JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 6 No 3 Edisi Mei - Agustus 2025, Page 2182-2194
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Analisis Sentimen Publik atas Kebijakan Efisiensi Anggaran 2025 dengan *Text Mining* dan *Natural Language Processing*

Vina Agustina[1], Asti Herliana[2]

[1,2,] Teknik Informatika, ARS University, Bandung, Indonesia
Email: [1]vinaagustina3711@gmail.com, [2]asti@ars.ac.id
Email Penulis Korespondensi: [1]vinaagustina3711@gmail.com

Abstrak− Kebijakan efisiensi anggaran merupakan langkah strategis pemerintah untuk mengoptimalkan penggunaan belanja negara. Langkah konkret terbaru di tahun 2025, yakni Instruksi Presiden Nomor 1 Tahun 2025 yang dikeluarkan pada tanggal 22 Januari 2025, menetapkan pemangkasan anggaran belanja negara sebesar Rp 306,69 triliun. Namun, implementasi kebijakan ini sering menimbulkan pro dan kontra di masyarakat. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kebijakan efisiensi anggaran tahun 2025 dengan pendekatan *Text Mining* dan *Natural Language Processing* (NLP). Data dikumpulkan dari media sosial Twitter menggunakan teknik web crawling berbasis *Python*, dengan kata kunci tertentu dan filter waktu tertentu, sehingga diperoleh 1.614 tweet yang relevan. Proses *pre-processing* meliputi pembersihan data, *case folding*, tokenisasi dan *stopword removal*. Data kemudian diberi label sentimen secara manual (positif, negatif, netral), dibagi menjadi data latih (70%) dan data uji (30%) dengan teknik stratified sampling, serta ditransformasikan menjadi bentuk numerik menggunakan metode TF-IDF. Hasil klasifikasi menggunakan algoritma *Naive Bayes* menunjukkan bahwa mayoritas sentimen masyarakat bersifat negatif (74,53%), dengan akurasi model mencapai 93,01%. Temuan ini menunjukkan bahwa masih terdapat ketidakpuasan publik terhadap kebijakan tersebut. Penelitian ini memberikan kontribusi dalam pemanfaatan teknologi untuk mendukung pengambilan keputusan berbasis data (*evidence-based policy*), serta dapat menjadi acuan bagi pemerintah dalam merumuskan strategi komunikasi publik yang lebih responsif.

Kata Kunci: Analisis Sentimen, Efisiensi Anggaran, *Text Mining*, NLP, *Naive Bayes*

Abstract− The budget efficiency policy is a strategic measure by the government to optimize the use of state expenditure. The most recent concrete step in 2025 is Presidential Instruction No. 1 of 2025, issued on January 22, 2025, which mandates a cut in state spending amounting to IDR 306.69 trillion. However, the implementation of this policy often generates public debate and controversy. This study aims to analyze public sentiment toward the 2025 budget efficiency policy using a Text Mining and Natural Language Processing (NLP) approach. Data were collected from the social media platform Twitter using a Python-based web crawling technique, with specific keywords and time filters, resulting in 1,614 relevant tweets. The pre-processing stage included data cleaning, case folding, tokenization and stopword removal. Sentiment labeling was performed manually (positive, negative, neutral), followed by data splitting into training (70%) and testing sets (30%) using stratified sampling, and numerical transformation using the TF-IDF method. Classification results using the Naive Bayes algorithm indicate that the majority of public sentiment is negative (74.53%), with a model accuracy of 93.01%. These findings suggest that there remains significant public dissatisfaction with the policy. This study contributes to the utilization of technology in supporting evidence-based policymaking and can serve as a reference for the government in formulating more responsive public communication strategies.

Keywords: Budget Efficiency, Naive Bayes, NLP, Sentiment Analysis, Text Mining

1. PENDAHULUAN

Kebijakan efisiensi anggaran merupakan strategi penting yang diambil oleh pemerintah untuk memastikan bahwa penggunaan sumber daya negara dapat dilakukan secara optimal, efisien, dan berkelanjutan. Langkah konkret terbaru di tahun 2025, yakni Instruksi Presiden Nomor 1 Tahun 2025 yang dikeluarkan pada tanggal 22 Januari 2025, menetapkan pemangkasan anggaran belanja negara sebesar Rp 306,69 triliun terdiri dari potongan Rp 256,1 triliun dari kementerian dan lembaga serta Rp 50,59 triliun dari transfer ke daerah sebagai upaya strategis untuk mengurangi pengeluaran non-prioritas dan menjaga keberlanjutan fiskal [1]. Dalam praktiknya, efisiensi anggaran biasanya diwujudkan melalui pemangkasan belanja pada sektor-sektor tertentu dan pengalokasian ulang anggaran berdasarkan prioritas pembangunan nasional. Tujuannya adalah menekan pemborosan, meningkatkan kualitas belanja publik, serta memperkuat disiplin fiskal negara [2]. Namun, implementasi kebijakan ini tidak jarang memunculkan pro dan kontra di tengah masyarakat. Sebagian mendukung karena dinilai sebagai langkah yang transparan dan bertanggung jawab, sedangkan sebagian lainnya mengkritik kebijakan tersebut karena dianggap berdampak negatif terhadap pelayanan publik, terutama pada sektor esensial seperti pendidikan, kesehatan, dan bantuan sosial [3].

Respon masyarakat terhadap kebijakan publik menjadi indikator penting dalam mengevaluasi efektivitas dan keberhasilan kebijakan tersebut. Tingkat penerimaan dan persepsi publik dipengaruhi oleh banyak faktor, mulai dari seberapa transparan pemerintah dalam menyampaikan tujuan kebijakan, sampai sejauh mana masyarakat merasakan

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **18/06/2025**; Accepted: **02/07/2025**; Published: **17/07/2025**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 6 No 3 Edisi Mei - Agustus 2025, Page 2182-2194
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

dampaknya secara langsung. Dalam hal ini, penting bagi pembuat kebijakan untuk memahami opini publik secara menyeluruh agar dapat mengambil keputusan yang lebih tepat dan akomodatif terhadap kebutuhan masyarakat. Salah satu pendekatan yang kini berkembang pesat untuk menggali opini publik adalah melalui pemanfaatan teknologi informasi, khususnya dengan memanfaatkan data dari media sosial [4].

Seiring meningkatnya penggunaan platform digital, media sosial seperti Twitter telah menjadi ruang ekspresi masyarakat yang sangat aktif dan terbuka [5]. Di platform ini, masyarakat bebas menyuarakan pandangannya terhadap isu-isu kebijakan yang sedang hangat, termasuk mengenai efisiensi anggaran. *Volume* data yang besar, keberagaman topik, serta sifatnya yang *real-time* menjadikan media sosial sebagai sumber data yang sangat kaya untuk dianalisis [6]. Namun, untuk mengolah data dalam jumlah besar dan tidak terstruktur ini, dibutuhkan pendekatan komputasional yang canggih, seperti *Text Mining* dan *Natural Language Processing* (NLP) [7]. *Text Mining* adalah proses mengekstrak informasi bermakna dari data teks, sedangkan NLP adalah cabang dari kecerdasan buatan yang memungkinkan mesin memahami, menginterpretasi, dan menghasilkan bahasa manusia [8], [9]. Salah satu penerapan utama dari kedua pendekatan ini adalah analisis sentimen, yaitu teknik untuk mengidentifikasi opini masyarakat dan mengklasifikasikannya ke dalam kategori sentimen tertentu seperti positif, negatif, atau netral [10] Analisis ini tidak hanya memberikan gambaran umum mengenai sikap publik terhadap suatu isu, tetapi juga dapat digunakan untuk mendeteksi kata-kata kunci yang paling sering digunakan serta mengungkap pola persepsi masyarakat secara menyeluruh.

Penelitian-penelitian sebelumnya telah membuktikan bahwa analisis sentimen berbasis *Text Mining* dan NLP efektif digunakan untuk mengkaji opini publik terhadap berbagai kebijakan. Handayani [11] dan Naraswati [12] menggunakan algoritma *Naïve Bayes* untuk mengelompokkan opini publik terhadap kebijakan penanganan pandemi COVID-19. Arsi dan Kusuma [13] mengadopsi metode Naive Bayes dalam menilai sentimen terhadap wacana pemindahan ibu kota negara. Di kancah internasional, Ali et al [14] menerapkan analisis sentimen dan topic modeling untuk menelusuri persepsi publik terhadap pemilihan presiden Tahun 2020 di Amerika Serikat. Herliana [15] bahkan membandingkan akurasi berbagai algoritma dalam menilai tanggapan terhadap kebijakan kuliah daring. Semua studi tersebut sama-sama menunjukkan bahwa pendekatan analisis sentimen dapat menjadi instrumen yang efektif dalam memahami opini masyarakat terhadap isu kebijakan publik. Berdasarkan kajian literatur yang ada, belum ditemukan studi yang secara khusus menganalisis sentimen masyarakat terhadap kebijakan efisiensi anggaran nasional di Indonesia, khususnya kebijakan tahun 2025 yang dirancang untuk mengatur ulang belanja negara di tengah tantangan fiskal global. Padahal, isu ini sangat relevan dan berdampak luas karena menyangkut pergeseran sumber daya dalam sektor publik. Celah inilah yang menjadi dasar dari penelitian ini. Dengan mengisi gap yang ada, penelitian ini bertujuan untuk memperkaya literatur dan menyediakan wawasan berbasis data mengenai persepsi publik terhadap efisiensi anggaran.

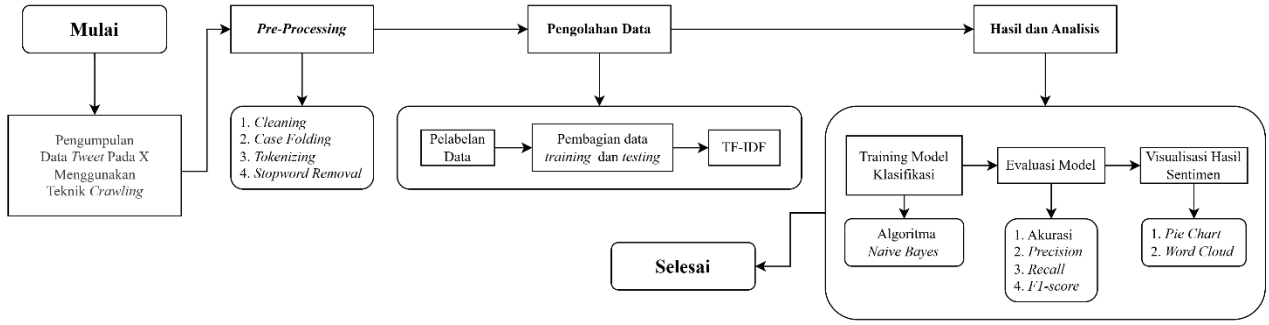
Berdasarkan pemaparan latar belakang tersebut maka, tujuan dari penelitian ini adalah untuk menerapkan metode *Text Mining* dan *Natural Language Processing* dalam menganalisis sentimen masyarakat terhadap kebijakan efisiensi anggaran tahun 2025. Secara khusus, penelitian ini akan mengumpulkan data opini publik dari Twitter yang berkaitan dengan isu efisiensi anggaran, melakukan pra-pemrosesan teks untuk membersihkan dan menstrukturkan data, mengklasifikasikan sentimen ke dalam kategori positif, negatif, dan netral, mengidentifikasi kata-kata kunci yang dominan dalam membentuk persepsi publik. Hasil dari penelitian ini diharapkan dapat menjadi masukan bagi pemerintah dalam menyusun strategi komunikasi kebijakan yang lebih efektif dan adaptif terhadap dinamika opini masyarakat, serta mendorong pengambilan kebijakan yang berbasis bukti (*evidence-based policy*) di era digital.

2. METODOLOGI PENELITIAN

Objek penelitian ini adalah opini masyarakat terkait kebijakan efisiensi anggaran 2025 yang diperoleh dari Twitter. Twitter dipilih karena merupakan media sosial yang aktif digunakan masyarakat untuk menyampaikan opini secara *real-time* dan cocok dianalisis dengan metode *text mining* dan NLP. Dalam penelitian ini, digunakan kombinasi alat dan bahan yang mendukung proses pengumpulan, pengolahan, dan analisis data secara efektif. Alat penelitian meliputi perangkat keras berupa laptop dengan spesifikasi minimum RAM 8 GB, prosesor Intel Core i5, dan penyimpanan SSD 256 GB, serta perangkat lunak seperti Visual Studio Code, RapidMiner Studio 9.10, Python 3.8, dan Google Chrome, didukung oleh berbagai library Python seperti pandas, logging, dan subprocess untuk keperluan crawling dan analisis data. Adapun bahan penelitian berupa data sekunder yang dikumpulkan dari Twitter menggunakan teknik *crawling* berbasis token autentikasi *cookie* tanpa API, dengan kata kunci seperti "efisiensi" dan "efisiensi anggaran." Data yang diperoleh mencakup 1.614 cuitan dalam rentang waktu Desember 2024 hingga April 2025, yang kemudian dianalisis menggunakan pendekatan *text mining* dan *Natural Language Processing* (NLP). Analisis ini bertujuan untuk menggambarkan sentimen publik terhadap kebijakan tersebut, sebagai masukan bagi pengambil kebijakan dalam mengevaluasi penerimaan masyarakat. Untuk memberikan gambaran yang lebih jelas mengenai tahapan-tahapan yang dilakukan dalam penelitian ini, berikut disajikan alur penelitian dalam bentuk kerangka pemikiran pada Gambar 1.

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **18/06/2025**; Accepted: **02/07/2025**; Published: **17/07/2025**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 6 No 3 Edisi Mei - Agustus 2025, Page 2182-2194
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin



Gambar 1. Kerangka Berfikir Penelitian

Berdasarkan Gambar 1, penelitian ini dimulai dari tahap pengumpulan data hingga tahap hasil dan analisis. Setiap tahapan dilaksanakan secara sistematis untuk memastikan bahwa data yang diperoleh valid, relevan dan siap untuk dianalisis secara optimal. Berikut ini merupakan penjelasan rinci dari alur penelitian yang telah disusun.

2.1 Pengumpulan Data

Data dalam penelitian ini dikumpulkan dari platform media sosial Twitter dengan menggunakan bahasa pemrograman *Python*. Proses pengumpulan dilakukan dengan teknik *crawling*, di mana token *cookies* Twitter diambil melalui inspeksi jaringan untuk memungkinkan akses data. Pencarian data dilakukan berdasarkan kata kunci yang telah ditentukan, seperti 'efisiensi' , ' efisiensi anggaran' dengan tambahan filter tanggal awal dan akhir dengan format since:YY-MM-DD until:YY-MM-DD. Data yang diperoleh kemudian disimpan dalam format CSV untuk keperluan pra-pemrosesan dan analisis lebih lanjut.

2.2 Tahap *Pre-Processing*

Tahap ini bertujuan untuk membersihkan dan menyiapkan data teks agar dapat diproses secara optimal oleh algoritma analisis sentimen. Adapun tahapan pre-processing yang dilakukan adalah sebagai berikut:
a. *Cleaning* (Pembersihan Data Teks): Membersihkan teks dari karakter-karakter yang tidak diperlukan, seperti: menghapus tanda baca (punctuation), menghapus angka, menghapus simbol atau karakter khusus (contoh: @, #, $, %), menghapus URL/link, menghapus emoji atau emotikon jika tidak dibutuhkan dan menghapus spasi berlebih. Proses ini penting untuk memastikan bahwa teks yang dianalisis hanya berisi informasi relevan.
b. *Case Folding*: Mengubah seluruh huruf dalam teks menjadi huruf kecil (lowercase) untuk menghindari perbedaan kata yang disebabkan oleh kapitalisasi.
c. *Tokenizing*: Memecah kalimat atau paragraf menjadi unit-unit kata atau token, sehingga dapat diproses lebih lanjut.
d. *Stopword Removal*: Menghapus kata-kata umum yang tidak memiliki makna penting dalam analisis, seperti "yang", "dan", "di", "ke", dan sebagainya.

2.3 Tahap Pengolahan Data

Setelah data mentah berhasil dikumpulkan, tahap selanjutnya adalah pengolahan data untuk mempersiapkan data agar dapat dianalisis secara optimal.

a. Pelabelan Data

Pada tahap ini, data yang telah dikumpulkan dari Twitter diberi label sentimen berdasarkan konteks isi cuitan. Setiap data teks diklasifikasikan ke dalam kategori positif, negatif atau netral sesuai dengan opini yang terkandung. Pelabelan ini dapat dilakukan secara manual atau semi-otomatis menggunakan algoritma dasar, kemudian diverifikasi kembali untuk memastikan akurasi. Proses pelabelan sangat penting untuk membangun model klasifikasi sentimen yang andal pada tahap analisis selanjutnya.

b. Pembagian Data Training dan Testing

Setelah proses pelabelan selesai, data dibagi menjadi dua subset, yaitu data *training* (70%) dan data *testing* (30%). Pembagian ini dilakukan secara *stratified*, yakni dengan menjaga proporsi kelas sentimen (positif, negatif, netral) tetap seimbang di kedua subset. Tujuannya adalah untuk memastikan bahwa model pembelajaran mesin mendapatkan representasi yang adil dari semua jenis sentimen selama pelatihan, serta dapat diuji secara objektif pada data baru yang belum pernah dikenali sebelumnya.

c. Ekstraksi Fitur dengan TF-IDF

Setelah data dilabeli, tahap berikutnya adalah mengubah teks menjadi representasi numerik yang dapat diproses oleh algoritma *machine learning*. Dalam penelitian ini, digunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk ekstraksi fitur. TF-IDF menghitung bobot setiap kata dalam dokumen berdasarkan frekuensi kemunculannya dan sejauh mana kata tersebut bersifat unik di antara semua dokumen. Dengan TF-IDF, teks mentah diubah menjadi

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **18/06/2025**; Accepted: **02/07/2025**; Published: **17/07/2025**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 6 No 3 Edisi Mei - Agustus 2025, Page 2182-2194
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin
matriks vektor angka yang merepresentasikan tingkat kepentingan setiap kata, sehingga memudahkan analisis sentimen berbasis mesin.

2.4 Tahap Hasil dan Analisis

Setelah data melalui tahap pra-pemrosesan dan pengolahan, selanjutnya dilakukan tahap analisis data untuk membangun model klasifikasi sentimen, mengevaluasi performa model, serta menyajikan hasil analisis dalam bentuk visual yang informatif. Tahapan ini terdiri dari beberapa langkah berikut:

a. Training Model Klasifikasi

Pada tahap ini, data yang telah diolah digunakan untuk melatih model klasifikasi sentimen. Penelitian ini menggunakan algoritma *machine learning Naive Bayes*, yaitu algoritma berbasis probabilistik yang mengasumsikan independensi antar fitur, cocok untuk klasifikasi teks karena sederhana dan efisien.

b. Evaluasi Model

Setelah model dilatih, dilakukan evaluasi performa untuk mengukur efektivitas masing-masing algoritma dalam mengklasifikasikan sentimen. Evaluasi dilakukan menggunakan empat metrik utama, yaitu:

1. *Accuracy*: Persentase prediksi yang benar terhadap seluruh data yang diuji.
2. *Precision*: Kemampuan model dalam memprediksi positif secara tepat dibandingkan jumlah prediksi positif yang dilakukan.
3. *Recall*: Kemampuan model dalam menemukan semua instance positif yang benar.
4. *F1-Score*: Rata-rata harmonis dari precision dan recall, memberikan gambaran keseimbangan antara keduanya.

Evaluasi ini bertujuan untuk menentukan algoritma mana yang menghasilkan performa terbaik dalam analisis sentimen.

2.5 Visualisasi Hasil Sentimen

Untuk memperjelas hasil analisis, dilakukan visualisasi data sentimen dengan dua jenis tampilan:

1. *Pie Chart*: Menampilkan proporsi distribusi sentimen (positif, negatif, netral) dalam bentuk diagram lingkaran untuk memudahkan interpretasi komposisi opini masyarakat.
2. *Word Cloud*: Menyajikan kata-kata yang paling sering muncul dalam cuitan, dengan ukuran font yang menunjukkan frekuensi relatif kata tersebut. Visualisasi ini membantu mengidentifikasi tema atau kata kunci dominan yang berkaitan dengan kebijakan efisiensi anggaran. Visualisasi ini tidak hanya mendukung penyajian hasil analisis secara informatif, tetapi juga memperkaya pemahaman terhadap sentimen publik secara umum.

3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan menganalisis sentimen masyarakat terhadap kebijakan efisiensi anggaran 2025 menggunakan pendekatan *text mining* dan NLP. Data diambil dari Twitter sebagai ruang publik digital yang aktif, memungkinkan analisis opini secara *real-time* dan objektif. Hasil analisis diharapkan memberi gambaran penerimaan masyarakat terhadap kebijakan tersebut dan menjadi masukan bagi pengambil kebijakan dalam merumuskan strategi yang lebih responsif.

3.1 Pengumpulan Data

Data dikumpulkan menggunakan teknik *web crawling* berbasis *Python* dengan kata kunci "efisiensi" dan "efisiensi anggaran" dalam rentang waktu tertentu, dengan menggunakan parameter since dan until. Pengambilan data dilakukan dari bulan Desember 2024 hingga bulan April 2025, dengan memilih satu tanggal secara acak pada setiap bulan sebagai sampel. Sebanyak 1.614 cuitan publik berhasil dikumpulkan dari Twitter, kemudian disimpan dalam format CSV untuk keperluan *pre-processing* dan analisis lebih lanjut.

Tabel 1. Sampel Data Mentah Hasil *Crawling* Twitter

No	Teks Tweet
1	Biasanya efisiensi anggaran akan memunculkan produktifitas. Kita liat saja apa sih yg bikin boros? Ato bakal muncul kreatifitas dari kepala dinas & pejabat2 ? Semoga di ekonomi skg ini @prabowo bisa langsung tindak tegas ASN TNI Polri yg msh berani Korupsi!
2	@pisang_gulai @RajaMandiri70 @yanuarnugroho Penghematan dengan alasan efisiensi anggaran padahal untuk menopang program jelek yang boros
3	gw lihat pemerintah udah efisiensi anggaran bahkan gw pun kena, gak ada lagi pembekalan internship di hotel dan nggak dikasih uang duduk. Bagus, tapi ngerasa anyep juga.

Tabel 1 merepresentasikan data mentah yang menjadi dasar dalam tahap-tahap *pre-processing* selanjutnya. Setiap cuitan mencerminkan opini masyarakat yang beragam, sehingga sangat relevan untuk dianalisis lebih lanjut menggunakan teknik klasifikasi sentimen.

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **18/06/2025**; Accepted: **02/07/2025**; Published: **17/07/2025**

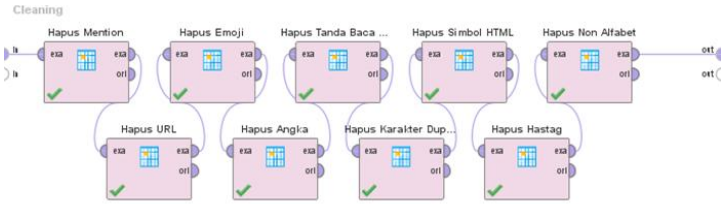
JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 6 No 3 Edisi Mei - Agustus 2025, Page 2182-2194
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

3.2 *Pre-Processing* Data

Pre-processing data merupakan tahap penting dalam pengolahan data teks, terutama pada analisis media sosial seperti Twitter. Tujuan utama *pre-processing* adalah untuk membersihkan teks dari unsur-unsur yang tidak relevan serta menyederhanakan teks agar dapat dianalisis lebih lanjut secara efektif, seperti dalam text mining, analisis sentimen, atau topik modeling. Berikut ini tahapan *pre-processing* yang dilakukan.

a. *Cleaning*

Cleaning merupakan tahap awal dalam *pre-processing* data teks yang bertujuan untuk membersihkan teks dari elemen-elemen yang tidak relevan atau tidak diperlukan dalam proses analisis. Dalam konteks data dari media sosial seperti Twitter, teks mentah cenderung mengandung berbagai noise seperti mention (@username), hyperlink, emoji, tanda baca berlebih, angka, hingga simbol atau karakter khusus lainnya. Elemen-elemen ini tidak membawa makna penting terhadap analisis sentimen maupun klasifikasi topik, sehingga perlu dibersihkan. Gambar 2 menunjukkan implementasi proses *cleaning* menggunakan *RapidMiner*.



Gambar 2. Proses *Cleaning* Menggunakan Rapid Miner

Pada gambar 2 beberapa operator digunakan secara berurutan dan setiap tahapan diatur secara sistematis untuk memastikan bahwa data yang dihasilkan bersih. Berikut Tabel 2 menampilkan perbandingan contoh data sebelum dan sesudah dilakukan *cleaning*.

Tabel 2. Contoh Data Sebelum dan Sesudah Dilakukan *Cleaning*

No	Sebelum *Cleaning*	Sesudah *Cleaning*
1	@kamentrader Yang harusnya dikenakan efisiensi...	Yang harusnya dikenakan efisiensi...
2	efisiensi 😤	efisiensi
3	astaghfirullahaladziim.....	astaghfirullahaladziim

Dengan proses *cleaning* yang menyeluruh, data teks menjadi lebih terstruktur dan siap untuk tahap berikutnya seperti *case folding, tokenization*, dan analisis yang lebih dalam. Tahap ini sangat krusial terutama untuk data dari media sosial yang sangat bebas dan tidak terstandar dalam penulisannya.

b. *Text Processing*

Text processing merupakan tahap penting dalam *pre-processing* data teks yang bertujuan untuk membersihkan dan menyiapkan data agar dapat dianalisis secara efektif oleh algoritma pembelajaran mesin. Dengan melakukan *text processing*, teks yang semula tidak terstruktur dapat diubah menjadi bentuk yang lebih seragam, terstandardisasi, dan representatif, sehingga meningkatkan akurasi dan efisiensi dalam proses ekstraksi fitur dan klasifikasi data. Gambar 3 menunjukan proses *text processing*.

Gambar 3. Proses *Text Processing*

1. *Case Folding*

Case folding adalah proses mengubah semua huruf dalam teks menjadi huruf kecil (*lowercase*). Tahapan ini sangat penting dalam *pre-processing* karena komputer memperlakukan huruf kapital dan huruf kecil sebagai karakter yang berbeda. Misalnya, kata "Efisiensi" dan "efisiensi" akan dianggap berbeda jika tidak dinormalisasi terlebih dahulu. Melalui *case folding*, kita memastikan bahwa analisis selanjutnya (seperti tokenisasi, pencocokan kata, dan pembobotan) dapat dilakukan secara konsisten. Ini terutama penting dalam bahasa Indonesia yang memiliki variasi penulisan dalam media sosial, termasuk kapitalisasi yang tidak standar. Tabel 3 menampilkan contoh hasil *case folding* pada beberapa tweet yang telah melalui tahap pembersihan (*cleaning*).

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **18/06/2025**; Accepted: **02/07/2025**; Published: **17/07/2025**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 6 No 3 Edisi Mei - Agustus 2025, Page 2182-2194
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Tabel 3. Contoh Data Sebelum dan Sesudah Dilakukan *Case Folding*

No	Sebelum *Case Folding*	Sesudah *Case Folding*
1	Yang Harusnya Dikenakan Efisiensi Itu Jabatan Giveaway	yang harusnya dikenakan efisiensi itu jabatan giveaway
2	Efisiensi Terbesar di KEMENPU Kemendikti Kemenkes yaAllah Ini Beneran	efisiensi terbesar di kemenpu kemendikti kemenkes yaallah ini beneran

Tabel di atas menunjukkan hasil dari proses *case folding*, di mana semua huruf dalam teks diubah menjadi huruf kecil (*lowercase*). Proses ini bertujuan untuk menyeragamkan format teks sehingga memudahkan dalam analisis lebih lanjut.

2. *Tokenizing*

Tokenizing adalah proses memecah teks menjadi bagian-bagian yang lebih kecil yang disebut token. Token biasanya berupa kata, frasa, atau karakter tergantung pada tujuan analisis. Dalam konteks pengolahan data teks dari media sosial seperti Twitter, tokenisasi dilakukan untuk memecah kalimat menjadi kata-kata individu agar memudahkan proses analisis selanjutnya seperti *stopword removal*. Tabel 4 menampilkan contoh hasil *tokenizing* dari data yang telah melalui tahap *cleaning* dan *case folding*.

Tabel 4. Contoh Data Sebelum dan Sesudah Dilakukan *Tokenizing*

No	Sebelum *Tokeninzing*	Sesudah *Tokenizing*
1	yang harusnya dikenakan efisiensi itu jabatan giveaway	[‘yang’, ‘harusnya’, ‘dikenakan’, ‘efisiensi’, ‘itu’, ‘jabatan’, ‘giveaway’]
2	efisiensi terbesar di kemenpu kemendikti kemenkes yaallah ini beneran	[‘efisiensi’, ‘terbesar’, ‘di’, ‘kemenpu’, ‘kemendikti’, ‘kemenkes’, ‘yaallah’, ‘ini’, ‘beneran’]

Tabel di atas menunjukkan bagaimana kalimat-kalimat utuh dipecah menjadi token-token individual. Tokenisasi seperti ini akan sangat membantu dalam proses analisis teks lanjutan, seperti identifikasi kata penting, *filtering stopword*, dan klasifikasi sentimen.

3. *Stopword Removal*

Stopword removal adalah proses menghapus kata-kata umum yang tidak memiliki makna penting dalam analisis teks, seperti “yang”, “dan”, atau “di”. Tahap ini penting untuk menyaring informasi relevan dan meningkatkan efisiensi analisis. Dalam teks berbahasa Indonesia, proses ini biasanya menggunakan pustaka seperti *Sastrawi* atau NLTK, dan membantu mengurangi dimensi data serta meningkatkan kinerja model analisis. Tabel 5 menampilkan contoh hasil *stopword removal* berdasarkan data yang telah melalui proses *cleaning, case folding,* dan *tokenizing* sebelumnya.

Tabel 5. Contoh Data Sebelum dan Sesudah Dilakukan *Stopword Removal*

No	Sebelum *Stopword Removal*	Sesudah *Stopword Removal*
1	kami sampaikan bahwa merujuk pada instruksi presiden ri nomor tahun tentang efisiensi belanja dalam pelaksanaan anggaran pendapatan ...	sampaikan merujuk instruksi presiden ri nomor tahun efisiensi belanja pelaksanaan anggaran pendapatan ...
2	punya anggaran jumbo dan gak kena efisiensi aja masih ada oknum seperti ini	punya anggaran jumbo gak kena efisiensi oknum

Tabel di atas memperlihatkan perbandingan antara teks sebelum dan sesudah proses *stopword removal*. Kata-kata yang tidak berkontribusi signifikan terhadap makna inti dihilangkan, sehingga hanya tersisa kata-kata utama yang lebih relevan untuk dianalisis lebih lanjut.

3.3 Pengolahan Data

Tahap pengolahan data mencakup pelabelan sentimen pada tweet, pembagian data menjadi data latih dan data uji, serta transformasi teks menjadi representasi numerik menggunakan TF-IDF. Tujuannya adalah menyiapkan data agar dapat diproses oleh algoritma pembelajaran mesin dalam pelatihan dan evaluasi model klasifikasi.

a. Pelabelan Data

Proses pelabelan dilakukan secara manual untuk mengklasifikasikan tweet ke dalam sentimen positif, negatif atau netral, guna memastikan akurasi interpretasi teks yang bersifat informal. Hasil pelabelan ini menjadi dasar bagi pelatihan dan pengujian model pembelajaran mesin. Tabel 6 menampilkan sampel data yang sudah dilabeli.

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **18/06/2025**; Accepted: **02/07/2025**; Published: **17/07/2025**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 6 No 3 Edisi Mei - Agustus 2025, Page 2182-2194
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Tabel 6. Contoh Data Hasil Pelabelan

No	Teks Tweet	Sentimen
1	Mantap! Perencanaan dan efisiensi anggaran penting buat kesuksesan program ini.	positif
2	Banyak cara efisiensi anggaran salah satunya kurangi jumlah kementerian yang ga esensial dan pos jabatan yang ga perlu berantas korupsi beneran yang bukan efisiensi itu mengebiri demokrasi	netral
3	Kenapa dpr ga kena efisiensi anggaran Kayak kerja ajaa upss	negatif

Dapat dilihat pada tabel 6 bahwa setiap tweet diklasifikasikan berdasarkan ekspresi sentimen yang terkandung di dalamnya. Proses pelabelan ini dilakukan secara manual guna memastikan akurasi data sebelum masuk ke tahap pemodelan

b. Pemisahan Data Training dan Testing

Setelah pelabelan, data sebanyak 1.614 tweet dipisahkan menjadi 1.130 data training dan 484 data testing. Pemisahan ini bertujuan untuk melatih model klasifikasi dengan data *training* dan menguji kemampuannya menggeneralisasi pada data baru melalui data *testing*. Tahapan ini penting untuk menghindari *overfitting* dan mengukur akurasi model secara objektif. Data dibagi dengan komposisi 70% training dan 30% testing secara *stratified*, agar proporsi kelas sentimen (positif, negatif, netral) seimbang. Tujuannya adalah memastikan representasi yang merata dalam pelatihan dan evaluasi model, sehingga hasil klasifikasi lebih akurat.

c. TF-IDF

Setelah pembagian data dilakukan, langkah selanjutnya adalah mengubah teks menjadi representasi numerik menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). Teknik ini bekerja dengan memberikan bobot yang lebih tinggi pada kata-kata yang bersifat unik dan jarang muncul di seluruh dokumen, tetapi sering muncul dalam satu dokumen tertentu, karena dianggap mewakili informasi penting. Dengan demikian, metode ini membantu mengurangi pengaruh kata-kata umum dan memungkinkan model pembelajaran mesin untuk lebih fokus pada kata-kata yang benar-benar relevan. Tabel 7 menampilkan hasil skor TF-IDF tertinggi.

Tabel 7. Kata dengan Skor TF-IDF Tertinggi

No	Kata	Jumlah Dokumen	Skor TF-IDF
1	efisiensi	1679	1775
2	anggaran	455	519
3	kena	300	323

Hasil TF-IDF menunjukkan kata-kata seperti "efisiensi", "anggaran", dan "negara" memiliki bobot tinggi, menandakan pentingnya kata-kata tersebut dalam membedakan konteks sentimen. Kata-kata ini menjadi fitur kunci untuk meningkatkan akurasi model klasifikasi.

3.4. Hasil dan Analisis

Tahap ini menganalisis performa model klasifikasi sentimen yang dibangun dengan *Naive Bayes* menggunakan data teks hasil *pre-processing*. Evaluasi dilakukan dengan *metrik accuracy, precision, recall,* dan *F1-score*. Hasil sentimen divisualisasikan lewat *pie chart* dan *word cloud* untuk memahami pola dan distribusi sentimen dalam data media sosial. Tahap ini penting untuk menilai keandalan model dan pola sentimen.

a. Training Model Klasifikasi

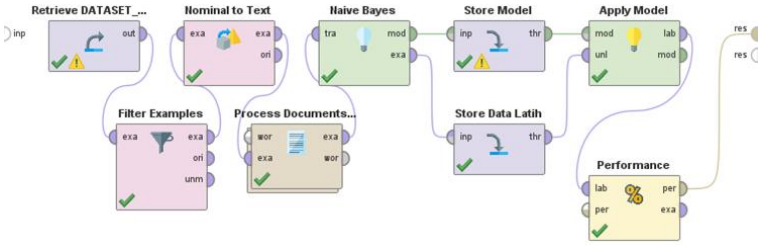
Penelitian ini menggunakan algoritma *Naive Bayes* yaitu algoritma *supervised learning* berbasis probabilitas yang efektif untuk klasifikasi teks. Pelatihan dilakukan pada data yang sudah diproses, dengan *Naive Bayes* menghitung probabilitas kelas berdasarkan distribusi kata. Algoritma ini cepat dan cocok untuk dataset besar seperti teks.

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **18/06/2025**; Accepted: **02/07/2025**; Published: **17/07/2025**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 6 No 3 Edisi Mei - Agustus 2025, Page 2182-2194
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

1. Penerapan Model

Pada tahap ini, model melakukan klasifikasi berdasarkan pola dan probabilitas yang telah dipelajari selama pelatihan. Proses ini sangat penting karena akan menentukan sejauh mana model dapat mengenali pola sentimen dari teks baru dan belum dikenal sebelumnya. Keakuratan model akan diuji dengan membandingkan hasil prediksi terhadap label aktual pada data uji yang telah disiapkan. Hasil dari penerapan ini akan menjadi dasar dalam mengevaluasi performa model pada tahap berikutnya. Gambar 4 menunjukkan proses lengkap penerapan algoritma *Naive Bayes* pada *RapidMiner*.

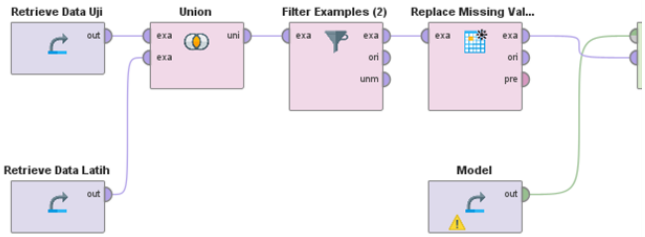


Gambar 4. Proses Penerapan Model Klasifikasi

Gambar 4 menggambarkan alur pelatihan model klasifikasi sentimen menggunakan algoritma *Naive Bayes* di *RapidMiner*. Proses dimulai dengan pengambilan dataset yang telah dipreproses, dilanjutkan dengan penyaringan data kosong menggunakan *Filter Examples*, dan konversi atribut nominal menjadi teks melalui *Nominal to Text*. Tahap utama dilakukan oleh *Process Documents from Data*, yang melibatkan *tokenisasi, case folding, stopword removal, dan transformasi* TF-IDF. Data yang telah diproses kemudian digunakan untuk membangun model *Naive Bayes*, yang hasilnya disimpan melalui *Store* Model*,* sementara dataset latih disimpan melalui *Store* Data Latih untuk evaluasi lanjutan.

2. Analisis Sentimen

Langkah selanjutnya adalah menerapkan model yang telah dilatih untuk melakukan prediksi terhadap data *testing*. Proses ini mencakup penggabungan data *training* dan data *testing*, penyesuaian atribut, serta penerapan model pada data yang telah dipersiapkan. Pada gambar 5 ditampilkan proses penerapan model analisis sentimen menggunakan algoritma *Naive Bayes* pada *RapidMiner*.



Gambar 5. Proses Analisis Sentimen

Gambar 5 menunjukkan alur proses analisis sentimen yang dimulai dari pengambilan data *training* dan data *testing* yang telah melalui tahap *pre-processing*. Data testing yang ditampilkan pada proses ini bukanlah data mentah, melainkan data yang sudah siap untuk digunakan, karena seluruh proses *text processing* (seperti *case folding, tokenizing* dan *stopword removal*) telah dilakukan sebelumnya. Hal ini dilakukan sebagai solusi atas keterbatasan pada versi terbaru *Rapid Miner*, di mana operator *Union* tidak dapat memproses data *training* dan data *testing* secara langsung jika tipe label berbeda. Masalah ini terjadi karena *RapidMiner* secara otomatis menyisipkan nilai 0 pada label kosong di data *testing* yang mengakibatkan label bertipe *real* dan menyebabkan *error* saat digabungkan dengan label bertipe *polynominal* dari data *training*.

Setelah kedua dataset digabung menggunakan operator *Union*, data kemudian difilter menggunakan *Filter Examples* untuk menyesuaikan atribut yang digunakan dalam pelatihan dan pengujian model yaitu atribut " *is missing* " atau artibut yang tidak berlabel. Selanjutnya, operator *Replace Missing Values* digunakan untuk menangani atribut yang mungkin masih memiliki nilai kosong. Terakhir, model yang sudah dilatih sebelumnya diterapkan pada data gabungan untuk melakukan prediksi sentimen terhadap data *testing*. Proses ini memungkinkan evaluasi performa model terhadap data baru secara efisien dan tanpa gangguan teknis dari perbedaan tipe data. Setelah model diterapkan pada data *testing*, sistem akan menghasilkan output berupa prediksi sentimen dari setiap entri data. Gambar 6 menunjukkan contoh hasil prediksi yang dihasilkan oleh model setelah melalui seluruh tahapan analisis.

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **18/06/2025**; Accepted: **02/07/2025**; Published: **17/07/2025**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 6 No 3 Edisi Mei - Agustus 2025, Page 2182-2194
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

sentimen	prediction(sentimen)	text
?	negatif	efisiensi uangnya hamburhamburkan korupsi munafik negara
?	netral	efisiensi efisiensi efisiensi efisiensi
?	netral	nungguin apbn efisiensi wung berubah ngitung spend infra beda ...
?	positif	efisiensi ahhaahahahaahaha
?	netral	mengecewakan karna efisiensi negara uang kena
?	positif	kena efisiensi
?	negatif	webnya kena efisiensi soalny apbn jeblok

Gambar 6. Contoh Hasil Prediksi Model Sentimen

Hasil prediksi yang ditampilkan pada gambar 6 menunjukkan label sentimen yang berhasil diprediksi oleh model *Naive Bayes* terhadap data *testing* yang telah diproses sebelumnya. Setiap baris merepresentasikan satu entri data dengan atribut asli beserta label hasil prediksi.

b. Evaluasi Model

Setelah proses pelatihan menggunakan algoritma *Naive Bayes*, langkah selanjutnya adalah mengevaluasi performa model terhadap data *training*. Evaluasi ini bertujuan untuk mengetahui seberapa baik model dalam mengklasifikasikan data ke dalam kategori sentimen yang benar yaitu positif, netral, dan negatif. Salah satu metode evaluasi yang umum digunakan adalah *confusion matrix*, yang menyajikan informasi mengenai jumlah prediksi benar dan salah untuk setiap kelas. Gambar 7 menampilkan hasil akurasi model *Naïve Bayes* dari model yang telah dilatih.

accuracy: 93.01%

	true netral	true positif	true negatif	class precision
pred. netral	187	0	47	79.91%
pred. positif	17	83	15	72.17%
pred. negatif	0	0	781	100.00%
class recall	91.67%	100.00%	92.65%	

Gambar 7. Hasil Akurasi Model *Naïve Bayes*

Evaluasi kinerja model dilakukan menggunakan data *training* dengan algoritma *Naive Bayes* dan menghasilkan tingkat akurasi sebesar 93,01%, yang menunjukkan performa klasifikasi yang sangat baik secara umum. Berdasarkan *confusion matrix*, model mampu mengenali kelas negatif dengan sangat baik, ditunjukkan oleh nilai *precision* 100%, artinya seluruh data yang diprediksi negatif memang benar-benar negatif, dan seluruh data negatif berhasil diklasifikasikan dengan benar. Untuk kelas positif, *precision* berada pada angka 72,17% dan *recall* mencapai 100%, menunjukkan bahwa model berhasil mendeteksi seluruh data positif, namun terdapat beberapa prediksi positif yang sebenarnya bukan positif (*false positive*). Sementara pada kelas netral, *precision*-nya cukup tinggi yaitu 79,91%, dan *recall*-nya 91,67%, menandakan model cukup baik dalam mengenali data netral, meskipun masih terdapat prediksi netral yang keliru terhadap kelas negatif.

Secara keseluruhan, model *Naive Bayes* menunjukkan performa yang sangat baik, khususnya dalam mendeteksi sentimen negatif, namun masih perlu peningkatan dalam membedakan data netral dan positif agar *precision* untuk kedua kelas tersebut meningkat. Sebagai pelengkap dari evaluasi yang telah dibahas, perhitungan metrik evaluasi seperti *accuracy, precision, recall,* dan *F1-score* juga dapat dilakukan secara manual berdasarkan *confusion matrix* yang dihasilkan. Hal ini bertujuan untuk memverifikasi hasil evaluasi otomatis serta memberikan pemahaman yang lebih mendalam mengenai performa model pada masing-masing kelas sentimen.

a. *Accuracy*

Akurasi dihitung sebagai rasio jumlah prediksi yang benar terhadap jumlah total prediksi:

$$Akurasi = \frac{187 + 83 + 781}{187 + 0 + 47 + 17 + 83 + 15 + 0 + 0 + 781} = \frac{1051}{1130} \approx 93.01\%$$

b. *Precision*

Presisi menunjukkan proporsi prediksi suatu kelas yang benar-benar termasuk dalam kelas tersebut.

1. Netral

$$Precision = \frac{TP}{TP + FP} = \frac{187}{187 + 17 + 0} = \frac{187}{204} \approx 91.67\%$$

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **18/06/2025**; Accepted: **02/07/2025**; Published: **17/07/2025**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 6 No 3 Edisi Mei - Agustus 2025, Page 2182-2194
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

2. Positif

$$Precision = \frac{TP}{TP + FP} = \frac{83}{83 + 0 + 0} = \frac{83}{83} \approx 100\%$$

3. Negatif

$$Precision = \frac{TP}{TP + FP} = \frac{781}{781 + 47 + 15} = \frac{781}{843} \approx 92.65\%$$

c. *Recall*

Recall (atau sensitivitas) menunjukkan seberapa banyak data yang sebenarnya termasuk dalam suatu kelas dapat dikenali dengan benar oleh model.

1. Netral

$$Recall = \frac{TP}{TP + FN} = \frac{187}{187 + 0 + 47} = \frac{187}{234} \approx 79.91\%$$

2. Positif

$$Recall = \frac{TP}{TP + FN} = \frac{83}{83 + 17 + 15} = \frac{83}{115} \approx 72.17\%$$

3. Negatif

$$Recall = \frac{TP}{TP + FN} = \frac{781}{781 + 0 + 0} = \frac{781}{781} \approx 100\%$$

d. *F1-Score*

F1-score adalah *harmonic mean* dari *precision* dan *recall*, memberikan ukuran keseimbangan antara keduanya.

1. Netral

$$F1 = 2\ x\ \frac{0.9167\ x\ 0.7991}{0.9167 + 0.7991} \approx 0.8547 \approx 85.47\%$$

2. Positif

$$F1 = 2\ x\ \frac{1.0\ x\ 0.7217}{1.0 + 0.7217} \approx 0.8378 \approx 83.78\%$$

3. Negatif

$$F1 = 2\ x\ \frac{0.9265\ x\ 1.0}{0.9265 + 1.0} \approx 0.9611 \approx 96.11\%$$

Dari hasil perhitungan manual berdasarkan *confusion matrix*, dapat disimpulkan bahwa model *Naive Bayes* memberikan performa yang sangat baik dengan akurasi mencapai 93.01%, serta nilai *precision* dan *recall* yang tinggi pada masing-masing kelas. Hal ini menunjukkan bahwa model mampu mengenali sentimen netral, positif, dan negatif dengan cukup akurat dan seimbang.

Hasil penelitian ini sejalan dengan penelitian sebelumnya oleh Isnain [16] yang juga menggunakan algoritma *Naive Bayes* dalam analisis sentimen publik terhadap kebijakan pemerintah, khususnya pada kebijakan *New Normal* selama pandemi COVID-19. Dalam penelitian tersebut, *Naive Bayes* berhasil mencapai tingkat akurasi sebesar 81% dalam mengklasifikasikan opini masyarakat berdasarkan data Twitter, dengan nilai precision 78%, recall 91% dan f1 -score 84%. Temuan ini menegaskan bahwa *Naive Bayes* merupakan algoritma yang andal untuk mengolah data teks tidak terstruktur dalam konteks opini publik di media sosial, khususnya pada isu-isu kebijakan pemerintah. Dengan demikian, model ini layak diterapkan untuk menganalisis respons masyarakat terhadap kebijakan efisiensi anggaran 2025 secara objektif dan berbasis data.

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **18/06/2025**; Accepted: **02/07/2025**; Published: **17/07/2025**

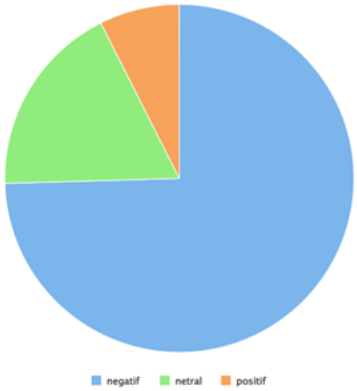
JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 6 No 3 Edisi Mei - Agustus 2025, Page 2182-2194
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

c. Visualisasi Hasil Sentimen

Setelah model dievaluasi dan dinyatakan memiliki performa yang cukup baik, langkah selanjutnya adalah visualisasi hasil sentimen. Visualisasi ini bertujuan untuk memberikan gambaran yang lebih intuitif dan mudah dipahami mengenai distribusi sentimen dalam data, serta kata-kata yang paling sering muncul pada setiap kategori sentimen.

1. *Pie Chart*

Dalam penelitian ini, visualisasi data sentimen menjadi bagian penting untuk memberikan gambaran yang lebih intuitif dan mudah dipahami mengenai persebaran opini masyarakat terhadap kebijakan efisiensi anggaran tahun 2025. Salah satu metode visualisasi yang digunakan adalah diagram *pie chart*, yang mampu menampilkan proporsi masing-masing kelas sentimen secara jelas. Visualisasi ini tidak hanya membantu dalam menganalisis kecenderungan sentimen, tetapi juga memperkuat hasil klasifikasi yang telah dilakukan sebelumnya. Gambar berikut menyajikan distribusi sentimen berdasarkan data yang telah dikumpulkan dari media sosial Twitter.



Gambar 8. *Pie Chart* Hasil Distribusi Sentimen

Gambar 8 menunjukkan visualisasi distribusi sentimen dalam bentuk *pie chart* terhadap opini masyarakat mengenai kebijakan efisiensi anggaran tahun 2025, berdasarkan data yang diambil dari media sosial Twitter. Dari total data yang telah dikumpulkan dan dianalisis, sentimen negatif mendominasi dengan proporsi sebesar 74,53%, yang mengindikasikan bahwa sebagian besar warganet menyampaikan kritik, keluhan atau ketidakpuasan terhadap kebijakan tersebut. Sentimen netral berada pada angka 18,02%, mencerminkan adanya tanggapan yang bersifat informatif atau tidak memihak. Sementara itu, sentimen positif hanya mencakup 7,43%, menunjukkan bahwa hanya sebagian kecil pengguna Twitter yang memberikan dukungan atau tanggapan optimis terhadap kebijakan efisiensi anggaran ini. Visualisasi ini memberikan gambaran awal tentang bagaimana persepsi publik di media sosial, khususnya Twitter, terhadap kebijakan yang sedang diberlakukan.

2. *World Cloud*

Untuk melengkapi analisis sentimen, digunakan visualisasi dalam bentuk *word cloud* guna mengidentifikasi kata-kata yang paling sering muncul dalam opini masyarakat terhadap kebijakan efisiensi anggaran tahun 2025. *Word cloud* memungkinkan peneliti untuk melihat secara langsung topik atau istilah yang dominan digunakan dalam percakapan di media sosial, khususnya Twitter. Visualisasi ini membantu memahami fokus perhatian dan kekhawatiran publik, serta memperkaya interpretasi dari klasifikasi sentimen yang telah dilakukan sebelumnya. Gambar 9 menunjukkan hasil *word cloud* dari data tweet yang telah diproses.

Gambar 9. *Word Cloud* Opini Publik terhadap Kebijakan Efisiensi Anggaran 2025

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **18/06/2025**; Accepted: **02/07/2025**; Published: **17/07/2025**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 6 No 3 Edisi Mei - Agustus 2025, Page 2182-2194
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Wordcloud pada gambar 9 menampilkan kata-kata yang paling sering muncul dalam tweet masyarakat terkait kebijakan efisiensi anggaran tahun 2025. Beberapa kata dominan seperti "efisiensi", "anggaran", "pemerintah", "negara", dan "presiden" mencerminkan fokus utama diskusi publik. Kata-kata seperti "rakyat", "gaji" dan "program" menunjukkan bahwa masyarakat banyak menyoroti dampak kebijakan ini terhadap kesejahteraan dan pelayanan publik. Munculnya nama "prabowo" serta kata-kata seperti "bikin", "makan", dan "kena" juga mengindikasikan keterkaitan opini publik dengan isu politik dan ekonomi sehari-hari. Visualisasi ini membantu menangkap tema-tema utama yang sedang hangat dibicarakan dan memberikan wawasan tambahan yang tidak selalu terlihat dari angka kuantitatif saja.

4. KESIMPULAN

Penelitian ini bertujuan untuk menjawab tantangan dalam memahami sentimen masyarakat terhadap kebijakan efisiensi anggaran tahun 2025 dengan memanfaatkan pendekatan *Text Mining* dan algoritma *Naive Bayes*. Hasilnya menunjukkan bahwa sebagian besar opini masyarakat di Twitter bersifat negatif (74,53%), yang mencerminkan adanya kekhawatiran atau ketidakpuasan publik terhadap kebijakan tersebut. Dengan akurasi mencapai 93,01%, model *Naive Bayes* terbukti efektif dalam mengklasifikasikan sentimen secara otomatis. Visualisasi data melalui *pie chart* dan *wordcloud* juga membantu memberikan gambaran yang lebih komprehensif mengenai kata kunci dan distribusi opini.

Selain itu, klaim efektivitas model tidak hanya didasarkan pada akurasi yang tinggi, tetapi juga ditunjang oleh distribusi data uji yang seimbang dan hasil evaluasi metrik lain seperti *precision* dan *recall* yang menunjukkan performa stabil di setiap kelas sentimen. Dengan demikian, simpulan bahwa pendekatan berbasis teknologi dapat menjadi alat yang andal untuk membantu pemerintah merespons dinamika opini publik bersifat valid dan berbasis pada data analisis yang memadai. Penelitian ini memberikan kontribusi penting dalam mendemonstrasikan bagaimana data sosial media dapat dimanfaatkan secara sistematis untuk menghasilkan wawasan berbasis bukti dalam proses pengambilan kebijakan publik.

Untuk penelitian selanjutnya, disarankan agar dilakukan perbandingan performa antara algoritma *Naive Bayes* dengan metode lain seperti SVM atau *Random Forest*, dengan mempertimbangkan berbagai metrik evaluasi seperti *presisi, recall*, dan efisiensi komputasi. Di sisi pengolahan data, perlu dilakukan filtrasi yang lebih ketat agar data lebih bersih dan relevan, misalnya dengan menghapus retweet, konten duplikat, serta tweet yang terlalu singkat. Selain itu, penggunaan metode representasi data yang lebih canggih seperti Word2Vec atau BERT dapat meningkatkan kualitas fitur. Penambahan sumber data dari platform lain di luar Twitter juga direkomendasikan agar hasil analisis sentimen lebih luas dan mencerminkan opini masyarakat secara lebih menyeluruh.

UCAPAN TERIMAKASIH

Penulis menyampaikan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan, bimbingan, dan kontribusi dalam penyusunan penelitian ini. Ucapan terima kasih secara khusus ditujukan kepada dosen pembimbing atas arahan dan masukan yang sangat berharga, serta kepada rekan-rekan yang turut membantu dalam proses pengumpulan data, analisis, dan penyusunan laporan. Tanpa dukungan dan kerja sama dari berbagai pihak, penelitian ini tidak akan dapat diselesaikan dengan baik.

REFERENCES

[1] Kementrian Sekretariat Negara, *Instruksi Presiden (Inpres) Nomor 1 Tahun 2025 tentang Efisiensi Belanja dalam Pelaksanaan Anggaran Pendapatan dan Belanja Negara dan Anggaran Pendapatan dan Belanja Daerah Tahun Anggaran 2025.* . https://peraturan.bpk.go.id/Details/313401/inpres-no-1-tahun-2025, 2025.

[2] S. Salman and M. Ikbal, "Analisis Efektivitas Kebijakan Efisiensi Anggaran: Ditinjau Dari Aspek Ekonomi," *Journal of Economics Development Research*, vol. 1, no. 2, pp. 68–72, 2025.

[3] R. A. Hafika, A. Hafiz, S. A. Waruwu, M. Y. Noor, Y. A. A. Sitohang, and K. Saputra, "EFISIENSI ANGGARAN DALAM WACANA PUBLIK: ANALISIS SENTIMEN PLATFORM X DENGAN NAÏVE BAYES," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 4, pp. 6093–6099, 2025.

[4] M. M. H. R. Pratama, "Comparison of Support Vector Machine (SVM) and Naïve Bayes Algorithm Performance in Analyzing Garuda Bird Design Sentiment in IKN," *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, vol. 8, no. 3Spc, pp. 1–8, 2025.

[5] F. Mas'ud, H. Jeluhur, K. Negat, A. Tefa, M. Uly, and M. Amtiran, "Etika Dalam Media Sosial Antara Kebebasan Ekspresi Dan Tanggung Jawab Digital," *Jimmi: Jurnal Ilmiah Mahasiswa Multidisiplin*, vol. 2, no. 2, pp. 235–246, 2025.

[6] M. Z. Maharani, "Analisis sentimen positif terhadap Avoskin sebagai Eco Friendly Brand di media sosial X dan TikTok," *Filosofi: Publikasi Ilmu Komunikasi, Desain, Seni Budaya*, vol. 1, no. 3, pp. 125–140, 2024.

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **18/06/2025**; Accepted: **02/07/2025**; Published: **17/07/2025**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 6 No 3 Edisi Mei - Agustus 2025, Page 2182-2194
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

[7] R. A. E. V. T. Sapanji, D. Hamdani, and P. Harahap, "Sentiment Analysis of the Top 5 E-commerce Platforms in Indonesia using Text Mining and Natural Language Processing (NLP)," *Journal of Applied Informatics and Computing*, vol. 7, no. 2, pp. 202–211, 2023.

[8] F. F. Mailoa, "Analisis sentimen data twitter menggunakan metode text mining tentang masalah obesitas di indonesia," *Journal of Information Systems for Public Health*, vol. 6, no. 1, pp. 44–51, 2021.

[9] F. Erlangga and I. P. Sari, "Perancangan Sistem Untuk Merekomendasikan Produk Skincare Menggunakan Metode NLP," *Portal Riset dan Inovasi Sistem Perangkat Lunak*, vol. 2, no. 4, pp. 1–11, 2024.

[10] M. Kholilullah, M. Martanto, and U. Hayati, "Analisis Sentimen Pengguna Twitter (X) Tentang Piala Dunia Usia 17 Menggunakan Metode Naive Bayes," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 392–398, 2024.

[11] E. T. Handayani and A. Sulistiyawati, "Analisis Sentimen Respon Masyarakat Terhadap Kabar Harian Covid-19 Pada Twitter Kementerian Kesehatan Dengan Metode Klasifikasi Naive Bayes," *J. Teknol. dan Sist. Inf*, vol. 2, no. 3, pp. 32–37, 2021.

[12] N. P. G. Naraswati, R. Nooraeni, D. C. Rosmilda, D. Desinta, F. Khairi, and R. Damaiyanti, "Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification," *Sistemasi: Jurnal Sistem Informasi*, vol. 10, no. 1, pp. 222–238, 2021.

[13] P. Arsi, B. A. Kusuma, and A. Nurhakim, "Analisis Sentimen Pindah Ibu Kota Berbasis Naive Bayes Classifier," *Jurnal Informatika Upgris*, vol. 7, no. 1, 2021.

[14] R. H. Ali, G. Pinto, E. Lawrie, and E. J. Linstead, "A large-scale sentiment analysis of tweets pertaining to the 2020 US presidential election," *J Big Data*, vol. 9, no. 1, p. 79, 2022.

[15] A. Herliana, "Analisis Sentimen Kuliah Daring Dengan Algoritma Naïve Bayes, K-Nn Dan Decision Tree," *Jurnal Responsif: Riset Sains Dan Informatika*, vol. 4, no. 1, pp. 70–80, 2022.

[16] A. R. Isnain, N. S. Marga, and D. Alita, "Sentiment analysis of government policy on corona case using naive bayes algorithm," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 1, pp. 55–64, 2021.

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **18/06/2025**; Accepted: **02/07/2025**; Published: **17/07/2025**