PERBANDINGAN METODE *NAÏVE BAYES* DAN *SUPPORT VECTOR MACHINE* DALAM ANALISIS SENTIMEN PROGRAM MAKAN BERGIZI GRATIS PADA MEDIA SOSIAL X

Dwi Retnosari, Dwi Hartanti, Aprilisa Arum Sari
Teknik Informatika, Universitas Duta Bangsa Surakarta
Jl. Bhayangkara No 55-57 Tipes, Serengan, Kota Surakarta, Jawa Tengah
220103264@mhs.udb.ac.id

ABSTRAK

Perkembangan media sosial, khususnya platform X, mendorong masyarakat untuk menyampaikan opini terhadap berbagai kebijakan pemerintah secara terbuka dan *real-time*. Salah satu kebijakan yang banyak menjadi perhatian publik adalah program makan bergizi gratis. Permasalahan dalam penelitian ini adalah bagaimana menentukan metode klasifikasi terbaik antara *Naïve Bayes* dan *Support Vector Machine* (SVM) dalam menganalisis sentimen masyarakat terhadap program tersebut. Penelitian ini bertujuan untuk membandingkan performa kedua algoritma dalam mengklasifikasikan sentimen positif, negatif, dan netral berdasarkan opini publik di media sosial X. Metode penelitian dilakukan melalui beberapa tahapan, yaitu pengumpulan data komentar dari media sosial X, pelabelan data berbasis leksikon, *preprocessing* data, pembobotan kata menggunakan TF-IDF, serta proses klasifikasi menggunakan algoritma *Naïve Bayes* dan SVM. Dataset kemudian dibagi menjadi data pelatihan dan data pengujian dengan rasio 80:20. Evaluasi model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa metode SVM memperoleh tingkat akurasi sebesar 85.10%, lebih tinggi dibandingkan *Naïve Bayes* sebesar 68.51%. Selain itu, visualisasi word cloud menunjukkan sentimen positif didominasi dukungan terhadap manfaat gizi program, sedangkan sentimen negatif lebih banyak menyoroti pendanaan dan pelaksanaan program.

Kata kunci : *analisis sentimen, Naïve Bayes, Support Vector Machine, TF-IDF, media sosial*

1. PENDAHULUAN

Pertumbuhan pesat teknologi informasi dan komunikasi secara langsung berkontribusi terhadap meluasnya penggunaan media sosial sebagai sarana ekspresi opini publik. Platform seperti X (Twitter) memungkinkan pengguna menyampaikan pandangan terhadap berbagai kebijakan pemerintah secara terbuka dan real-time [1].

Fenomena ini menghasilkan volume data yang sangat besar, yang dapat dimanfaatkan untuk menggali opini publik melalui teknik analisis sentimen [2].

Analisis sentimen merupakan salah satu cabang dalam penambangan data yang bertujuan untuk mengelompokkan opini masyarakat ke dalam kategori positif, negatif, dan netral [3].

Metode ini banyak digunakan untuk memahami persepsi masyarakat terhadap suatu isu secara sistematis dan terukur [4].

Seiring dengan perkembangan teknologi, penerapan analisis sentimen semakin meningkat, terutama dengan dukungan metode pembelajaran mesin yang terus berkembang.

Dalam implementasinya, analisis sentimen sering menggunakan metode *Naïve Bayes* karena kemudahan penerapan serta kemampuannya dalam mengolah data teks secara efisien [5].

Di sisi lain, *Support Vector Machine* (SVM) dikenal memiliki tingkat akurasi yang tinggi, khususnya dalam menangani data berdimensi besar, sehingga menjadi salah satu metode yang banyak digunakan dalam klasifikasi teks. Oleh karena itu, diperlukan suatu perbandingan antara kedua metode tersebut untuk menentukan model yang paling optimal dalam mengklasifikasikan opini publik [6].

Permasalahan utama dalam penelitian ini adalah bagaimana menentukan metode terbaik antara *Naïve Bayes* dan SVM dalam menganalisis sentimen masyarakat terhadap suatu kebijakan pemerintah berbasis data dari media sosial.

Penelitian ini berfokus pada analisis sentimen terhadap program makan bergizi gratis yang dicanangkan oleh pemerintah, yang memiliki potensi besar dalam meningkatkan kesehatan masyarakat, khususnya dalam pemenuhan kebutuhan gizi [7].

Dengan menganalisis opini publik, penelitian ini bertujuan untuk mengetahui tingkat dukungan maupun kritik masyarakat terhadap program tersebut. Selain itu, penelitian ini juga bertujuan untuk membandingkan kinerja algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM) dalam mendeteksi sentimen atau emosi publik berdasarkan data yang diperoleh dari media sosial X.

Sebagai rencana penyelesaian masalah, penelitian ini dilakukan melalui beberapa tahapan, yaitu pengumpulan data dari media sosial X, pelabelan data sentimen, *preprocessing* data, pembobotan menggunakan metode TF-IDF, serta implementasi algoritma *Naïve Bayes* dan SVM untuk proses klasifikasi. Hasil dari kedua metode kemudian akan dievaluasi menggunakan metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score* untuk menentukan metode yang memiliki performa terbaik. Temuan dari

penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan metode analisis sentimen serta menjadi referensi bagi pihak-pihak yang membutuhkan informasi terkait opini publik terhadap suatu kebijakan.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Dengan memanfaatkan kriteria yang telah ditentukan sebelumnya, seperti positif, negatif, dan netral, analisis sentimen adalah metode penambangan teks yang dapat mengidentifikasi, mengekstrak, dan mengklasifikasikan opini atau sentimen yang disampaikan dalam data teks [7].

Melihat data dari platform media sosial adalah cara umum untuk memanfaatkan strategi ini untuk menemukan bagaimana publik melihat suatu isu [8].

Analisis sentimen menggunakan pendekatan pembelajaran mesin untuk meningkatkan akurasi klasifikasi opini.

Alat penting lainnya untuk pengambilan keputusan berbasis data adalah analisis sentimen, yang dapat secara sistematis menilai sejumlah besar data. Menemukan pola dalam opini publik adalah salah satu cara analisis sentimen bisa bantu memahami isu sosial lebih baik.

2.2. Transformasi Data (*TF-IDF*)

Teknik pembobotan kata TF-IDF guna menentukan signifikansi suatu kata dalam teks dalam kaitannya dengan keseluruhan kumpulan kata [9]. Untuk menerapkan strategi ini, ia menghitung berapa kali setiap kata muncul dalam sebuah dokumen dan memberikan bobot yang lebih rendah kepada istilah yang muncul dalam banyak makalah. Akibatnya, TF-IDF dapat ubah teks jadi format numerik yang bisa dipahami algoritma klasifikasinya [10].

Karena bisa memisahkan kata-kata dengan signifikansi yang signifikan dalam sebuah dokumen, TF-IDF telah berguna dalam meningkatkan kinerja model kategorisasi teks. Di sisi lain, pengaturan semantik yang rumit, seperti ironi atau sarkasme, berada di luar cakupan pendekatan kami..

$$\mathrm{TF} - IDF_{(\mathrm{t,d})} = TF_{(\mathrm{t,d})} \times IDF_{(\mathrm{t})}$$

Keterangan :

$f_{t,d}$:frekuensi kata $t$dalam dokumen d
N : total seluruh dokumen
$df(t)$: total dokumen ada kata t

2.3. *Naïve bayes*

Strategi klasifikasi yang dikenal sebagai *Naïve Bayes* menggunakan Teorema Bayes dan bergantung pada probabilitas, beserta asumsi tiap fitur independen. Pada analisis sentimen, metode ini sering digunakan karena reputasinya yang sederhana dan perhitungannya cepat, yang sangat berguna ketika menangani sejumlah besar input teks [11].

Meskipun model *Naïve Bayes* cukup sederhana, namun tetap mampu menghasilkan hasil yang wajar.

Dalam praktiknya, *Naïve Bayes* sering digunakan dengan metode pembobotan seperti TF-IDF untuk meningkatkan akurasi klasifikasi. Kelemahan utama algoritma ini adalah bergantung pada gagasan independensi fitur, yang tidak selalu akurat ketika berurusan dengan data teks. [1].

$$P(x|y) = \frac{P(y|x)\,P(x)}{P(y)}$$

Ket :

P(x|y) : probabilitas kelas x terhadap data y
P(y|x) : probabilitas data y kelas x
P(x) : probabilitas awal kelas
P(y) : probabilitas data

2.4. *Support Vector machine*

Menemukan hyperplane maksimal membagi data pada kelas beda menjadi tugas dari SVM, cara belajar mesin guna klasifikasinya. SVM unggul dalam menangani data teks yang dihasilkan dari TF-IDF dan dataset berdimensi tinggi lainnya [12].

Kemampuan *Support Vector Machine* (SVM) untuk menghasilkan model yang akurat dan sangat dapat digeneralisasi adalah kekuatan utamanya, terutama ketika berurusan dengan dataset yang rumit [1].

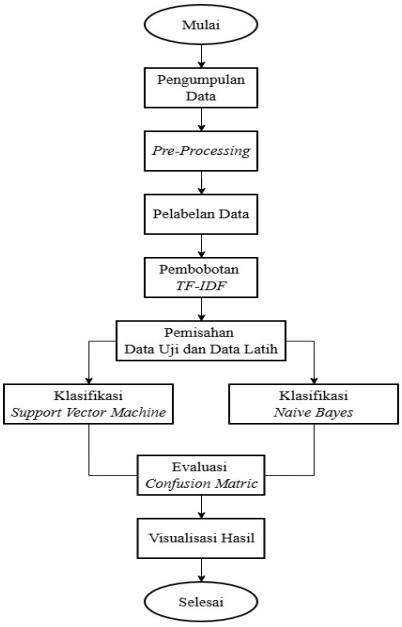
Karena alasan ini, algoritma lain, termasuk *Naïve Bayes*, sering diukur SVM pada studi analisis sentimen.

$$f\ (xd) = \sum_{i=1}^{ns} aiyi\overrightarrow{xi}\ \overrightarrow{xd}\ +\ b$$

Ket :

ns : total *support vector*
ai : skor bobot tiap titik data
yi : kelas data
$\overrightarrow{xi}$: *variable SVM*
$\overrightarrow{xd}$: data nantinya diklasifikasikan
b : skor error atau bias

3. METODE PENELITIAN



Gambar 1. Alur Penelitian

Guna mudahkan aktualisasi studi, dirancanglah kerangka fase studi guna bisa dilaksanakan tertsruktur membuatnya jauh efisien selaras telaksanaanya studi. Kerangkanya digambarkan dibawah.

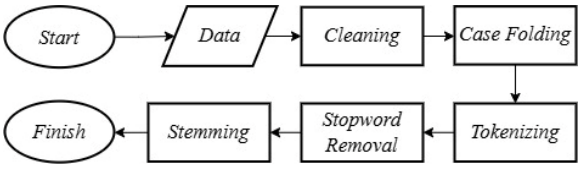
3.1. Pengumpulan Data

Untuk memperoleh wawasan berharga dari penelitian atau analisis data, ini jadi cara pertama wajib dilaksanakan. Guna memastikan hasil berkualitas tinggi, persiapan menyeluruh diperlukan sebelum mengumpulkan data yang relevan, aspek integral dari setiap proses penelitian. Informasi untuk penelitian ini dikumpulkan dari evaluasi program Makanan Bergizi Gratis di platform media sosial X [7] [13].

3.2. Pre-Processing

Langkah pra-pemrosesan tidak dapat lengkap tanpa terlebih dahulu memurnikan data. Data komentar publik di jejaring media sosial X tersedia dalam format .csv dan mencakup beberapa pernyataan yang tidak biasa [14].

Data harus menjalani teknik pra-pemrosesan termasuk pembersihan, pelipatan huruf besar/kecil, tokenisasi, penghapusan stopword, dan stemming [9][15].



Gambar 2. Langkah *Pre-Processing*

3.3. Pelabelan Data

```
def labeling(text):
    text = text.lower()
    score = 0

    for word in positive_words:
        if word in text:
            score += 1
    for word in negative_words:
        if word in text:
            score -= 1

    if score > 0:
        return "positif"
    elif score < 0:
        return "negatif"
    else:
        return "netral"
```

Gambar 3. Pelabelan Data

Dapat dilihat kode merupakan fungsi *labeling(text)* yang digunakan untuk melakukan pelabelan sentimen secara sederhana berbasis kamus kata *(lexicon-based)*. Pada awal fungsi, teks diubah menjadi huruf kecil menggunakan *text.lower()* untuk memastikan proses pencocokan kata tak dapat keberpengaruhan bedanya huruf besar serta kecil. Selanjutnya, variabel score diinisialisasi dengan nilai nol sebagai penampung skor sentimen.

Proses pelabelan dilakukan dengan dua perulangan. Perulangan pertama mengecek setiap kata dalam daftar *positive_words*; jika kata tersebut ditemukan dalam teks, maka skor akan ditambah satu. Sebaliknya, pada perulangan kedua, setiap kata dalam daftar *negative_words* diperiksa, dan jika ditemukan dalam teks, maka skor akan dikurangi satu. Dengan demikian, skor akhir mencerminkan kecenderungan sentimen dari teks berlandaksan total kata positif serta negatif terdeteksi.

Di tahap akhir, dilakukan pengambilan keputusan berdasarkan nilai score. Jika skor > nol, teks dikatakan "positif"; bilamana skor < nol, dikatakan "negatif"; sedangkan bilamana skor sama dengan nol, teks dikategorikan sebagai "netral". Pendekatan ini merupakan metode sederhana yang efektif untuk pelabelan awal data sebelum digunakan dalam proses analisis sentimen lebih lanjut, berikut sampel hasil pelabelan.

Tabel 1. Sampel Hasil Pelabelan Data

	clean_text	label
1678	saya sih tuju dengan orang ini bahwa rakyat su...	negatif
1890	salut buat prabowo yang sadar penting gizi ana...	netral
1343	makan siang ngga ada yg gratis kecuali program...	netral
2839	suasana ceria sambut datang presiden prabowo d...	netral
2213	makan gizi gratis harga bahan pokok aman semua...	positif
3111	gibran mah tugas jumpa fans makan gizi gratis ...	netral
1888	terus support prabowo yang peduli dengan sehat...	netral
2696	nasihat presiden tuh ada ga sih harus ngasih t...	negatif
2575	cerdas dan sehat ala prabowo dukung dong progr...	positif
672	berapa biaya program makan siang gratis per bu...	netral

label	count
netral	1000
positif	759
negatif	317

3.4. Pembobotan *TF-IDF*

Fase ini ialah cara mengganti data teks bentuk data numeriknya guna hitung jenisnya tiap kata pun fitur. Guna memantik TF-IDF seperti.

PEMBOBOTAN TF-IDF

```
from sklearn.feature_extraction.text import TfidfVectorizer

tfidf = TfidfVectorizer(max_features=5000)
X = tfidf.fit_transform(df_balanced['clean_text'])

y = df_balanced['label']
```

Gambar 4. Pembobotan *TF-IDF*

a. Baris *from sklearn.feature_extraction.text import TfidfVectorizer* digunakan untuk mengimpor kelas *TfidfVectorizer* dari pustaka *scikit-learn*, yang berfungsi guna ubah data teks jadi mewakili numerik memakai *TF-IDF*.
b. Baris *tfidf = TfidfVectorizer(max_features=5000)* bertujuan untuk membuat objek *TF-IDF* dengan batas maksimal 5.000 fitur (kata). Pembatasan ini dilakukan agar hanya kata-kata yang paling penting yang digunakan, sehingga dapat meningkatkan efisiensi pemrosesan dan kinerja model.
c. Baris*X=tfidf.fit_transform(df_balanced['clean_text'])* digunakan untuk melakukan dua proses sekaligus, yaitu:
   - *Fitting* : mempelajari kosakata dan menghitung bobot *TF-IDF* dari data teks.
   - *Transforming* : mengubah data teks pada *kolom clean_text* menjadi matriks numerik. Hasilnya disimpan dalam variabel X sebagai fitur yang akan digunakan dalam model.
d. Baris *y = df_balanced['label']* digunakan untuk mengambil data label atau kelas dari setiap teks pada dataset. Variabel y ini berperan sebagai target dalam proses pembelajaran model.
e. Secara keseluruhan, kode tersebut berfungsi untuk mempersiapkan data dalam bentuk fitur (X) dan label (y) yang siap dipakai difase pelatihan klasifikasi.

3.5. Pemisahan Data Uji dan Data Latih

Dataset dibagi menjadi set pelatihan serta pengujian dengan rasio 80:20 guna pastikan validitas pengujian [16].

Dengan menilai kinerja algoritma memakai metrik akurasi, presisi, *recall*, serta F1-score, efektivitas *Naïve Bayes* serta (SVM) [8] [11].

3.6. Klasifikasi *Naïve Bayes*

```
==================================
HASIL EVALUASI MODEL NAÏVE BAYES
==================================
Akurasi   : 68.51%
Precision : 73.95%
Recall    : 68.51%
F1-Score  : 63.72%
```

Gambar 5. Akurasi *Naïve Bayes*

Diatas menampilkan evaluasi kinerja model klasifikasi menggunakan metode *Naïve Bayes*. Berdasarkan output yang ditunjukkan, diperoleh nilai akurasi sebesar 68,51%.

Persentase akurasi perlihatkan model *Naïve Bayes* bisa klasifikasi 68,51 persen data yang diberikan. Lebih dari dua pertiga prediksi model sesuai dengan label sebenarnya, dengan kata lain.

Temuan memperlihatkan kemampuan pengenalan pola model; meskipun demikian, tingkat akurasinya masih rendah.. Karenanya, masih terdapat peluang untuk meningkatkan performa model, misalnya melalui pemilihan fitur yang lebih relevan, penyeimbangan data, atau penggunaan metode klasifikasi lain sebagai pembanding.

3.7. Klasifikasi *Support Vector Machine*

```
==================================
HASIL EVALUASI SUPPORT VECTOR MACHINE
==================================
Akurasi   : 85.10%
Precision : 86.71%
Recall    : 85.10%
F1-Score  : 84.46%
```

Gambar 6. Akurasi *SVM*

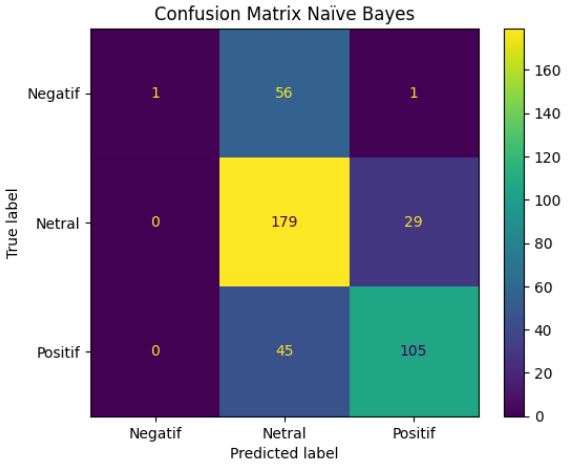
Guna menilai seberapa baik model kategorisasi bekerja, (SVM) digunakan, seperti Gambar 6. Ditemukan akurasinya 85,10%.

Berlandaskan angka akurasi ini, dapat didapa simpulan (SVM) berhasil mengkategorikan sekitar 85,10 persen dari data yang diuji. Sebagian besar prediksi model konsisten dengan label aktual, dengan kata lain.

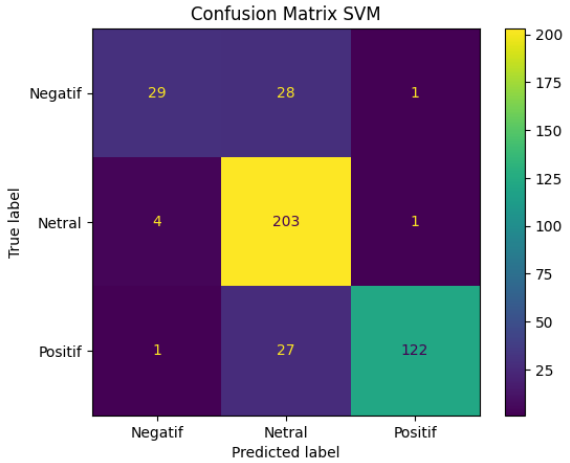
SVM mengungguli strategi klasifikasi *Naïve Bayes*, yang memiliki akurasi lebih rendah, dalam mendeteksi pola dalam data. Penerapan metode SVM hasilkan hasil yang diinginkan untuk teknik kategorisasi studi.

3.8. Evaluasi *Confusion Matrix*

Alat pengukuran berbasis matriks, adalah teknik yang sangat baik untuk mengamati seberapa baik kinerja sistem klasifikasi. Perbandingan hasil dari kedua metode tersebut dirangkum dalam Gambar 7 dan 8. guna melihat hasil evaluasi untuk algoritma *Naïve Bayes* serta SVM pada Gambar 9 serta 10.



Gambar 7. *Confusion Matrix Naïve bayes*



Gambar 8 . *Confusion Matrix Support Vector Machine*

```
============================================================
CLASSIFICATION REPORT
============================================================
              precision    recall  f1-score   support

     Negatif       1.00      0.02      0.03        58
      Netral       0.64      0.86      0.73       208
     Positif       0.78      0.70      0.74       150

    accuracy                           0.69       416
   macro avg       0.81      0.53      0.50       416
weighted avg       0.74      0.69      0.64       416
```

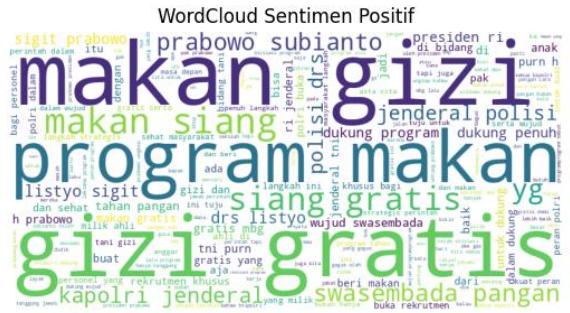
Gambar 9. Hasil Nilai Evaluasi *Naïve Bayes*

```
============================================================
CLASSIFICATION REPORT
============================================================
              precision    recall  f1-score   support

     Negatif       0.85      0.50      0.63        58
      Netral       0.79      0.98      0.87       208
     Positif       0.98      0.81      0.89       150

    accuracy                           0.85       416
   macro avg       0.87      0.76      0.80       416
weighted avg       0.87      0.85      0.84       416
```

Gambar 10. Hasil Nilai Evaluasi *Support Vector Machine*

4. HASIL DAN PEMBAHASAN

4.1. Sentimen Positif



Gambar 11. Hasil Sentiment Positif

Data yang diberi label dengan sikap positif dan kemudian diurutkan ke dalam kategori positif menggunakan analisis sentimen dapat dilihat pada Gambar 11. Data sentimen positif didasarkan pada penentuan frekuensi frasa tertentu dalam dataset. Jauh lebih banyak kemunculan kata-kata "makan," "bergizi," "gratis," dan "program" dibandingkan yang lain. Tampaknya frasa-frasa ini sering digunakan dalam literatur yang lebih optimis. Selain itu, istilah-istilah seperti "makanan," "masyarakat," "anak-anak," dan "kemandirian" membantu memberikan konteks untuk membahas kemampuan program dalam memenuhi kebutuhan nutrisi.

Makanan bergizi gratis merupakan daya tarik utama, sehingga masuk akal untuk berasumsi sebagian besar orang percaya inisiatif fantastis yang bermanfaat bagi masyarakat dan anggotanya. Awan kata ini jelas visi serta kebermanfaatan program jadi pendorong utama suasana hati yang positif.

4.2. Sentimen Negatif

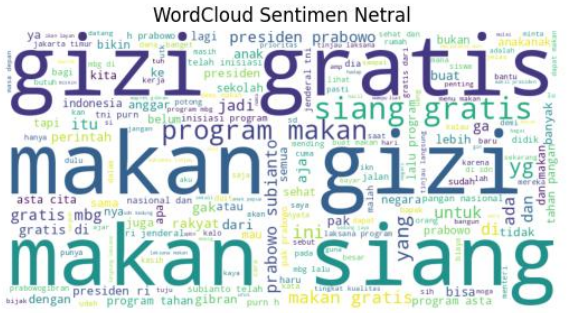
Gambar 12. Hasil Sentiment Negatif

Setelah melakukan analisis sentimen pada data sentimen negatif yang diambil dari label, hasilnya ditunjukkan pada Gambar 12. Untuk mendapatkan data sentimen negatif, kami menghitung jumlah frasa yang sering digunakan dalam sampel.

Kata-kata seperti "yang," "makan," dan "bebas" termasuk di antara opini negatif yang paling sering digunakan. Selain itu, kata-kata seperti "tidak," "kurang," "janji," "bendatan," dan "masalah" menunjukkan bahwa program yang dibahas tidak diterima dengan baik atau sedang dikritik.

Tampaknya sebagian besar kritik terhadap program tersebut berputar pada kekurangannya, implementasi yang tidak merata, dan janji yang tidak terpenuhi, mengingat frekuensi istilah-istilah ini. Awan kata semacam ini menggambarkan korelasi antara permusuhan publik serta ketidakpuasan terhadap efektivitas serta implementasi programt.

4.3. Sentimen Netral



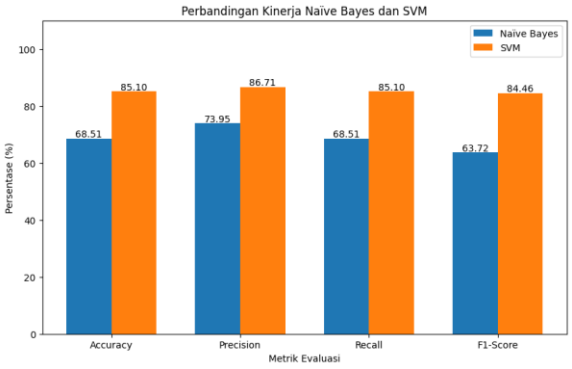
Gambar 13. Hasil Sentiment Netral

Data sentimen netral yang dikumpulkan dari pelabelan dipindahkan ke kategori netral menggunakan analisis sentimen, seperti yang diilustrasikan pada Gambar 13. Data sentimen netral dihasilkan dengan menganalisis frekuensi frasa dalam dataset.

Frekuensi data berbanding lurus dengan ukuran, sehingga istilah yang lebih besar seperti "makan," "siang," dan "gratis" tampak lebih besar dalam bentuk ini. Akibatnya, ungkapan-ungkapan ini sering kali digunakan dalam konteks ketika nadanya netral.

Beberapa kata lain, seperti "program," "anak," "gizi," dan "sekolah," menyiratkan bahwa topik yang dibahas adalah program makan siang gratis. Poin utama biasanya diangkat dalam pertemuan netral tanpa mengungkapkan sentimen kokoh, positif serta negatif.

4.4. Diagram Batang



Gambar 14. Diagram Batang

Naïve Bayes serta (SVM), dua pendekatan klasifikasi, disandingkan Gambar 14, sebuah grafik batang yang menampilkan akurasi, presisi, recall, dan skor F1, dua kriteria evaluasi. Metode SVM berkinerja lebih baik daripada *Naïve Bayes* dalam setiap evaluasi. Disandingkan skor *Naïve Bayes* sekitar 68,51 % skor akurasi SVM mendekati 85,10 %. Demikian pula, setelah melihat *recall*, presisi, dan skor F1, jelas bahwa SVM adalah pendekatan terbaik. berlandaskan percobaan ini, disimpulkan metode SVM mencapai hasil klasifikasi yang lebih baik daripada *Naïve Bayes*.

5. KESIMPULAN DAN SARAN

Berdasarkan temuan pengkajian serta pembahasan yang telah dilaksanakan, dapat disimpulkan bahwa penerapan metode *Support Vector Machine* (SVM) terbukti lebih unggul dibandingkan dengan *Naïve Bayes* dalam mengklasifikasikan sentimen terkait program makan bergizi gratis pada platform X, dengan selisih akurasi sebesar 16.59%. Algoritme SVM juga menunjukkan stabilitas performa yang lebih baik pada seluruh metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*, khususnya dalam pengolahan data berdimensi tinggi hasil pembobotan TF-IDF.

Selain itu, distribusi sentimen masyarakat terhadap program tersebut didominasi oleh klasifikasi netral sebanyak 1.000 data, diikuti oleh sentimen positif sebanyak 759 data, serta sentimen negatif sebanyak 317 data. Secara umum, opini publik yang bersifat positif banyak berkaitan dengan aspek pemenuhan gizi dan bantuan sosial, sedangkan pandangan negatif masyarakat cenderung menyoroti isu transparansi anggaran serta teknis pelaksanaan di lapangan.

Berdasarkan hasil penelitian yang telah dilakukan, diharapkan studi selanjutnya dapat memperluas cakupan sumber data, tidak hanya terbatas pada platform X, tetapi juga mencakup media sosial lain seperti Instagram dan YouTube guna memperoleh perspektif publik yang lebih mendalam. Selain itu, penggunaan teknik ekstraksi fitur yang lebih modern seperti *Word2Vec* dan *FastText* perlu dicoba untuk menangkap korelasi semantik antar kata secara lebih komprehensif.

Mengingat masih terdapat kendala dalam mendeteksi sarkasme pada metode berbasis kamus, disarankan untuk mengintegrasikan pendekatan pembelajaran mendalam, seperti *Long Short-Term Memory* (LSTM) atau BERT, guna meningkatkan akurasi klasifikasi pada teks yang kompleks. Di samping itu, perlu dilakukan pemutakhiran secara berkala terhadap daftar kosakata positif dan negatif dalam proses pelabelan agar tetap selaras dengan perkembangan bahasa gaul (slang) serta istilah teknis terbaru yang digunakan masyarakat di media sosial.

DAFTAR PUSTAKA

[1] F. D. Samuel, P. D. Atika, and S. Setiawati, “Analisis Sentimen Masyarakat Terhadap Perkuliahan Daring Di Twitter Menggunakan Algoritma Naive Bayes Dan Support Vector Machine,” *J. Students‘ Res. Comput. Sci.*, vol. 4, no. 2, pp. 261–272, 2023, doi: 10.31599/jsrcs.v4i2.3253.

[2] L. Rahmawati, A. Pratama, and B. Sutrisno, “Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Media Sosial,” *J. Inform. dan Sist. Inf.*, vol. 10, no. 3, pp. 120–129, 2023.

[3] Gishella Septania Al-Husna, Dian Asmarajati, Iman Ahmad Ihsannuddin, and Rina Mahmudati, “Perbandingan Metode Naïve Bayes Dan Support Vector Machine Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi Linkedin,” *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 3, no. 2, pp. 139–144, 2024, doi: 10.55123/storage.v3i2.3602.

[4] M. R. P. Srg, “Global Research and Innovation Journal (GREAT) ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP PENEGAKAN HUKUM DI INDONESIA TAHUN 2024 MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM) Analisis Sentimen , Twitter , Penegakan Hukum , Support Vector ,” vol. 1, pp. 43–51, 2025.

[5] P. N. Amelia, N. B. Aatiqah, and A. Hasibuan, “Public sentiment analysis of government

subsidy policies on Twitter using the Naïve Bayes classifier," vol. 1, no. 1, pp. 1–15, 2026.

[6] N. Hazrati and A. Rahmatulloh, "Comparison of SVM , Naive Bayes , and Logistic Regression for LinkedIn Reviews Sentiment Analysis," pp. 1–11, 2026.

[7] I. T. Morales Eguez, "済無No Title No Title No Title," *Pap. Knowl. . Towar. a Media Hist. Doc.*, vol. 5, no. 2, pp. 40–51, 2014.

[8] R. H. Wijaya, "Comparison of F1-Score Naive Bayes , Logistic Regression , K-Nearest Neighbors , and SVM for Sentiment Classification X in Police Institutions," vol. 6, no. 3, pp. 4196–4206, 2026.

[9] R. T. Aldisa and P. Maulana, "Analisis Sentimen Opini Masyarakat Terhadap Vaksinasi Booster COVID-19 Dengan Perbandingan Metode Naive Bayes, Decision Tree dan SVM," *Build. Informatics, Technol. Sci.*, vol. 4, no. 1, pp. 106–109, 2022, doi: 10.47065/bits.v4i1.1581.

[10] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *J. KomtekInfo*, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.

[11] W. Ningsih, B. Alfianda, R. Rahmaddeni, and D. Wulandari, "Perbandingan Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 556–562, 2024, doi: 10.57152/malcom.v4i2.1253.

[12] S. Ratnaswari, N. C. Wibowo, and D. S. Y. Kartika, "Analisis Sentimen Menggunakan Metode Lexicon-Based Dan Support Vector Machine Pada Presiden Dan Wakil Presiden Indonesia Periode 2024–2029," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 1, pp. 362–368, 2025, doi: 10.23960/jitet.v13i1.5604.

[13] A. A. Firdausi, D. Hartanti, and A. A. Sari, "MENGGUNAKAN METODE CONTENT BASED FILTERING PADA PRODUK," vol. 10, no. 2, pp. 232–237, 2025.

[14] J. Fathurahman, D. Hartanti, and S. Sopingi, "Sentiment Analysis of Joglo Wifi UMKM Service with Naive Bayes Method," *J. Inotera*, vol. 9, no. 2, pp. 370–377, 2024, doi: 10.31572/inotera.vol9.iss2.2024.id379.

[15] P. Pramono, R. Dwi Irawan, and A. Arum Sari, "Comparative Analysis of SQL Injection Attack Clasification Using Naïve Bayes Method And Support Vector Machine (SVM).," *Proceeding Int. Conf. Sci. Heal. Technol.*, pp. 208–217, 2024, doi: 10.47701/icohetech.v5i1.4178.

[16] K. Yoga Pratama, S. Achmadi, and K. Aulia Sari, "Analisis Sentimen Terhadap Makan Bergizi Gratis Pada Sosial Media X Menggunakan Metode Decicion Tree Dengan Pembanding Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 6, pp. 11061–11067, 2026, doi: 10.36040/jati.v9i6.15525.