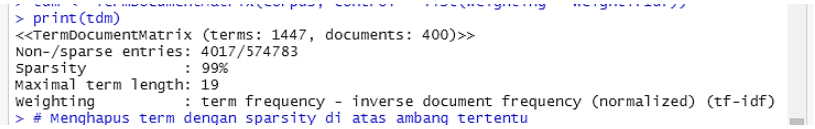
Gambar 2. Hasil *Output* tentang korpus

Berdasarkan gambar 2 diatas, Output tersebut menunjukkan informasi dasar tentang korpus yang terdiri dari Metadata dimana corpus specific “0” ini berarti korpus Anda tidak memiliki metadata spesifik yang terkait dengan seluruh korpus. document level (indexed): 0: Ini menunjukkan bahwa dokumen dalam korpus Anda tidak memiliki metadata individual yang diindeks, seperti judul atau tanggal. Sealin itu pada gambar juga terdapat “Content: documents: 400:” yang berarti korpus Anda terdiri dari 400 dokumen teks, yang merupakan isi dari korpus tersebut.

Secara keseluruhan, ini adalah cara R memberi tahu Anda bahwa korpus berhasil dibuat, terdiri dari 400 dokumen teks, dan tidak memiliki metadata tambahan. Anda bisa melanjutkan dengan pemrosesan lebih lanjut seperti membangun Term-Document Matrix atau melatih model Naive Bayes.

3.2 Term-Document Matrix (TDM) dengan TF-IDF:

TDM dibangun menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk memberi bobot pada term yang lebih signifikan. Output TDM dapat dilihat pada gambar 3 berikut.



Gambar 3. *Output* dari *print*(tdm)

Output dari print (tdm) memberikan informasi tentang Term-Document Matrix (TDM) yang telah Anda buat dari korpus teks, seperti “TermDocumentMatrix (terms: 1447, documents: 400)” yang berarti bahwa ada 1447 istilah unik (kata atau token) dalam TDM Anda. Selain itu TDM mencakup 400 dokumen, sesuai dengan jumlah dokumen dalam korpus Anda. Pada gambar 3 tersebut juga menjelaskan Non-/sparse entries “4017/574783” dimana Non-sparse entries (4017) berarti jumlah entri non-nol dalam TDM. Ini adalah jumlah pasangan term-dokumen yang memiliki nilai non-nol. Selain itu ada juga Sparse entries (574783) yang berarti

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 1033-1042
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5826

jumlah entri nol dalam TDM. Ini adalah jumlah pasangan term-dokumen yang tidak ada (kata tersebut tidak muncul dalam dokumen tersebut). Bagian "Sparsity 99%" yang berarti tingkat kelangkaan TDM, yang menunjukkan persentase entri nol di dalamnya. Dalam kasus Anda, TDM memiliki 99% sparsity, yang berarti sebagian besar pasangan term-dokumen adalah nol (kata tersebut tidak muncul di sebagian besar dokumen). Sealin itu, "Maximal term length: 19" menunjukkan panjang maksimal dari term (kata atau token) yang ada dalam TDM adalah 19 karakter. Bagian terakhir Weighting menunjukkan bobot dalam TDM menggunakan TF-IDF, yang memberikan nilai berdasarkan frekuensi term dalam dokumen dan seberapa umum term tersebut di seluruh korpus. TF-IDF membantu untuk menyoroti term yang lebih penting dalam konteks tertentu.

Tingginya sparsity "Sparsity yang Tinggi (99%)" menunjukkan bahwa sebagian besar kata muncul di sangat sedikit dokumen, yang merupakan hal umum dalam analisis teks. Ini menunjukkan bahwa TDM sangat jarang diisi, dan ada banyak kata yang hanya muncul di sedikit dokumen. Karena sparsity mencapai 99 % maka kita akan mencoba menguranginya.

```
> print(tdm_reduced)
<<TermDocumentMatrix (terms: 28, documents: 400)>>
Non-/sparse entries: 1078/10122
Sparsity           : 90%
Maximal term length: 8
Weighting          : term frequency - inverse document frequency (normalized) (tf-idf)
```

Gambar 4 *Sparsity*

Setelah menerapkan pengurangan sparsity menggunakan removeSparseTerms, hasilnya adalah Term-Document Matrix (TDM) yang berukuran lebih kecil dan lebih terkonsentrasi pada istilah-istilah yang lebih relevan. Berikut penjelasannya. Jumlah istilah unik (kata atau token) dalam TDM Anda telah berkurang menjadi 28 setelah pengurangan sparsity. Documents masih mencakup 400 dokumen, yang merupakan jumlah dokumen dalam korpus Anda. Non-sparse entries (1078) jumlah entri non-nol dalam TDM. Ini adalah jumlah pasangan term-dokumen yang memiliki nilai non-nol, yaitu kata yang muncul dalam dokumen. Sparse entries (10122) jumlah entri nol dalam TDM. Ini adalah jumlah pasangan term-dokumen yang tidak ada (kata tidak muncul dalam dokumen tersebut). Sparsity tingkat kelangkaan TDM Anda sekarang adalah 90%, yang berarti sebagian besar pasangan term-dokumen masih nol, tetapi sudah lebih baik daripada 99% sebelumnya. Maximal term length 8, maximal term length: Panjang maksimal dari term (kata atau token) yang ada dalam TDM adalah 8 karakter.Weighting: term frequency - inverse document frequency (normalized) (tf-idf), weighting bobot dalam TDM menggunakan TF-IDF, yang memberikan nilai berdasarkan frekuensi term dalam dokumen dan seberapa umum term tersebut di seluruh korpus.

$$TF(t,d) = \frac{Jumlah\ kemunculan\ istilah\ t\ dalam\ dokumen\ d}{Jumlah\ total\ istilah\ dalam\ dokumen\ d}$$

Menghitung Inverse Document Frequency (IDF)

$$IDF(t) = log\left(\frac{Jumlah\ total\ dokumen}{Jumlah\ dokumen\ yang\ mengandung\ istilah\ t}\right)$$

Menghitung TF-IDF

$$TF - IDF(t,d) = TF(t,d) \times IDF(t)$$

UNTUK DOKUMEN 1 (Kata "Aja")

$$TF(t,d) = \frac{1}{1}$$

$$IDF(t) = log\left(\frac{400}{120}\right)$$

$$TF - IDF(t,d) = \frac{1}{1} \times log\left(\frac{400}{120}\right) = 0.5253338$$

UNTUK DOKUMEN 2 (Kata "aja" "anggaran" dan "belum")
Kata "aja"

$$TF(t,d) = \frac{1}{3}$$

$$IDF(t) = log\left(\frac{400}{18}\right)$$

$$F - IDF(t,d) = \frac{1}{3} \times log\left(\frac{400}{18}\right) = 0.4502862$$

Kata "anggaran"

$$TF(t,d) = \frac{1}{3}$$

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 1033-1042
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5826

$$IDF(t) = log\left(\frac{400}{19}\right)$$

$$F - IDF(t,d) = \frac{1}{3} \times log\left(\frac{400}{19}\right) = 0.4413239$$

Kata "belum"

$$TF(t,d) = \frac{1}{3}$$

$$IDF(t) = log\left(\frac{400}{8}\right)$$

$$F - IDF(t,d) = \frac{1}{3} \times log\left(\frac{400}{8}\right) = 0.5798420$$

UNTUK DOKUMEN 3 (Kata "aja")

$$TF(t,d) = \frac{1}{1}$$

$$IDF(t) = log\left(\frac{400}{8}\right)$$

$$F - IDF(t,d) = \frac{1}{1} \times log\left(\frac{400}{65}\right) = 0.7880008$$

Lakukan hal yang serupa seperti tahapan diatas terhadap seterusnya sampai 400 dokumen yang ada, berikut gambaran dokuneb yang diperlukan.

RStudio Source Editor — tdm_matrix_reduced

	ada	aja	anak	anggaran	apa	belum	bergizi	bisa	cuman	dan	dapat	dari	dengan	dipangka	grati	ini	itu	jadi	kalo	lagi
1	0.0000000	0.5253338	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
2	0.0000000	0.4502862	0.0000000	0.4413239	0.0000000	0.5798420	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
3	0.0000000	0.7880008	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
4	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
5	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.3584924	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
6	0.0000000	0.3940004	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
7	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.3636364	0.0000000
8	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
9	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
10	0.0000000	0.5253338	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
11	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.2523917	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
12	0.0000000	0.5253338	0.0000000	0.0000000	0.6228276	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.6481614
13	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
14	0.0000000	0.0000000	0.6625193	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.19418741	0.3441705	0.0000000	0.0000000	0.0000000	0.0000000
15	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
16	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
17	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
18	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.6764823	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
19	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
20	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.2459773	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.2666667	0.0000000
21	0.0000000	0.0000000	0.2143445	0.0000000	0.2198215	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
22	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.2347360	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
23	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
24	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
25	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000

	ada	aja	anak	anggaran	apa	belum	bergizi	bisa	cuman	dan	dapat	dari	dengan	dipangka	grati	ini	itu	jadi	kalo	lagi
376	0.0000000	0.5253338	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.35601026	0.0000000	0.0000000	0.0000000	0.6666667	0.0000000
377	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	3.2515388	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
378	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
379	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.4444444	0.0000000
380	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.8000000	0.0000000
381	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.4612075	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
382	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
383	0.0000000	0.6304006	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
384	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.35601026	0.0000000	0.0000000	0.5597423	0.0000000	0.6481614
385	0.8503078	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.42721231	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
386	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
387	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
388	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.35601026	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
389	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.2955944	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.58256224	0.0000000	0.0000000	0.3053140	0.0000000	0.0000000
390	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.2322528	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
391	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.21360615	0.0000000	0.0000000	0.3358454	0.4000000	0.0000000
392	0.0000000	0.6304006	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
393	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
394	0.0000000	0.0000000	0.2429237	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.2705929	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
395	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.23734017	0.0000000	0.0000000	0.3731616	0.0000000	0.4321076
396	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.30515165	0.0000000	0.0000000	0.0000000	0.5714286	0.0000000
397	0.4723932	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.3731616	0.0000000	0.0000000
398	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.4293004	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.4861211
399	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.6764823	0.0000000	0.0000000	0.0000000	0.00000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
400	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.35601026	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000

Gambar 5. Perhitungan DokumenTF-IDF

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 1033-1042
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5826

3.3 Analisis Model Naive Bayes:

Model Naive Bayes dilatih menggunakan data yang sudah diproses, lalu digunakan untuk memprediksi sentimen pada data uji.

```
44
45 # Membagi data menjadi training dan testing set
46 set.seed(123)
47 trainIndex <- sample(1:nrow(tdm_matrix_reduced), 0.8*nrow(tdm_matrix_reduced))
48 trainData <- tdm_matrix_reduced[trainIndex, ]
49 trainLabels <- WISDA$sentiment[trainIndex]
50 testData <- tdm_matrix_reduced[-trainIndex, ]
51 testLabels <- WISDA$sentiment[-trainIndex]
52
53 # Membuat model Naive Bayes
54 model <- naiveBayes(trainData, as.factor(trainLabels))
55 print(model)
56
```

Gambar 6. Prediksi Sentimen Data Uji

Disini kita membagi data dengan perbandingan 80:20, yaitu sebesar 320 data training dan 80 data testing. Berikut adalah wordcloude yang ditambilkan

Gambar 7. Hasil Wordcloud

```
                        apa
as.factor(trainLabels)        [,1]       [,2]
               negatif 0.036049667 0.13528153
               netral  0.021301751 0.07355560
               positif 0.004772625 0.02570137

                        belum
as.factor(trainLabels)       [,1]      [,2]
               negatif 0.01952128 0.1029420
               netral  0.04660935 0.1423978
               positif 0.00000000 0.0000000

                        bergizi
as.factor(trainLabels)       [,1]      [,2]
               negatif 0.05276631 0.2430708
               netral  0.01628006 0.0569047
               positif 0.04871689 0.1402800

                        bisa
as.factor(trainLabels)       [,1]       [,2]
               negatif 0.02366838 0.11687476
               netral  0.01833173 0.07689532
               positif 0.04739825 0.13152140

                        cuman
as.factor(trainLabels)        [,1]       [,2]
               negatif 0.036770405 0.13557878
               netral  0.010147234 0.07175178
               positif 0.006361902 0.03425989

                        dan
as.factor(trainLabels)       [,1]       [,2]
               negatif 0.02068678 0.06981625
               netral  0.03741574 0.09239186
               positif 0.05862871 0.12648701
```

Gambar 8. Output Model Naive Bayes

Hasil dari print(model) menunjukkan output dari model Naive Bayes yang telah Anda buat berdasarkan data pelatihan. Berikut adalah penjelasan dari beberapa bagian penting dalam output tersebut:

1. A-priori probabilities
   Ini adalah probabilitas awal (prior) untuk setiap kelas sebelum mempertimbangkan data. Dalam hal ini, ada tiga kelas sentimen diantaranya negatif, netral dan positif. 75.31% dari data pelatihan memiliki label sentimen negatif, 15.63% dari data pelatihan memiliki label sentimen netral dan 9.06% dari data pelatihan memiliki label sentimen positif. Hal tersebut dapat dilihat pada tabel 1 dibawah ini.

Tabel 1. Hasil probabilitas awal (prior) 1

Sentimen	A-priori
Negative	75.31%
Netral	15.63%

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 1033-1042
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5826

Sentimen	A-priori
positif	9.06%

a. Probabilitas kata "aja" dalam setiap kategori:
   1) Negatif:

      $Likelihood: P("aja" \mid negatif)$=0.044683587

   2) Netral:

      $Likelihood: P("aja" \mid netral)$=0.037825500

   3) Positif:

      $Likelihood: P("aja" \mid positif) = 0.009880887$

b. Menggunakan Naive Bayes untuk Mengklasifikasikan Dokumen Baru:

   $P(Kategori \mid "aja") = P(Kategori) \times P("aja" \mid Kategori)$

   1) Negatif:

      $P(Negatif \mid "aja") = 0.753125 \times 0.044683587 = 0.033644$

   2) Netral:

      $P(Netral \mid "aja") = 0.156250 \times 0.037825500 = 0.005907$

   3) Positif:

      $P(Positif \mid "aja") = 0.090625 \times 0.009880887 = 0.000895$

   Dari perhitungan di atas, dokumen baru ini paling mungkin diklasifikasikan sebagai Negatif karena probabilitasnya yang paling tinggi.

2. Conditional probabilities:
Hal ini adalah probabilitas bersyarat untuk setiap istilah (kata) dalam data pelatihan, diberikan kelas tertentu (negatif, netral, atau positif). Untuk setiap istilah yang muncul dalam data, ada dua kolom yaitu "[,1]" menandakan bahwa probabilitas bahwa istilah tidak muncul dalam dokumen yang memiliki label sentimen tertentu. Selain itu "[,2]" yang menjelaskan probabilitas bahwa istilah muncul dalam dokumen yang memiliki label sentimen tertentu.

Probabilitas kata "ada" tidak muncul dalam dokumen dengan sentimen negatif adalah 0.02117248, dan probabilitas kata "ada" muncul adalah 0.09980221, selain itu Probabilitas kata "ada" tidak muncul dalam dokumen dengan sentimen netral adalah 0.02145122, dan probabilitas kata "ada" muncul adalah 0.07514039, dan Positif: Probabilitas kata "ada" tidak muncul dalam dokumen dengan sentimen positif adalah 0.01832560, dan probabilitas kata "ada" muncul adalah 0.09868637.

A-priori probabilities dalam 'Interpretasi" menunjukkan bahwa sebagian besar dokumen dalam set pelatihan memiliki sentimen negatif, diikuti oleh netral, dan paling sedikit memiliki sentimen positif. Kemudian Conditional probabilities memberikan wawasan tentang bagaimana kemunculan kata tertentu dapat mempengaruhi klasifikasi ke dalam kelas sentimen tertentu. Misalnya, kata "bergizi" memiliki probabilitas tinggi muncul dalam dokumen dengan sentimen negatif dan positif, tetapi lebih jarang dalam dokumen dengan sentimen netral.

3.4 Memprediksi Data Testing

Dalam memprediksi data testing ini kita bisa lihat pada gambar dibawah ini, yang hasilnya bisa berupa postif atau negatif

```
> print(predictions)
 [1] negatif negatif positif positif positif positif positif positif positif positif
[11] positif positif negatif positif positif positif positif positif positif positif
[21] positif positif positif negatif negatif negatif positif positif positif positif
[31] negatif positif negatif positif negatif positif negatif positif positif positif
[41] positif positif positif positif negatif positif positif negatif positif negatif
[51] positif positif negatif positif positif positif positif positif netral  positif
[61] negatif positif netral  positif positif negatif positif positif positif negatif
[71] positif positif negatif positif positif positif positif positif netral  positif
Levels: negatif netral positif
```

Gambar 9. Prediksi Data Testing

Hasil prediksi menunjukkan bahwa model Naive Bayes yang Anda latih telah memprediksi kelas sentimen untuk 80 data uji. Kelas sentimen yang diprediksi adalah negatif, netral, dan positif, sesuai dengan

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 1033-1042
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5826

label yang telah ditentukan. Dari hasil prediksi terlihat bahwa lebih banyak sentiment positif, sedangkan dari data sebenarnya menunjukkan lebih banyak sentiment negative.

3.5 Confusion Matrix , Akurasi, dan metric evaluasi

Confusion matrix dihitung untuk melihat performa prediksi dari model, selain itu akurasi adalah rasio antara prediksi yang benar dengan jumlah total prediksi.

```
> print(evaluation)
Confusion Matrix and Statistics

          Reference
Prediction negatif netral positif
   negatif      12      4       2
   netral        1      1       1
   positif      44     12       3

Overall Statistics

               Accuracy : 0.2
                 95% CI : (0.1189, 0.3044)
    No Information Rate : 0.7125
    P-Value [Acc > NIR] : 1

                  Kappa : -0.0304

 Mcnemar's Test P-Value : 1.043e-10

Statistics by Class:

                     Class: negatif Class: netral Class: positif
Sensitivity                  0.2105        0.05882         0.50000
Specificity                  0.7391        0.96825         0.24324
Pos Pred Value               0.6667        0.33333         0.05085
Neg Pred Value               0.2742        0.79221         0.85714
Prevalence                   0.7125        0.21250         0.07500
Detection Rate               0.1500        0.01250         0.03750
Detection Prevalence         0.2250        0.03750         0.73750
Balanced Accuracy            0.4748        0.51354         0.37162
> |
```

Gambar 10 NLP (Natural Language Processing)

Hasil evaluasi model Naive Bayes Anda menunjukkan beberapa informasi penting terkait kinerja model pada data uji bahwa Model memprediksi 12 data dengan benar sebagai negatif, namun ada 4 yang seharusnya netral dan 2 yang seharusnya positif. Selain oru, Netral Hanya 1 data yang diprediksi benar sebagai netral, dengan 1 data salah diprediksi sebagai negatif dan 1 sebagai positif. Positif: Model memprediksi sebagian besar data (44) sebagai positif, namun dari 44 tersebut, hanya 3 yang benar-benar positif.

Akurasi sebesar 20% - menunjukkan bahwa model hanya memprediksi dengan benar 20% dari keseluruhan data uji. Akurasi ini rendah, yang berarti model mungkin kurang efektif dalam menangani tugas klasifikasi ini. Kappa: -0.0304 - Kappa adalah statistik yang mengukur tingkat kesepakatan antara prediksi model dan label sebenarnya setelah mengoreksi peluang. Nilai negatif menunjukkan bahwa model memiliki kinerja lebih buruk dari prediksi acak. Mcnemar's Test P-Value: 1.043e-10 - Ini menunjukkan perbedaan signifikan antara distribusi kesalahan prediksi, yang berarti bahwa model memiliki bias yang kuat dalam prediksi.

Sensitivity digunakan mengukur kemampuan model untuk benar-benar mendeteksi kelas tertentu. Model memiliki sensitivitas rendah terutama untuk kelas netral dan positif. Specificity: Mengukur kemampuan model untuk mendeteksi data yang tidak termasuk dalam kelas tertentu. Pos Pred Value: Nilai prediksi positif yang rendah terutama untuk kelas positif (0.05085) menunjukkan bahwa model sering salah mengklasifikasikan data sebagai positif. Balanced Accuracy: Akurasi yang disesuaikan dengan mempertimbangkan ketidakseimbangan kelas. Nilai ini juga menunjukkan bahwa model kesulitan untuk menangani ketidakseimbangan data di antara kelas-kelas tersebut. Model Naive Bayes saat ini memiliki kinerja yang kurang memadai, terutama dalam menangani kelas netral dan positif. Menghitung Metrik untuk Setiap Kelas:

True Positives (TP): 12, False Positives (FP): 1 (1 data netral salah diprediksi sebagai negatif) + 44 (44 data positif salah diprediksi sebagai negatif) = 45. False Negatives (FN): 4 (4 data netral yang benar-benar negatif) + 2 (2 data positif yang benar-benar negatif) = 6

$$Precision = \frac{12}{12+45} = \frac{12}{57} \approx 0.2105$$

$$Recall = \frac{12}{12+6} = \frac{12}{18} \approx 0.6667$$

$$F1 - Score = 2 \times \frac{0.2105 \times 0.6667}{0.2105 + 0.6667} \approx 0.3182$$

True Positives (TP): 1, False Positives (FP): 4 (4 data negatif salah diprediksi sebagai netral) + 12 (12 data positif salah diprediksi sebagai netral) = 16. False Negatives (FN): 1 (1 data netral yang benar-benar negatif) + 1 (1 data netral yang benar-benar positif) = 2

$$Precision = \frac{1}{1+16} = \frac{1}{17} \approx 0.0588$$

$$Recall = \frac{1}{1+2} = \frac{1}{3} \approx 0.3333$$

$$F1 - Score = 2 \times \frac{0.0588 \times 0.3333}{0.0588 + 0.3333} \approx 0.1000$$

Kelas Positif:

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 1033-1042
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5826

True Positives (TP): 3, False Positives (FP): 2 (2 data negatif salah diprediksi sebagai positif) + 1 (1 data netral salah diprediksi sebagai positif) = 3. False Negatives (FN): 44 (44 data negatif yang benar-benar positif) + 12 (12 data netral yang benar-benar positif) = 56

$$Precision = \frac{3}{3+3} = \frac{3}{6} = 0.5000$$

$$Recall = \frac{3}{3+56} = \frac{3}{59} \approx 0.0508$$

$$F1 - Score = 2 \times \frac{0.5000 \times 0.0508}{0.5000 + 0.0508} \approx 0.0930$$

Tabel 2 Hasil Ringkasan

Sentiment	Precision	Recall	F1-Score
Negative	21%	66%	31%
Positif	50%	5%	9%
Netral	31%	33%	10%

Nilai-nilai ini memberikan gambaran tentang seberapa baik model dalam mengidentifikasi dan mengklasifikasikan setiap kelas sentimen. Kelas positif memiliki presisi lebih tinggi, tetapi recall yang sangat rendah, sementara kelas negatif memiliki keseimbangan yang lebih baik antara presisi dan recall. Berdasarkan hasil evaluasi dan metrik yang dihitung.

4. KESIMPULAN

Pada penelitian ini peneliti dapat menyimpulkan Akurasi dan Kinerja Umum: Model memiliki akurasi keseluruhan yang rendah (20%), yang menunjukkan bahwa model tidak berhasil memprediksi sentimen dengan baik secara keseluruhan. Akurasi ini jauh di bawah nilai "No Information Rate" (NIR), menunjukkan bahwa model kurang efektif dibandingkan dengan prediksi acak. Kinerja per Kelas: Negatif: Model cukup baik dalam mendeteksi kelas negatif dengan recall yang lebih tinggi (66.67%) tetapi memiliki presisi rendah (21.05%). Ini berarti model sering salah mengklasifikasikan data dari kelas lain sebagai negatif. Netral: Model sangat buruk dalam mendeteksi kelas netral, dengan presisi yang sangat rendah (5.88%) dan recall yang juga rendah (33.33%). Model hampir tidak dapat membedakan data netral dari kelas lain. Positif: Model menunjukkan presisi yang lebih tinggi untuk kelas positif (50%) tetapi dengan recall yang sangat rendah (5.08%). Ini menunjukkan bahwa model jarang benar-benar mendeteksi data positif dan sering mengabaikannya. Dan Nilai F1-score untuk semua kelas sangat rendah, terutama untuk kelas netral dan positif. F1-score mengukur keseimbangan antara presisi dan recall, dan nilai rendah menunjukkan bahwa model mengalami kesulitan dalam mendeteksi kelas-kelas tersebut dengan baik. Pertimbangkan untuk menggunakan model lain selain Naive Bayes, seperti Support Vector Machines (SVM), Random Forest, atau model berbasis neural network yang mungkin lebih cocok untuk data teks ini. Secara keseluruhan, model saat ini memerlukan perbaikan signifikan untuk mencapai kinerja yang memadai dalam klasifikasi sentimen.

REFERENCES

[1] N. Afifa, R. E. Saputra, and R. A. Nugrahaeni, "Implementasi NLP Pada Chatbot Layanan Akademik Dengan Algoritma Bert," *e-Proceedings Eng.*, vol. 10, no. 1, pp. 383–387, 2023.

[2] E. H. Muktafin, K. Kusrini, and E. T. Luthfi, "Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing," *J. Eksplora Inform.*, vol. 10, no. 1, pp. 32–42, 2020, doi: 10.30864/eksplora.v10i1.390.

[3] A. Info, "Analisis Sentimen Pada Ulasan Produk Dengan Metode Natural Language Processing (NLP) (Studi Kasus Zalika Store 88 Shopee)," vol. 17, no. 1, pp. 120–128, 2024.

[4] 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, 2020.

[5] M. Prasetya, M. Wulandari, and S. A. Nikmah, "Implementasi NLP (Natural Language Processing) Dasar pada Analisis Sentiment Review Spotify," *Stain. (Seminar Nas. Teknol. Sains)*, vol. 3, no. 1, pp. 145–153, 2024.

[6] N. Nurwanda, N. Suarna, and W. Prihartono, "Penerapan Nlp (Natural Language Processing) Dalam Analisis Sentimen Pengguna Telegram Di Playstore," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1841–1846, 2024, doi: 10.36040/jati.v8i2.8469.

[7] F. Rayyan, *Pengembangan chatbot untuk aplikasi online chat telegram dengan pendekatan klasifikasi emosi pada teks menggunakan metode indobert-lite*. in Repository.Uinjkt.Ac.Id. 2022.

[8] Nugroho, D. G., Chrisnanto, Y. H., & Wahana, A. (2015). *Analisis Sentimen Pada Jasa Ojek Online menggubakan metode navie bayes (Nugroho dkk.)* 156–161.

[9] A. Mukti, A. D. Hadiyanti, A. Nurlaela, and J. Panjaitan, "Sistem Analisa Sentiment Bakal Calon Presiden 2024 Menggunakan Metode NLP Berbasis Web," *Soscied*, vol. 6, no. 1, p. p-ISSN, 2023.

[10] D. Suryani, A. Yulianti, E. L. Maghfiroh,"Quality Classification of Palm Oil Products Using Naïve Bayes Method," *Sist. J. Sist.* 2021.

[11] Wahyu Sejati, Ankur Singh Bist, and Amirsyah Tambunan, "Pengembangan Analisis Sentimen dalam Rekayasa

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 4, August 2024, Page 1033-1042
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i4.5826

Software Engineering menggunakan tinjauan literatur sistematis," *J. MENTARI Manajemen, Pendidik. dan Teknol. Inf.*, vol. 2, no. 1, pp. 95–103, 2023, doi: 10.33050/mentari.v2i1.377.

[12] M. K. Insan, U. Hayati, and O. Nurdiawan, "Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di," *J. Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 478–483, 2023.

[13] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Comparison of Naïve Bayes and Support Vector Machine Methods in Twitter Sentiment Analysis," *Smatika J.*, vol. 10, no. 02, pp. 71–76, 2020.

[14] S. Fanissa, M. A. Fauzi, and S. Adinugroho, "Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 8, pp. 2766–2770, 2022, [Online]. Available: http://j-ptiik.ub.ac.id

[15] A. Mukti, A. D. Hadiyanti, A. Nurlaela, and J. Panjaitan, "Sistem Analisa Sentiment Bakal Calon Presiden 2024 Menggunakan Metode NLP Berbasis Web The Sentiment Analysis System For the 2024 Presidential Candidates Uses Web-Based NLP Method," vol. 6, no. 1, 2024.

[16] G. F. Grandis, Y. Arumsari, and Indriati, "Seleksi Fitur Gain Ratio pada Analisis Sentimen Kebijakan Pemerintah Mengenai Pembelajaran Jarak Jauh dengan K-Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 8, pp. 3507–3514, 2021.

[17] D. Darmanto, N. I. Pradasari, and E. Wahyudi, "Sistem Deteksi Plagiarisme Tugas Akhir Mahasiswa Berbasis Natural Language Processing Menggunakan Algoritma Jaro-Winkler dan TF-IDF," *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 13, no. 1, pp. 201–211, 2024, doi: 10.30591/smartcomp.v13i1.6375.

[18] K. S. Putri, I. R. Setiawan, and A. Pambudi, "Analisis Sentimen Terhadap Brand Skincare Lokal Menggunakan Naïve Bayes Classifier," *Technol. J. Ilm.*, vol. 14, no. 3, p. 227, 2023, doi: 10.31602/tji.v14i3.11259.

[19] R. Sari and R. Y. Hayuningtyas, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Pada Wisata TMII Berbasis Website," *Indones. J. Softw. Eng.*, vol. 5, no. 2, pp. 51–60, 2020, doi: 10.31294/ijse.v5i2.6957.

[20] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional," *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.