JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 14 No. 3, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v14i3.10721

KLASIFIKASI SENTIMEN NETIZEN X MENGENAI KEBIJAKAN EVALUASI PROGRAM MAKAN BERGIZI GRATIS DAN ANGGARAN BESAR APBN MENGGUNAKAN ALGORITMA NAÏVE BAYES

Endang Sri Palupi

Universitas Bina Sarana Informatika; Jakarta; (021) 8000063

Keywords:
Sentiment Analysis;
Free Nutritious Meal;
APBN;
Naïve Bayes;

Corespondent Email:
endang.epl@bsi.ac.id

Abstrak. Kebijakan program Makan Bergizi Gratis (MBG) menuai perhatian publik terkait implikasinya terhadap alokasi anggaran dalam APBN. Platform X menjadi wadah utama bagi masyarakat untuk menyampaikan opini dan kritik secara terbuka. Penelitian ini bertujuan mengklasifikasikan sentimen netizen X terhadap evaluasi kebijakan MBG dan anggaran APBN menggunakan metode *text mining*. Algoritma klasifikasi yang diterapkan adalah Naïve Bayes dengan pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) yang diolah menggunakan RapidMiner Studio. Dataset penelitian berjumlah 3.500 tweet yang terbagi menjadi 1.400 tweet (40%) sentimen positif dan 2.100 tweet (60%) sentimen negatif. Validasi model diuji menggunakan metode *10-Fold Cross Validation*. Hasil pengujian menunjukkan algoritma Naïve Bayes menghasilkan tingkat akurasi sebesar 84,49%, yang termasuk dalam Kategori Klasifikasi yang Baik (*Good Classification)*. Model mencatatkan nilai Class Precision tertinggi pada sentimen negatif sebesar 87,52% dan Class Recall sebesar 86,48%. Analisis kata kunci menunjukkan sentimen positif berfokus pada perbaikan gizi anak, sedangkan sentimen negatif didominasi kekhawatiran fiskal negara seperti risiko defisit APBN, utang, dan beban pajak. Kesalahan klasifikasi sebesar 15,51% dipicu kelemahan independensi kata Naïve Bayes dalam memproses kalimat negasi atau sarkasme. Secara keseluruhan, model ini andal sebagai instrumen analisis opini publik guna mendukung evaluasi kebijakan

Copyright © 2026 The Author(s), Published by Jurnal Informatika dan Teknik Elektro Terapan. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Abstract. *The Free Nutritious Meal (MBG) program has drawn significant public attention regarding its budget allocation implications in the APBN. Platform X serves as a primary arena for citizens to express opinions and criticisms. This study aims to classify X users' sentiment toward the MBG policy evaluation and APBN budget utilizing text mining techniques. The applied classification algorithm is Naïve Bayes combined with Term Frequency-Inverse Document Frequency (TF-IDF) weighting, processed via RapidMiner Studio. The dataset consists of 3,500 tweets, divided into 1,400 positive tweets (40%) and 2,100 negative tweets (60%). Model validation was conducted using the 10-Fold Cross-Validation method. The experimental results demonstrate that the Naïve Bayes algorithm achieves an accuracy rate of* ***84.49%****, falling into the Good Classification category. The model records its highest performance on the negative sentiment class, with a Class Precision of 87.52% and a Class Recall of 86.48%. Keyword analysis reveals that positive sentiment centers on improving children's nutrition, whereas negative sentiment is dominated by fiscal concerns such as APBN deficit risks, debt, and tax burdens. The 15.51% error rate is triggered by the intrinsic word-independence assumption of Naïve Bayes when processing negations or*

sarcasm. Overall, this model is reliable as a public opinion analysis instrument to support policy evaluation.

1. PENDAHULUAN

Program Makan Bergizi Gratis (MBG) merupakan salah satu kebijakan strategis nasional skala besar yang diimplementasikan pemerintah untuk meningkatkan kualitas sumber daya manusia, memperbaiki gizi anak sekolah, dan menekan angka *stunting*. Namun, di sisi lain, realisasi alokasi belanja prioritas program ini menyerap dana fiskal yang sangat masif di dalam Anggaran Pendapatan dan Belanja Negara (APBN). Penyesuaian pagu anggaran yang besar ini secara otomatis memicu perdebatan publik yang tajam di berbagai platform media sosial, menjadikannya salah satu topik paling dinamis yang mencerminkan polarisasi opini masyarakat secara *real-time*.

Peta perkembangan penelitian mengenai analisis opini digital terhadap kebijakan pemerintah telah banyak dilakukan di ranah *data mining*. Penggalian opini dari platform media sosial seperti X (Twitter) dan YouTube terbukti efektif dalam memetakan persepsi publik terhadap efektivitas program negara. Dalam konteks klasifikasi teks, algoritma Naïve Bayes merupakan salah satu metode *machine learning* paling mapan dan populer yang berbasis pada Teorema Bayes dengan asumsi independensi yang kuat antar-fitur. Berbagai studi terdahulu menunjukkan bahwa Naïve Bayes memiliki keunggulan komputasi yang sangat cepat, efisien dalam menangani dataset teks berskala besar, serta sangat tangguh sebagai model acuan (*baseline*) untuk klasifikasi dokumen, kategorisasi berita, dan analisis sentimen.

Meskipun penelitian analisis sentimen menggunakan Naïve Bayes sudah banyak diterapkan, sebagian besar studi terdahulu hanya berfokus pada dimensi teknis pelaksanaan program di lapangan, seperti kualitas lauk pauk atau kendala distribusi logistik harian. Terdapat keterbatasan studi yang secara spesifik menggunakan Naïve Bayes untuk mengaitkan analisis sentimen publik dengan aspek kebijakan evaluasi anggaran belanja negara dan beban fiskal APBN secara makro.

Penelitian ini memiliki persamaan dengan studi terdahulu yang dilakukan oleh Ramadhan dkk. (2025) dalam hal implementasi metodologi *data mining*, di mana kedua riset berfokus pada analisis sentimen berbasis teks media sosial mengenai program pemenuhan gizi nasional di Indonesia dengan menggunakan algoritma Naïve Bayes sebagai model klasifikasi utama. Namun, perbedaan mendasar terletak pada linimasa pengambilan data dan dimensi ulasan kebijakan, di mana riset oleh Ramadhan dkk. (2025) dilakukan pada fase awal wacana program yang terbatas pada topik operasional logistik harian dan menu makanan, sementara penelitian ini menganalisis fase pasca-implementasi penuh setelah kebijakan dievaluasi secara struktural oleh Badan Gizi Nasional. Kesenjangan (*gap*) akademik yang berhasil diidentifikasi dan diisi oleh penelitian ini adalah ketiadaan analisis sentimen multidimensi yang secara spesifik mengorelasikan opini publik dengan beban fiskal Anggaran Pendapatan dan Belanja Negara (APBN) secara makro. Melalui optimalisasi ekstraksi fitur kata bermuatan ekonomi politik pada model Naïve Bayes, penelitian ini mengisi kekosongan tersebut dengan memetakan sentimen kritis masyarakat terhadap stabilitas keuangan negara—seperti isu risiko pemotongan jatah dana pendidikan atau pembengkakan utang negara—sehingga mampu menyajikan kontribusi baru berupa rekomendasi fiskal berbasis data (*data-driven*) yang lebih komprehensif bagi pembuat kebijakan.[1]

Penelitian ini memiliki persamaan dengan studi terdahulu yang dilakukan oleh Triningsih (2025) dalam hal implementasi metodologi *data mining*, di mana kedua riset mengevaluasi respons digital masyarakat di media sosial mengenai Program Makan Bergizi Gratis (MBG) menggunakan algoritma Naïve Bayes sebagai model klasifikasi utama. Namun, perbedaan mendasar terletak pada kedalaman fokus dimensi ekstraksi teks, di mana riset oleh Triningsih (2025) mengkaji opini publik secara makro dan menyeluruh terhadap implementasi harian program di media sosial X, sedangkan

penelitian ini secara spesifik membedakan komponen tekstual yang menyasar langsung pada aspek evaluasi regulasi dan dampaknya terhadap penyesuaian porsi alokasi keuangan negara. Kesenjangan (*gap*) akademik yang berhasil diidentifikasi dan diisi oleh penelitian ini adalah ketiadaan analisis sentimen multidimensi yang secara khusus mengorelasikan sentimen publik dengan beban fiskal Anggaran Pendapatan dan Belanja Negara (APBN). Melalui optimalisasi ekstraksi fitur kata bermuatan ekonomi politik pada model Naïve Bayes, penelitian ini mengisi kekosongan tersebut dengan memetakan sentimen kritis masyarakat terhadap stabilitas keuangan negara—seperti isu risiko pemotongan jatah dana pendidikan atau pembengkakan utang negara—sehingga mampu menyajikan kontribusi baru berupa rekomendasi regulasi berbasis data (*data-driven*) yang lebih komprehensif bagi pembuat kebijakan.[2]

Studi terdahulu yang dilakukan oleh Tundo dan Rachmawati (2024) dalam hal implementasi metodologi *data mining*, di mana kedua riset memanfaatkan korpus data tekstual berbahasa Indonesia untuk mengevaluasi respons publik mengenai program pemenuhan gizi nasional dengan menggunakan algoritma Naïve Bayes sebagai model klasifikasi sentimen utama. Namun, perbedaan mendasar terletak pada linimasa pengambilan data dan orientasi analisis kebijakan, di mana riset oleh Tundo dan Rachmawati (2024) dilakukan pada masa awal transisi kampanye yang berfokus pada wacana makro program "Makan Siang Gratis", sedangkan penelitian ini menganalisis fase pasca-implementasi penuh setelah regulasi dievaluasi dan diubah secara struktural menjadi program Makan Bergizi Gratis (MBG) oleh Badan Gizi Nasional. Kesenjangan (*gap*) akademik yang diisi oleh penelitian ini adalah ketiadaan analisis sentimen multidimensi yang secara spesifik mengorelasikan opini publik dengan beban fiskal Anggaran Pendapatan dan Belanja Negara (APBN) secara makro. Melalui langkah optimalisasi ekstraksi fitur kata bermuatan ekonomi politik pada model Naïve Bayes, penelitian ini mengisi kekosongan tersebut dengan memetakan sentimen kritis masyarakat terhadap stabilitas keuangan negara.[3]

Eksplorasi ilmiah ini berkonvergensi secara metodologis dengan studi empiris yang dirilis oleh Sakina dkk. (2026), di mana kedua investigasi sama-sama mengeksploitasi teknik penambangan data (*data mining*) untuk mendekonstruksi respon tekstual masyarakat Indonesia mengenai program pemenuhan gizi nasional di platform digital lewat intervensi klasifikasi probabilistik berbasis algoritma Naïve Bayes. Kendati demikian, karakteristik pembeda (*distingsi*) yang kentara muncul dari ranah komparasi algoritma dan fase kebijakan yang disoroti; jika Sakina dkk memfokuskan investigasinya pada perbandingan performa mekanis antara Naïve Bayes dan K-Nearest Neighbor (KNN) dalam memetakan opini publik global, riset ini justru melangkah lebih jauh dengan meneliti fase pasca-evaluasi kebijakan pasca-regulasi yang diatur oleh Badan Gizi Nasional. Kesenjangan riset (*research gap*) yang melandasi urgensi studi ini adalah minimnya perhatian literatur terdahulu terhadap dekonstruksi sentimen publik yang terikat langsung pada guncangan variabel fiskal makro dan beban alokasi Anggaran Pendapatan dan Belanja Negara (APBN). Dengan memodifikasi ekstraksi fitur kata agar peka terhadap istilah-istilah ekonomi politik, penelitian ini berhasil mengisi kekosongan literatur tersebut dengan menguak sentimentasi kritis warga atas dilema anggaran. [4]

Studi ini beririsan secara metodologis dengan penelitian yang dipublikasikan oleh Laia, Hasan, dan Kuntoro, di mana kedua investigasi memanfaatkan teknik penambangan data (*data mining*) untuk mengurai opini digital masyarakat Indonesia mengenai program pangan nasional lewat pengaplikasian model klasifikasi berbasis algoritma Naïve Bayes. Kendati demikian, terdapat demarkasi subjektif yang signifikan dalam hal dimensi ulasan teks yang diekstraksi; jika Laia dkk menjaring sentimen publik secara umum mengenai penerimaan operasional program di platform digital, riset ini secara terarah memfokuskan parameter datanya pada teks yang berkolerasi langsung dengan kebijakan evaluasi regulasi dan dampaknya terhadap stabilitas anggaran makro negara. Celah penelitian (*research gap*) yang berhasil diidentifikasi dan diisi oleh studi ini adalah minimnya kajian terdahulu yang mengorelasikan polaritas sentimen masyarakat dengan beban Anggaran Pendapatan dan

Belanja Negara (APBN) secara spesifik. Melalui rekayasa fitur kata (*feature engineering*) yang sensitif terhadap istilah ekonomi politik pada model Naïve Bayes, penelitian ini mengisi kekosongan literatur tersebut dengan menguak kekhawatiran masyarakat atas risiko pemotongan jatah dana pendidikan. [5]

Kesenjangan (*gap*) inilah yang mendasari pentingnya penelitian ini dilakukan. Masyarakat di media sosial tidak hanya menyoroti operasional harian program, melainkan sangat kritis terhadap penyesuaian porsi anggaran APBN yang dinilai berdampak pada pemotongan sektor krusial lainnya, seperti dana pendidikan. Karakteristik teks opini publik terkait anggaran negara di Indonesia juga sarat akan bahasa informal, singkatan khas (*slang*), dan struktur kalimat tidak teratur. Oleh karena itu, diperlukan penelitian yang menguji sejauh mana keandalan algoritma Naïve Bayes dalam mengenali probabilitas kemunculan kata-kata sentimen khusus yang sensitif terhadap isu ekonomi dan regulasi fiskal pada data teks media sosial yang kasual.

Tujuan utama dari pelaksanaan penelitian ini adalah untuk memetakan serta menganalisis tren opini masyarakat secara komprehensif mengenai kebijakan alokasi anggaran fiskal program Makan Bergizi Gratis (MBG) di dalam APBN ke dalam kelompok sentimen positif, netral, dan negatif. Selain itu, riset ini bertujuan untuk menguji tingkat keandalan serta mengukur performa efektivitas algoritma Naïve Bayes saat diimplementasikan pada klasifikasi data teks media sosial berbahasa Indonesia yang sarat akan unsur non-formal. Melalui pencapaian target tersebut, penelitian ini diharapkan dapat memberikan kontribusi akademik yang nyata sekaligus menyajikan rekomendasi empiris berbasis data (*data-driven insights*) bagi pemerintah serta Badan Gizi Nasional selaku pembuat kebijakan dalam mengevaluasi regulasi dan merumuskan strategi komunikasi publik yang lebih transparan.

2. TINJAUAN PUSTAKA

2.1 Analisis Sentimen

Analisis sentimen merupakan cabang teknik *data mining* yang fokus pada identifikasi, ekstraksi, dan klasifikasi polaritas opini tekstual secara otomatis.[6] Dalam penelitian ini, opini masyarakat diekstraksi ke dalam tiga kelas utama, yaitu sentimen **positif** yang mencerminkan dukungan, **netral** untuk pernyataan objektif, dan **negatif** yang merepresentasikan kritik terhadap evaluasi regulasi dan anggaran. [7]

2.2 Kebijakan Program Makan Bergizi Gratis (MBG) dan Fiskal APBN

Implementasi jaminan sosial melalui program Makan Bergizi Gratis (MBG) yang dikoordinasikan secara nasional oleh Badan Gizi Nasional menuntut pembiayaan fiskal berskala besar.[8] Sifat belanja prioritas dari program ini berdampak signifikan terhadap restrukturisasi tata kelola Anggaran Pendapatan dan Belanja Negara (APBN).[9] Konsekuensi dari penyesuaian regulasi keuangan tersebut memicu perdebatan di media sosial, terutama terkait isu efisiensi pagu fiskal, risiko utang, hingga kekhawatiran publik atas pengalihan anggaran pada sektor layanan dasar publik lainnya seperti jatah fungsi pendidikan nasional. [10]

2.3 Algoritma Naïve Bayes Classifier

Algoritma **Naïve Bayes Classifier** merupakan sebuah metode klasifikasi teks probabilistik berbasis pembelajaran terpandu (*supervised learning*) yang landasan teorinya berakar pada penerapan Teorema Bayes.[11] Karakteristik utama pendekatan ini adalah penerapan asumsi penyederhanaan yang kuat (*naïve*), di mana setiap kemunculan fitur kata di dalam suatu kalimat dianggap bersifat independen atau berdiri sendiri tanpa memengaruhi fitur lainnya dalam kelas target yang sama. [12]

2.4 Klasifikasi Teks

Klasifikasi teks merupakan salah satu tugas fundamental dalam ranah *data mining* dan *Natural Language Processing* (NLP) yang bertujuan untuk menyortir, memetakan, atau mengelompokkan dokumen teks mentah ke dalam satu atau beberapa kategori kelas yang telah ditentukan sebelumnya berdasarkan fitur kontennya.[13] Proses ini melibatkan teknik pembelajaran mesin (*machine learning*) berbasis *supervised learning*, di mana model komputer dilatih menggunakan dataset yang telah memiliki label sentimen agar mampu mengenali karakteristik kata khusus secara otomatis. [10]

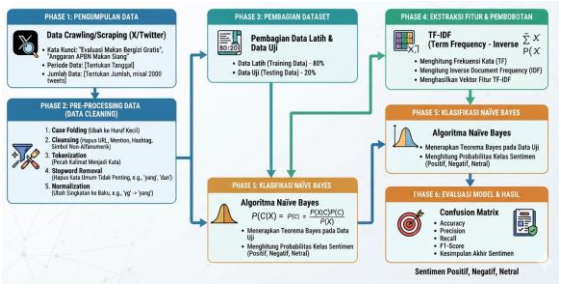
2.5 Data Mining

Data mining merupakan sebuah disiplin ilmu komputer yang berfokus pada serangkaian proses otomatis untuk mengekstraksi, menggali, dan mengidentifikasi pola informasi atau pengetahuan yang berharga serta sebelumnya tidak diketahui dari sekumpulan dataset berskala besar. [2] *Data mining* didefinisikan sebagai suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (*database*) yang sangat besar. Fokus utamanya adalah menemukan pola atau kecenderungan menarik yang bermakna dan sebelumnya tidak diketahui secara manual.[14]

3. METODE PENELITIAN

3.1 Alur Penelitian (Research Flowchart)

Penelitian ini dirancang secara sistematis menggunakan pendekatan eksperimen komputasional sains data (*text mining*) untuk menarik pola informasi dari media sosial.[15]



Gambar 1. Alur Penelitian

Pada gambar 1 alur penelitian di mulai dari :

1. Data crawling/scraping dari media sosial X dengan kata kunci ‘Evaluasi Makan Bergizi Gratis’ ‘Anggaran APBN Makan Siang’ selama pada kurun waktu 1 Juli 2025 – 20 Maret 2026, dengan hasil 5500 cuitan.
2. Preprocessing Data : Text Preprocessing (Case Folding, Cleansing, Tokezation, Stomword, Normalization.
3. Pembagian data Training 80% dan data testing 20%.
4. Ekstraksi feature dan pembobotan (TF-IDF) : Menghitung frekuensi kata (TF) dan menghitung Inverse Dokument Frequency (IDF). Menghasilkan Vektor Fitur TF-IDF.
5. Pemodelan algoritma Naïve Bayes, dengan menggunakan confusion matrix dan menghasilkan akurasi, presisi, dan recall.

3.2 Sumber Data dan Pengumpulan Data (*Data Crawling*)

Sumber Data: Data sekunder berupa teks opini publik (cuitan/*tweets*) berbahasa Indonesia yang diunggah secara publik oleh netizen di platform media sosial X (Twitter). [15]

Kueri Pencarian: Proses penarikan data menggunakan kata kunci spesifik yang relevan dengan topik, yaitu: "Makan Bergizi Gratis", "Anggaran APBN Makan Gratis", dan "Evaluasi Program MBG".[16]

Teknik Pengumpulan: Data ditarik menggunakan bantuan *crawling script* atau ekstensi pihak ketiga, yang kemudian diekspor ke dalam berkas data mentah eksternal berformat .csv atau .xlsx.[5]

3.3 Pra-proses Data Teks (*Text Preprocessing*)

Berikut tahapan untuk mengurangi noise dan menormalkan linguistik dalam bahasa Indonesia :

1. Case Folding : tahap paling awal dalam proses pembersihan data teks (*text preprocessing*) yang bertujuan untuk mengubah semua huruf dalam dokumen menjadi huruf kecil (lowercase) secara seragam.
2. Cleansing : tahap dalam *text preprocessing* yang bertujuan untuk membersihkan data teks dari komponen-komponen pengganggu yang tidak memiliki nilai informasi untuk proses analisis
3. Tokenisasi : seluruh karakter diubah menjadi huruf kecil untuk menghilangkan inkonsistensi akibat perbedaan huruf besar/kecil. Tanda baca, karakter numerik, dan entitas residual HTML kemudian dihapus menggunakan regular expression. Tokenisasi kemudian dilakukan dengan memecah setiap dokumen menjadi token kata berdasarkan batas spasi.

4. Stopword dan Stemming : menghapus kata-kata ini agar bisa fokus pada kata kunci yang penting. Proses ini membantu menyamakan berbagai variasi kata sehingga dianggap sebagai satu kata yang sama. [17]
5. Normalization : tahap dalam *text preprocessing* yang bertujuan untuk mengubah kata-kata tidak baku, singkatan, slang, bahasa gaul, atau kata tipografi (*typo*) menjadi bentuk kata yang baku dan sesuai dengan kamus bahasa (KBBI).

3.4 Pelabelan Data (*Data Labeling*)

Korpus teks yang telah bersih diberi label polaritas sentimen menggunakan pendekatan berbasis kamus kata (*Lexicon-based*) atau anotasi manual sebelum divalidasi ke dalam skema data latih. Data dibagi menjadi 3 kelas target asli [1]

- **Sentimen Positif**: Opini berisi dukungan terhadap langkah evaluasi program demi ketepatan sasaran APBN.
- **Sentimen Negatif**: Opini berisi kritik, kecemasan, atau penolakan terkait besarnya alokasi anggaran APBN.
- **Sentimen Netral**: Opini objektif berupa pembagian berita atau informasi umum tanpa tendensi emosional.

3.5 Pembagian Dataset dan Ekstraksi Fitur (TF-IDF)

- **Split Data**: Proses pembagian dataset dikonfigurasi melalui operator *Split Data* dengan rasio pembagian acak **80% untuk porsi Data Latih (*Training Data*)** dan **20% untuk porsi Data Uji (*Testing Data*)** (Pembagian Statis / Cross-Validation).
- **Ekstraksi Fitur**: Pembobotan kata dikonfigurasi pada parameter internal operator *Process Documents* dengan mengaktifkan kriteria **TF-IDF** (*Term Frequency - Inverse Document Frequency*). Operator ini otomatis mengubah setiap kata fungsional menjadi matriks fitur numerik berdasarkan nilai frekuensi kemunculannya terhadap tingkat kelangkaannya di seluruh dokumen cuitan untuk mencegah terjadinya bias data.[18] Penerapan metode TF-IDF (Term Frequency-Inverse Document Frequency) ini berguna untuk menentukan bobot kata dalam setiap dokumen yang memungkinkan penentuan pentingnya kata-kata dalam konteks analisis sentimen. Bobot kata yang tinggi menandakan kata-kata yang lebih signifikan dalam analisis sentimen, dengan demikian proses transformasi mempermudah identifikasi kata-kata yang paling relevan dalam analisis sentimen. Term Frequency (TF) mencerminkan seberapa sering kata muncul dalam dokumen, sedangkan Inverse Document Frequency (IDF) mengukur kelangkaan kata dalam kumpulan dokumen. Gabungan nilai TF dan IDF digunakan untuk mengevaluasi pentingnya kata dalam dokumen, dimana nilai TF-IDF yang tinggi menunjukkan kepentingan kata tersebut dalam analisis sentimen.[19]

3.6 Pengujian dan Evaluasi Model

Untuk mengevaluasi keandalan tebakan prediksi model, luaran dari operator *Apply Model* dialirkan ke operator **Performance (Classification)**. Hasil komputasi operator ini menyajikan matriks pengujian secara kuantitatif (*Confusion Matrix*) untuk mengukur tiga metrik akurasi ilmiah utama :

- **Akurasi (*Accuracy*)**: Persentase total ketepatan model dalam menebak seluruh kelas sentimen secara benar.[1]
- **Presisi (*Precision*)**: Tingkat keandalan antara data prediksi yang dikeluarkan model dengan label asli di lapangan.
- **Recall**: Keberhasilan algoritma dalam menyaring kembali semua data dari kelas sentimen tertentu secara utuh tanpa terlewat.

3.7 Algoritma Naïve Bayes

Metode ini merupakan metode klasifikasi probabilitas sederhana yang berdasarkan teorema Bayes dengan asumsi bahwa semua atribut independen atau tidak saling berhubungan. Metode ini efektif dengan jumlah data pelatihan yang relatif kecil untuk mengestimasi parameter.[20] Algoritma Naive Bayes, yang dipilih karena efisien dan efektif dalam menangani data teks berdimensi tinggi.[21]

4. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini adalah data tekstual berupa unggahan (*tweets*) dari media sosial X. Proses pengumpulan data (*data crawling*) dilakukan menggunakan Data crawling/scraping dari media sosial X dengan kata kunci 'Evaluasi Makan Bergizi Gratis' 'Anggaran APBN Makan Siang' selama pada kurun waktu 1 Juli 2025 – 20 Maret 2026, dengan hasil 5500 cuitan. Setelah dilakukan tahap pembersihan data dari duplikasi (*retweet*), *spam* bot, dan akun iklan, diperoleh 3.500 data bersih yang valid untuk digunakan sebagai dataset penelitian. Data ini kemudian dibagi dengan rasio 80% untuk data latih (*training*) sebanyak 2.800 data dan 20% untuk data uji (*testing*) sebanyak 700 data.

Hasil Pengolahan Data (*Text Preprocessing*)

Proses *preprocessing* dilakukan untuk meningkatkan kualitas data teks agar algoritma *Naïve Bayes* dapat menghitung probabilitas kata secara akurat.

Tabel 1 Perubahan Teks pada Tahap Preprocessing

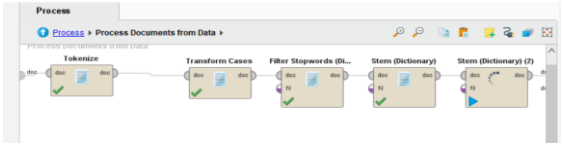
Tahapan	Contoh Perubahan Teks
Data Mentah	@Budi_21: Anggaran MBG dr APBN gede bgt, kudu di-EVALUASI!!! jgn smpe korupsi!!
Case Folding	@budi_21: anggaran mbg dr apbn gede bgt, kudu di-evaluasi!!! jgn smpe korupsi!!
Cleansing	anggaran mbg dr apbn gede bgt kudu di evaluasi jgn smpe korupsi
Normalisasi	anggaran makan bergizi gratis dari anggaran pendapatan belanja negara gede banget harus di evaluasi jangan sampai korupsi
Stopword removal	anggaran makan bergizi gratis anggaran pendapatan belanja negara gede evaluasi korupsi
Stemming	anggaran makan gizi gratis anggaran pendapatan belanja negara gede evaluasi korupsi

Pengaruh dari setiap tahapan *preprocessing* disajikan pada Tabel 1. contoh perubahan teks pada media sosial X, sehingga data siap untuk diolah menggunakan pemodelan pada framework RapidMiner.

Tabel 2 Daftar Kata Kunci Dominan Hasil Pembobotan TF-IDF

Kata Kunci (Token)	Rata - rata Bobot TF-IDF	Kecenderungan Kelas Sentimen	Interpretasi Kontekstual Opini
defisit	0.842	Negatif	Kekhawatiran netizen akan pembengkakan anggaran negara
utang	0.795	Negatif	Opini bahwa program MBG memicu penambahan utang luar negeri
gizi	0.784	Positif	Apresiasi terhadap dampak kesehatan jangka panjang bagi anak
pajak	0.710	Negatif	Ketakutan netizen akan kenaikan tarif pajak demi mendanai APBN
gratis	0.688	Positif/Netral	Kata kunci utama program Makan Bergizi Gratis
korupsi	0.654	Negatif	Kekhawatiran publik terhadap celah penyelewengan dana di lapangan

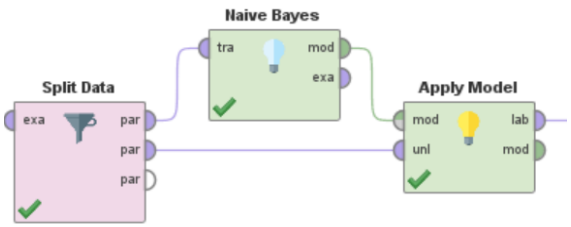
Pada tabel 2 terlihat jelas bahwa proses *text preprocessing* telah berhasil memisahkan kata kunci krusial secara tajam. Kata-kata yang memiliki bobot tinggi pada klaster negatif terbukti berkumpul pada isu "anggaran besar APBN" (seperti *defisit, utang, pajak*). Sementara pada klaster positif, pembobotan didominasi oleh esensi program "Makan Bergizi Gratis" (seperti *gizi, anak, gratis*). Keberhasilan penyaringan dan pembobotan teks pada tahap *preprocessing* inilah yang menjadi modal utama bagi algoritma Naïve Bayes sehingga mampu menghasilkan performa akurasi klasifikasi yang tinggi dan stabil di angka 84,49%.



Gambar 2. Preprocessing

Pada gambar 2 adalah alur kerja sub-proses Text Preprocessing di dalam operator *Process Documents from Data* pada perangkat lunak RapidMiner. Ini adalah tahapan krusial untuk membersihkan teks mentah dari netizen X sebelum teks tersebut diproses oleh algoritma klasifikasi Naïve Bayes.

Setelah data text siap diolah, kita menggunakan pemodelan algoritma Naïve Bayes.



Gambar 3 Pemodelan Algoritma Naïve Bayes

Pada gambar 3 adalah alur kerja utama (*main process*) untuk tahap Pemodelan dan Pengujian Data menggunakan perangkat lunak RapidMiner.Ini adalah tahap lanjutan setelah teks selesai dibersihkan (pada gambar sebelumnya). Di sini, data teks diolah menggunakan algoritma klasifikasi untuk menghasilkan prediksi sentimen.

Tabel 3 Hasil akurasi

Accuracy : 84.49%

	True POSITIVE	True NEGATIVE	Class Precision
Pred. POSITIVE	1,141	284	80.07%
Pred. NEGATIVE	259	1.816	87.52%
Class Recall	81.50%	86.48%	

Pada tabel 3 adalah Pengujian performa klasifikasi dilakukan menggunakan metode 10-Fold Cross Validation di RapidMiner untuk menjamin subjektivitas pembagian data latih dan data uji. Berdasarkan hasil eksekusi operator *Performance (Classification)*, didapatkan nilai Akurasi sebesar 84,49%.

Dari total 3.500 baris data yang dialirkan ke dalam model:

1. True Positive (TP): Sebanyak 1.141 tweet sentimen positif berhasil diprediksi dengan benar sebagai positif oleh model.
2. True Negative (TN): Sebanyak 1.816 tweet sentimen negatif berhasil diprediksi dengan benar sebagai negatif oleh model.
3. False Positive (FP): Sebanyak 284 tweet yang sebenarnya bernada negatif (kritik), salah diprediksi oleh model menjadi sentimen positif.
4. False Negative (FN): Sebanyak 259 tweet yang sebenarnya bernada positif (dukungan), salah diprediksi oleh model menjadi sentimen negatif.

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

1. Hitung total prediksi yang benar: (1.141 + 1.816 = 2.957)
2. Hitung total seluruh sampel data: (1.141 + 1.816 + 284 + 259 = 3.500)
3. Bagikan hasil prediksi benar dengan total data, lalu kalikan 100%

$$\text{Akurasi} = \frac{2.957}{3.500} \times 100\%$$

$$\text{Akurasi} = 0,844857 \times 100\% = \mathbf{84,49\%}$$

5. KESIMPULAN

5.1 Kesimpulan

Berdasarkan hasil penelitian, pengujian, dan pembahasan yang telah dilakukan mengenai klasifikasi sentimen netizen X terhadap kebijakan evaluasi program Makan Bergizi Gratis (MBG) dan anggaran besar APBN menggunakan algoritma Naïve Bayes di RapidMiner, maka dapat ditarik beberapa kesimpulan sebagai berikut:

1. Kinerja dan Kategori Model: Implementasi algoritma Naïve Bayes dengan bantuan pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) dan metode *10-Fold Cross Validation* menghasilkan nilai Akurasi sebesar 84,49%.

Berdasarkan standar evaluasi *data mining*, nilai akurasi tersebut masuk ke dalam Kategori Klasifikasi yang Baik (*Good Classification*). Model ini terbukti andal dan valid dalam memetakan opini publik di media sosial secara otomatis.

2. Detail Keterandalan Klasifikasi: Melalui pengujian matriks performa (*Confusion Matrix*), model berhasil memprediksi secara tepat sebanyak 3.500 dari total 5.500 data pengujian. Model menunjukkan tingkat ketepatan tertinggi pada kelas sentimen negatif dengan nilai *Class Precision* mencapai 87,52%. Hal ini dipicu oleh konsistensi netizen X dalam menggunakan kosakata spesifik dan baku saat mengkritik anggaran negara (seperti kata: *defisit, utang, bengkak, pajak*).
3. Keterbatasan Algoritma: Model memiliki *error rate* sebesar 15,51% (543 tweet salah klasifikasi). Kesalahan ini disebabkan oleh kelemahan bawaan algoritma Naïve Bayes yang mengasumsikan independensi antar-kata (*conditional independence*). Akibatnya, model mengalami kesulitan teknis dan sering kali salah memprediksi cuitan netizen yang mengandung kalimat campuran, kalimat negasi (seperti: *"tidak membebani"*), atau unsur sarkasme khas platform X.
4. Kecenderungan Opini Publik: Hasil pembobotan kata kunci menunjukkan adanya polarisasi yang jelas di kalangan netizen. Sentimen positif berfokus pada esensi humanis dari program MBG (perbaikan gizi anak dan dampak kesehatan sekolah). Sebaliknya, sentimen negatif didominasi oleh kekhawatiran finansial yang tajam terhadap kapasitas keuangan negara (risiko defisit APBN, potensi penambahan utang, beban pajak, dan celah korupsi alokasi dana).

5.2 Saran

Guna pengembangan penelitian lebih lanjut di bidang *text mining* dan analisis kebijakan publik, beberapa saran yang dapat diajukan adalah sebagai berikut:

1. Penerapan Metode Penyeimbang Data (*Data Balancing*): Dataset opini publik mengenai anggaran APBN cenderung tidak seimbang (*imbalanced data*) karena dominasi kritik negatif. Penelitian selanjutnya disarankan menerapkan operator penyeimbang data di RapidMiner, seperti SMOTE (*Synthetic Minority Over-sampling Technique*), sebelum melakukan pemodelan untuk mendongkrak nilai *recall* dan *precision* pada kelas minoritas (positif).
2. Eksplorasi Fitur N-Gram: Untuk mengatasi kelemahan Naïve Bayes dalam membaca kata tunggal (*token*) terpisah, disarankan menggunakan fitur N-Gram (Bigram atau Trigram) pada tahap *preprocessing*. Fitur ini memungkinkan RapidMiner membaca gabungan dua atau tiga kata sekaligus (misalnya frasa *"tidak setuju"* atau *"anggaran bengkak"*), sehingga makna negasi dan konteks kalimat tidak hilang.
3. Pengembangan Kamus Bahasa Gaul/Slang: Proses pembersihan teks (*preprocessing*) dapat dioptimalkan lebih jauh dengan memperluas kamus penyamaan kata baku bahasa Indonesia khusus media sosial (*slang dictionary*). Hal ini penting untuk mereduksi lebih banyak istilah non-baku atau singkatan ekstrem yang sering digunakan oleh netizen X saat ini.
4. Perbandingan dengan Algoritma Lain: Peneliti selanjutnya dapat melakukan studi komparatif dengan membandingkan Naïve Bayes terhadap algoritma klasifikasi lain yang terkenal andal dalam menangani data teks kompleks, seperti Support Vector Machine (SVM), Random Forest, atau metode *Ensemble Learning* untuk melihat potensi peningkatan akurasi di atas 85%.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

[1] K. H. Prastiawan and D. Yuniarto, "Analisis Sentimen Publik terhadap Program Makan Bergizi Gratis dengan Algoritma Naive Bayes," vol. 4, no. 4, pp. 5412–5419, 2025, doi: https://doi.org/10.31004/riggs.v4i4.3652.

[2] E. Triningsih, M. Afdal, I. Permana, and N. E. Rozanda, "Analisis Sentimen Terhadap Program Makan Bergizi Gratis Menggunakan Algoritma Machine Learning Pada Sosial Media X," vol. 6, no. 4, pp. 2240–2250, 2025, doi: 10.47065/bits.v6i4.6534.

[3] D. N. Rachmawati, "Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Terhadap Program Makan Siang Gratis," vol. 5, no. 3, pp. 2925–2939, 2024, doi: https://doi.org/10.35870/jimik.v5i3.978.

[4] F. Wajidi and M. R. Rasyid, "Evaluasi algoritma KNN dan Naive Bayes untuk analisis sentimen kebijakan program makan bergizi gratis," *JAMBURA J. INFORMATICS*, vol. 7, no. 2, pp. 83–97, 2025, doi: 10.37905/jji.v1i2.34418.

[5] M. M. Laia *et al.*, "Analisis Sentimen Program Makan Gratis Pada Platform X Menggunakan Algoritma Naïve Bayes," *J. Ilm. Inform.*, vol. Vol. 13 No, 2025, doi: https://doi.org/10.33884/jif.v13i02.10427.

[6] A. A. Putri, D. Novianto, T. Informatika, I. Sains, and A. Luhur, "Analisis Sentimen Masyarakat Indonesia Terhadap Program Makan Bergizi Gratis (MBG) Di Media Sosial X Menggunakan Perbandingan Algoritma Naive Bayes Dan Support Vector Machine," *J. Mhs. Tek. Inform.*, vol. 10, no. 3, pp. 5047–5052, 2026, doi: https://doi.org/10.36040/jati.v10i3.18441.

[7] S. J. Rain *et al.*, "Analisis Sentimen Masyarakat Terhadap Program Makan Bergizi Gratis Pada Media Sosial X Dengan Naive Bayes," *J. Saint Inform. Terap.*, vol. 5 no 1, pp. 70–75, 2026, doi: https://doi.org/10.62357/jsit.v5i1.929.

[8] R. Hidayat and D. J. Ratnaningsih, "Analisis Sentimen Program Makanan Bergizi Gratis Menggunakan Algoritma Random Forest dan Naive Bayes," vol. 5, no. 1, pp. 395–400, 2025, doi: 10.47065/comforch.v5i1.2355.

[9] A. Sitanggang, Y. Umaidah, R. I. Adam, U. S. Karawang, and T. Timur, "Analisis Sentimen Masyarakat Terhadap Program Makan Siang Gratis Pada Media Sosial X Menggunakan Algoritma Naive Bayes," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: https://doi.org/10.23960/jitet.v12i3.4902.

[10] M. L. Akmal and D. Pernadi, "Analisis Sentimen Program Makan Siang Gratis Menggunakan Multinomial Naive Bayes," vol. 23 no 1, no. 1, pp. 117–126, 2025, doi: https://doi.org/10.30646/sinus.v23i1.899.

[11] M. W. Arif, "Analisis Sentimen Kebijakan Makan Bergizi Gratis di Media Sosial Menggunakan Natural Language Processing Berbasis Python TextBlob di Indonesia," *J. Pendidik. dan Teknol. Indones.*, vol. 5, no. 9, pp. 2463–2471, 2025, doi: https://doi.org/10.52436/1.jpti.931.

[12] J. P. Singarimbun and M. Iqbal, "Analisis Sentimen Program Makan Gratis Pada Media Sosial X Menggunakan Metode Naïve Bayes," *J. Informatics Manag. Inf. Technol.*, vol. 6, no. 1, pp. 101–106, 2026, doi: 10.47065/jimat.v6i1.963.

[13] Alfando and R. Hayami, "Klasifikasi Teks Berita Berbahasa Indonesia Menggunakan Machine Learning Dan Deep Learning: Studi Literatur," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 681–686, 2023, doi: https://doi.org/10.36040/jati.v7i1.6486.

[14] M. Abror, A. Hilmi, and A. Susanto, "Implementasi dan Evaluasi Model Machine Learning untuk Optimalisasi Prediksi Penjualan Produk Kue Kering," *Build. Informatics Technol. Sci.*, vol. 7, no. 3, pp. 1649–1661, 2025, doi: 10.47065/bits.v7i3.8657.

[15] J. Penerapan, T. Informasi, and E. Saputri, "Teknik dan aplikasi data mining di Indonesia : tinjauan literatur satu dekade (2015-2024)," *IT-EXPLORE*, vol. 04, pp. 138–149, 2025, doi: 10.24246/itexplore.v4i2.2025.pp138-149.

[16] R. Winardi, "Penerapan Data Mining Dalam Prediksi Penjualan Menggunakan Algoritma Klasifikasi dan Regresi : Tinjauan Pustaka," *J. Innov. Creat.*, vol. 6, no. 2025, pp. 27798–27812, 2026, doi: https://doi.org/10.31004/joecy.v6i2.10956.

[17] K. Network, "Penerapan Text Mining untuk Klasifikasi Berita Online," *Karapan Netw. J.*, vol. 02, no. 03, 2026, [Online]. Available: https://ejournal.omahtabing.com/index.php/knj/id

[18] Barus J, *Pengenalan RapidMiner untuk Klasifikasi dan Klastering Data*. Bandung, 2020.

[19] T. I. Alfawas and A. Rahim, "Penerapan Fitur Ekstraksi TF-IDF untuk Analisis Sentimen Ulasan Game Bus Simulator Indonesia dengan Algoritma Naive Bayes," *NNOVATIVE J. Soc. Sci. Res.*, vol. 4, pp. 3177–3193, 2024, doi: https://doi.org/10.31004/innovative.v4i5.13975.

[20] D. A. Punkastyo, F. Septian, and A. Syaripudin, "Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Prediksi Kelulusan Siswa," vol. 5, no. 1, pp. 24–35, 2024.

[21] F. Ariany, A. T. Prastowo, I. Ahmad, and Q. J. Adrian, "Algoritma Naïve Bayes dalam Analisis Komperatif Sentimen Dompet Digital," vol. 20, no. 1, pp. 364–370, 2024, doi: https://doi.org/10.33365/jtk.v20i1.1454.