JUTIK : Jurnal Teknologi Informasi dan Komputer
Oktober-2025, Vol. 11, No.2, Hal.243-253
ISSN(P): 2442-241X; ISSN(E): 2528-5211

OPINI MASYARAKAT TERHADAP BONUS DEMOGRAFI PADA KANAL YOUTUBE DENGAN METODE TF-IDF, NAÏVE BAYES DAN SMOTE

Samuel Effendi Pratama[1*], Jolyn Lucretia[2], Hafiz Irsyad[3], Abdul Rahman[4]
Universitas Multi Data Palembang, Palembang, Sumatera Selatan, Indonesia[1]
Email*: samueleffendipratama_2226250024@mhs.mdp.ac.id
Universitas Multi Data Palembang, Palembang, Sumatera Selatan, Indonesia[2]
Email: jolynlucretia_2226250055@mhs.mdp.ac.id
Universitas Multi Data Palembang, Palembang, Sumatera Selatan, Indonesia[3]
Email: hafizirsyad@mdp.ac.id
Universitas Multi Data Palembang, Palembang, Sumatera Selatan, Indonesia[4]
Email: arahman@mdp.ac.id

(*) *Corresponding Author*

ABSTRAK

Penelitian ini mengkaji opini masyarakat terhadap isu bonus demografi yang diungkapkan melalui komentar pada kanal YouTube dengan menggunakan metode TF-IDF, Naïve Bayes, dan SMOTE. Data yang digunakan terdiri dari 870 komentar yang telah dilabeli secara manual menjadi sentimen positif dan negatif. Tahapan penelitian meliputi pra-pemrosesan data berupa *case folding*, penghapusan karakter non-alfabet, *stopword removal*, dan *stemming*, kemudian ekstraksi fitur menggunakan TF-IDF untuk mengubah teks menjadi representasi numerik yang dapat diproses oleh algoritma. Penelitian ini membandingkan performa model klasifikasi sentimen Naïve Bayes dalam dua skenario, yaitu tanpa dan dengan penerapan SMOTE. Teknik SMOTE digunakan untuk mengatasi ketidakseimbangan data antar kelas sentimen agar hasil klasifikasi lebih seimbang dan tidak bias. Hasil evaluasi menunjukkan bahwa model tanpa SMOTE menghasilkan akurasi sebesar 70% namun memiliki *recall* yang sangat rendah pada kelas positif. Setelah diterapkan SMOTE, akurasi meningkat menjadi 77%, dengan *precision* tertinggi sebesar 0,89 pada kelas negatif dan recall tertinggi sebesar 0,92 pada kelas positif. Visualisasi *word cloud* memperlihatkan kata-kata dominan yang mencerminkan pola opini masyarakat terkait bonus demografi secara jelas dan informatif. Hasil penelitian ini dapat memberikan gambaran kuantitatif terhadap persepsi publik serta menjadi bahan pertimbangan bagi pembuat kebijakan. Ke depan, metode ini dapat dikembangkan lebih lanjut dengan algoritma lain dan data dari berbagai platform media sosial untuk meningkatkan akurasi dan representativas analisis sentimen.
Kata kunci: bonus demografi, opini publik, Naïve Bayes, sentimen, TF-IDF

ABSTRACT

This study examines public opinion on the demographic bonus issue expressed through comments on YouTube channels using the TF-IDF, Naïve Bayes, and SMOTE methods. The data used consists of 870 comments that have been manually labeled into positive and negative sentiments. The research stages include data pre-processing in the form of case folding, removal of non-alphabetic characters, stopword removal, and stemming, then feature extraction using TF-IDF to convert text into numeric representations that

Submitted: 22 Mei 2025
Accepted: 15 September 2025
Published: 10 Oktober 2025

can be processed by the algorithm. This study compares the performance of the Naïve Bayes sentiment classification model in two scenarios, namely without and with the application of SMOTE. The SMOTE technique is used to overcome data imbalance between sentiment classes so that the classification results are more balanced and unbiased. The evaluation results show that the model without SMOTE produces an accuracy of 70% but has a very low recall in the positive class. After applying SMOTE, the accuracy increased to 77%, with the highest precision of 0.89 in the negative class and the highest recall of 0.92 in the positive class. The word cloud visualization shows the dominant words that reflect the pattern of public opinion regarding the demographic bonus clearly and informatively. The results of this study can provide a quantitative picture of public perception and be a consideration for policy makers. In the future, this method can be further developed with other algorithms and data from various social media platforms to improve the accuracy and representativeness of sentiment analysis.
Keywords: demographic bonus, Naïve Bayes, public opinion, sentiment, TF-IDF.

1. PENDAHULUAN

Bonus Demografi adalah suatu kondisi dimana jumlah penduduk yang usianya produktif (15-64 tahun) lebih banyak dibandingkan dengan penduduk yang usianya non-produktif, sehingga kondisi ini menciptakan peluang untuk mendorong pertumbuhan ekonomi secara signifikan [1]. Indonesia sendiri diperkirakan akan mencapai puncak bonus demografi pada tahun 2030 sampai 2040, menjadikan kondisi ini sangat penting pada perencanaan kebijakan dalam pembangunan jangka panjang dan menengah [2].

Seiring berjalannya perkembangan teknologi terutama media sosial, YouTube telah menjadi media interaktif dan informatif yang banyak dimanfaatkan untuk menyampaikan berbagai opini masyarakat [3], termasuk isu nasional strategis seperti bonus demografi. Komentar-komentar pada media sosial YouTube merepresentasikan beragam dan dinamisnya persepsi publik terhadap isu nasional tersebut [4]. Oleh karena itu, analisis opini publik berbasis data menjadi penting untuk mengidentifikasikan pola-pola persepsi masyarakat secara sistematis [5].

Dalam bidang analisis teks dan opini masyarakat, metode TF-IDF (*Term Frequency-Inverse Document Frequency*) terbukti efektif untuk mengubah teks menjadi representasi numerik berbobot yang menunjukan tingkatan urgensi sebuah kata dalam sebuah dokumen maupun keseluruhan korpus [6][7]. Sedangkan untuk klasifikasi opini, algoritma Naïve Bayes sering digunakan karena kecepatan pemrosesan dan tingkat akurasi yang tinggi dalam proses pengolahan data berupa teks [8].

Salah satu tantangan dalam analisis sentimen adalah ketidakseimbangan kelas, ketika opini publik cenderung didominasi oleh satu jenis sentimen, misalnya negatif, yang dapat menyebabkan model klasifikasi menjadi bias [9]. Untuk mengatasinya, teknik *Synthetic Minority Oversampling Technique* (SMOTE) digunakan guna menyeimbangkan distribusi data dengan membuat sampel sintetis dari kelas minoritas. Beberapa penelitian telah berhasil menggabungkan TF-IDF*,* Naïve Bayes*, dan* SMOTE dalam analisis data yang tidak seimbang dengan hasil yang baik [10], namun studi mengenai opini masyarakat terhadap isu bonus demografi khususnya di platform YouTube masih sangat terbatas.

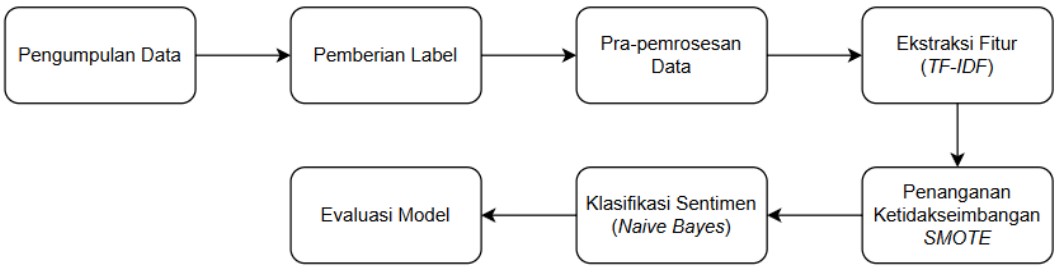
Penelitian ini bertujuan untuk mengklasifikasikan opini masyarakat mengenai isu bonus demografi dengan menggunakan data komentar dari kanal YouTube. Metode yang diterapkan meliputi TF-IDF sebagai teknik ekstraksi fitur, Naïve Bayes untuk proses klasifikasi, serta SMOTE untuk mengatasi masalah ketidakseimbangan data. Diharapkan hasil dari penelitian ini dapat memberikan gambaran kuantitatif mengenai persepsi

masyarakat sekaligus menjadi bahan pertimbangan bagi para pemangku kebijakan dalam merespons opini publik secara tepat.

2. METODE

Tahapan penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode pengolahan data berbasis pengklasifikasian teks untuk menganalisis opini masyarakat terkait isu bonus demografi. Data diperoleh melalui teknik text mining berupa pengumpulan komentar dari platform YouTube, kemudian dilakukan pelabelan sentimen secara manual menjadi dua kelas, yaitu positif dan negatif. Selanjutnya, dilakukan tahap pra-pemrosesan data yang meliputi normalisasi label, pembersihan teks menggunakan case folding, penghilangan karakter non-alfabet, penghapusan kata-kata tidak penting, serta proses stemming. Setelah teks dibersihkan, dilakukan ekstraksi fitur menggunakan metode TF-IDF untuk mengubah teks menjadi representasi numerik [11]. Untuk mengatasi masalah ketidakseimbangan jumlah data antar kelas, diterapkan teknik SMOTE pada data latih. Tahap akhir yaitu klasifikasi dilakukan menggunakan algoritma Naïve Bayes, yang dinilai memiliki kinerja baik untuk data teks [12]. Diagram alur proses dapat dilihat pada Gambar 1.



Gambar 1. Diagram alur penelitian

Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari 870 komentar yang dikumpulkan dari platform YouTube dengan topik yang berkaitan dengan isu bonus demografi. Komentar-komentar tersebut diperoleh melalui teknik *web scraping* menggunakan metode *text mining*. Proses pelabelan sentimen dilakukan secara manual oleh satu anotator, yaitu penulis sendiri. Penentuan label sentimen dilakukan dengan membaca setiap komentar dan menilai maknanya berdasarkan kata-kata yang memiliki konotasi positif atau negatif menurut Kamus Besar Bahasa Indonesia (KBBI). Tidak digunakan pedoman pelabelan formal atau anotator tambahan, namun untuk menjaga konsistensi, setiap komentar ditinjau ulang secara menyeluruh, dan penilaian diberikan secara hati-hati dengan merujuk pada arti kata menurut KBBI sebagai standar utama. Label sentimen diklasifikasikan ke dalam dua kategori, yaitu positif sebanyak 272 komentar dan negatif sebanyak 598 komentar. Dataset berlabel ini kemudian digunakan sebagai dasar dalam pelatihan dan pengujian model klasifikasi sentimen.

Pra-pemrosesan data

Pra-pemrosesan data (*preprocessing*) dilakukan untuk membersihkan dan menyiapkan data teks sebelum masuk ke tahap ekstraksi fitur dan klasifikasi. Tahapan ini penting untuk memastikan bahwa model hanya menerima informasi yang relevan dan

terstruktur. Langkah pertama dalam *preprocessing* adalah *case folding*, yaitu mengubah seluruh huruf dalam komentar menjadi huruf kecil agar konsistensi kata tetap terjaga. Selanjutnya dilakukan penghapusan karakter non-alfabet seperti angka, tanda baca, dan simbol khusus yang tidak memiliki nilai semantik dalam analisis sentimen [13]. Setelah itu, dilakukan *stopword removal* menggunakan daftar stopword Bahasa Indonesia dari pustaka Sastrawi untuk menghilangkan kata-kata umum yang tidak membawa makna penting, seperti "yang", "dan", "di", dan sebagainya. Proses berikutnya adalah *stemming*, yaitu mengubah kata menjadi bentuk dasarnya menggunakan algoritma stemming dari Sastrawi. Hasil akhir dari tahapan *preprocessing* ini adalah teks komentar yang lebih bersih dan siap untuk diekstraksi menjadi fitur numerik menggunakan metode TF-IDF.

Algoritma Naïve Bayes

Dalam penelitian ini, digunakan algoritma Naïve Bayes untuk melakukan klasifikasi sentimen terhadap komentar-komentar yang telah melewati proses pra-pemrosesan. Naïve Bayes merupakan metode klasifikasi berbasis probabilitas yang mengacu pada Teorema *Bayes*, dengan asumsi bahwa setiap kata atau fitur dalam data tidak saling bergantung satu sama lain [14]. Meskipun asumsi tersebut terbilang sederhana, algoritma ini dikenal cukup handal dan efisien dalam mengelola data teks yang bersifat jarang (*sparse*) dan memiliki dimensi tinggi, seperti yang umum dijumpai pada analisis sentimen.

Penelitian ini menggunakan varian *Multinomial* Naïve Bayes, yang sering diaplikasikan dalam klasifikasi teks karena mampu memanfaatkan frekuensi kemunculan kata dalam dokumen sebagai informasi penting [15]. Komentar-komentar yang telah melalui tahap pembersihan kemudian diubah ke dalam bentuk representasi numerik menggunakan metode TF-IDF, yang selanjutnya digunakan sebagai input untuk model klasifikasi. Sebelum proses pelatihan dimulai, diterapkan teknik SMOTE guna mengatasi ketidakseimbangan jumlah data antara kelas positif dan negatif. Teknik ini menghasilkan data sintetik untuk kelas minoritas sehingga distribusi data menjadi lebih merata. Dengan pendekatan ini, model Naïve Bayes diharapkan mampu mengenali pola sentimen dari kedua kelas secara lebih seimbang dan meningkatkan ketepatan prediksi.

Evaluasi

Evaluasi kinerja model dilakukan dengan menggunakan metrik akurasi dan *confusion matrix* untuk menilai seberapa baik model dalam mengklasifikasikan sentimen. Akurasi mengukur proporsi prediksi yang benar terhadap keseluruhan data uji, sementara *confusion matrix* memberikan informasi rinci mengenai jumlah prediksi yang tepat dan meleset pada masing-masing kelas. Untuk analisis yang lebih mendalam, digunakan pula metrik *precision*, *recall*, dan *F1-score*, yang sangat berguna terutama ketika menghadapi data yang tidak seimbang antar kelas.

Selain itu, visualisasi *word cloud* digunakan untuk memperlihatkan kata-kata yang paling sering muncul dalam komentar berlabel positif dan negatif. Hal ini membantu dalam mengidentifikasi pola atau kecenderungan kata yang mendominasi masing-masing sentimen [16]. Adapun akurasi didefinisikan sebagai rasio antara jumlah prediksi yang benar (baik kelas positif maupun negatif) terhadap total jumlah data, yang mencerminkan sejauh mana model dapat melakukan klasifikasi dengan tepat [17]. Nilai *accuracy* dapat dihitung dengan menggunakan persamaan(1).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision merupakan rasio yang menunjukkan seberapa banyak prediksi positif yang benar dibandingkan dengan seluruh prediksi yang diklasifikasikan sebagai positif. *Precision* menggambarkan tingkat ketepatan model dalam menghasilkan data yang sesuai dengan yang diharapkan [17]. Nilai *precision* dapat dihitung menggunakan persamaan(2).

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall adalah rasio antara jumlah prediksi positif yang benar dengan total data yang sebenarnya termasuk dalam kelas positif. Metrik ini menunjukkan sejauh mana model mampu menangkap atau mengenali seluruh data yang relevan sebagai positif [17]. Nilai *recall* dapat dihitung dengan menggunakan persamaan(3).

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F1 Score merupakan kombinasi harmonis antara nilai *precision* dan *recall* yang menghasilkan satu ukuran evaluasi [17]. F1 Score dapat dihitung dengan menggunakan persamaan(4).

$$F1\ Score = 2\ x \frac{Precision\ x\ Recall}{Precision+Recall} \quad (4)$$

3. HASIL DAN PEMBAHASAN

Deskripsi data

Pengujian dilakukan dengan mengukur tingkat akurasi dari hasil analisis sentimen terhadap opini masyarakat yang dianalisis menggunakan model yang telah dikembangkan. Berdasarkan hasil pengujian, dapat diketahui parameter mana yang memberikan akurasi terbaik. Penelitian ini menggunakan total 870 komentar yang diambil dari kanal YouTube. Semua komentar tersebut telah diberi label secara manual, dengan 598 komentar dikategorikan sebagai sentimen negatif dan 272 komentar sebagai sentimen positif.

Pra-pemrosesan data

Pra-pemrosesan dilakukan untuk merubah teks komentar yang masih mentah menjadi lebih terstruktur dan bersih, sehingga memudahkan proses ekstraksi fitur dan klasifikasi. Hasil dari tahapan pra-pemrosesan dapat dilihat pada Tabel 1.

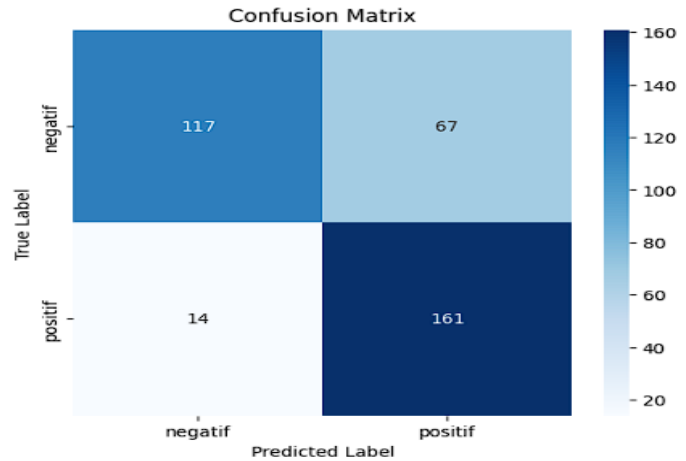
Tabel 1. Hasil Pra-pemrosesan

Komentar	*Cleaning*	Tokenisasi	Stop Removal	Stemming	Hasil
benar.. bonus demografi harus jadi berkat utk negara ini... lihat lah bagaimana negara2 lain yg KEKURANGAN PENDUDUK... berupaya	benar bonus demografi harus jadi berkat utk negara ini lihat lah bagaimana negara lain yg kekurangan penduduk berupaya mengajak penduduk	['benar', 'bonus', 'demografi', 'harus', 'jadi', 'berkat', 'utk', 'negara', 'ini', 'lihat', 'lah', 'bagaimana', 'negara', 'lain', 'yg', 'kekurangan', 'penduduk', 'berupaya', 'mengajak', 'penduduk', 'penduduk', 'berupaya', 'negara', 'mengajak', 'penduduk', 'penduduk',	['benar', 'bonus', 'demografi', 'berkat', 'negara', 'lihat', 'bagaimana', 'negara', 'kekurangan', 'penduduk', 'berupaya', 'mengajak', 'penduduk', 'penduduk', 'negara', 'datang', 'tinggal',	['benar', 'bonus', 'demograf i', 'berkat', 'negara', 'lihat', 'bagaiman a', 'negara', 'kurang', 'duduk', 'upaya', 'ajak', 'duduk', 'negara', 'negara',	benar bonus demograf i berkat negara lihat bagaiman a negara kurang duduk upaya ajak duduk negara datang

mengajak penduduk negara lain utk datang dan tinggal di negara mereka supaya negara mereka TIDAK PUNAH... DUKUNG MEMAJUKAN SUMBERDAYA GENERASI MUDA Ã°Å¸â€™Âª Ã°Å¸ÂÂ¼	negara lain utk datang dan tinggal di negara mereka supaya negara mereka tidak punah dukung memajukan sumberdaya generasi muda	'negara', 'lain', 'utk', 'datang', 'dan', 'tinggal', 'di', 'negara', 'mereka', 'supaya', 'negara', 'mereka', 'tidak', 'punah', 'dukung', 'memajukan', 'sumberdaya', 'generasi', 'muda']	'negara', 'negara', 'punah', 'dukung', 'memajukan', 'sumberdaya', 'generasi', 'muda']	'datang', 'tinggal', 'negara', 'negara', 'punah', 'dukung', 'maju', 'sumberdaya', 'generasi', 'muda']	tinggal negara negara punah dukung maju sumberdaya generasi muda

Model analisis sentimen

Dalam proses analisis sentimen, data teks terlebih dahulu melalui tahapan preprocessing seperti pembersihan, tokenisasi, *stopword removal*, dan *stemming*. Setelah itu, data teks diubah menjadi representasi numerik menggunakan metode TF-IDF untuk memudahkan pemrosesan oleh algoritma pembelajaran mesin. Selanjutnya, dilakukan dua pendekatan dalam pembangunan model klasifikasi. Pendekatan pertama menggunakan algoritma Naïve Bayes secara langsung pada data yang telah direpresentasikan dalam bentuk TF-IDF. Sementara pada pendekatan kedua, diterapkan teknik *oversampling* SMOTE terlebih dahulu untuk menangani ketidakseimbangan jumlah data antar kelas sebelum dilakukan pelatihan model dengan algoritma yang sama. Setelah pemodelan selesai, data dibagi menjadi dua bagian, yakni 70% untuk pelatihan dan 30% untuk pengujian, guna memastikan evaluasi performa model yang adil dan representatif.

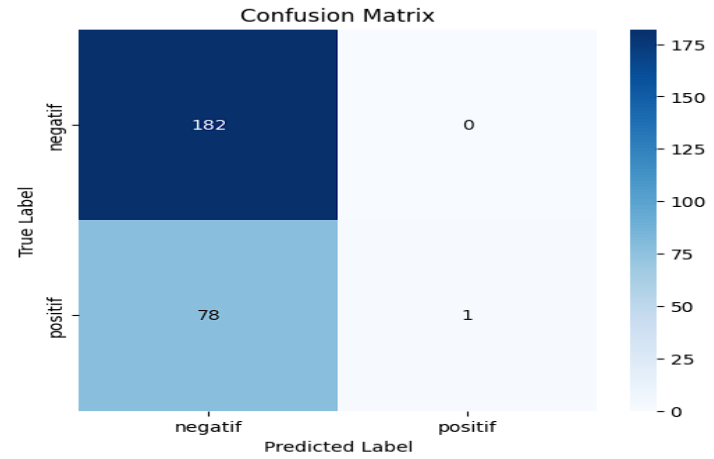


Classification Report:

	precision	recall	f1-score	support
negatif	0.89	0.64	0.74	184
positif	0.71	0.92	0.80	175
accuracy			0.77	359
macro avg	0.80	0.78	0.77	359
weighted avg	0.80	0.77	0.77	359

Gambar 2. Confusion Matriks Naïve Bayes dengan SMOTE

Gambar 2 menunjukkan Confusion Matriks dari model Naïve Bayes dengan penerapan SMOTE. Model berhasil mengklasifikasikan 117 data negatif dan 161 data positif dengan benar, meskipun masih terdapat kesalahan prediksi. Hasil evaluasi menunjukkan akurasi sebesar 77,44%, dengan *recall* yang tinggi pada kelas positif (0,92), menandakan model efektif dalam mengenali kelas minoritas. Penerapan SMOTE terbukti membantu meningkatkan keseimbangan performa model terhadap kedua kelas.



Classification Report:

	precision	recall	f1-score	support
negatif	0.70	1.00	0.82	182
positif	1.00	0.01	0.03	79
accuracy			0.70	261
macro avg	0.85	0.51	0.42	261
weighted avg	0.79	0.70	0.58	261

Gambar 3. Confusion Matriks Naïve Bayes tanpa SMOTE

Gambar 3 menunjukkan Confusion Matrix dari model Naïve Bayes tanpa penerapan SMOTE. Model sangat baik dalam mengenali kelas mayoritas (negatif),

dengan 182 prediksi benar dan tanpa kesalahan. Namun, performa pada kelas minoritas (positif) sangat rendah, dengan hanya 1 prediksi benar dari 79 data positif. Hal ini terlihat dari nilai *recall* kelas positif yang hanya 0,01 dan *f1-score* sebesar 0,03. Meskipun akurasi keseluruhan tampak tinggi (70%), nilai ini menyesatkan karena model gagal mengenali sebagian besar data positif. Tanpa SMOTE, ketidakseimbangan kelas menyebabkan model bias terhadap kelas mayoritas.

Evaluasi

Model Naïve Bayes dianalisis dalam dua kondisi, yaitu tanpa dan dengan penerapan metode SMOTE. Pada kondisi tanpa SMOTE, model menghasilkan akurasi sebesar 70%, namun angka ini tidak mencerminkan performa yang sesungguhnya karena model cenderung bias terhadap kelas mayoritas. Hal tersebut tercermin dari nilai *recall* yang sangat rendah pada kelas positif (0,01), yang mengindikasikan kegagalan model dalam mengenali sebagian besar data positif. Setelah dilakukan penyeimbangan data menggunakan SMOTE, kinerja model menunjukkan peningkatan yang cukup signifikan, terutama dalam mendeteksi kelas minoritas. Akurasi meningkat menjadi 77,44%, dan *recall* untuk kelas positif juga melonjak menjadi 0,92. Hasil ini menunjukkan bahwa penerapan SMOTE efektif dalam mengurangi bias terhadap kelas mayoritas dan meningkatkan keseimbangan klasifikasi antar kelas. Perbandingan dari dua pendekatan (Naïve Bayes tanpa SMOTE dan Naïve Bayes dengan SMOTE) dapat dilihat pada Tabel 2.

Tabel 2. Perbandingan Performa Model

Performa	Naïve Bayes	Naïve Bayes + SMOTE
Accuracy	70.11%	77.44%
Precision	0.85	0.80
Recall	0.51	0.78
F1-Score	0.42	0.77

Visualisasi

Untuk mendapatkan gambaran umum mengenai kata-kata yang paling sering muncul dalam komentar yang dianalisis, dilakukan visualisasi menggunakan *wordcloud*. Visualisasi ini membantu dalam mengidentifikasi tema atau topik yang dominan dalam kumpulan data, baik untuk sentimen positif maupun negatif. *Word Cloud* dibuat dengan menghitung frekuensi kemunculan kata-kata setelah melalui tahap pra-pemrosesan, seperti *case folding*, penghapusan karakter non-alfabet, penghilangan kata umum (*stopword removal*), serta proses *stemming*. Hasil visualisasi memperlihatkan bahwa kata-kata tertentu muncul dengan ukuran lebih besar, yang menandakan frekuensi kemunculannya lebih tinggi dibandingkan kata lain. Dengan demikian, *word cloud* dapat memberikan wawasan awal mengenai fokus pembicaraan atau opini masyarakat terkait isu bonus demografi sebelum dilakukan analisis yang lebih mendalam melalui metode klasifikasi sentimen. Hasil visualisasi ini dapat dilihat pada Gambar 4 untuk sentimen positif dan negatif.

Gambar 4. *Word cloud* sentimen positif dan negatif

KESIMPULAN DAN SARAN

Hasil evaluasi menunjukkan bahwa model tanpa SMOTE menghasilkan akurasi sebesar 70%, namun kinerjanya kurang optimal karena sangat bias terhadap kelas mayoritas, terlihat dari nilai *recall* kelas positif yang sangat rendah (0,01). Setelah diterapkan SMOTE, performa model meningkat secara signifikan, akurasi naik menjadi 77,44% dan *recall* pada kelas positif meningkat drastis menjadi 0,92. Ini menunjukkan bahwa SMOTE efektif dalam menangani ketidakseimbangan data dan meningkatkan kemampuan model dalam mengenali opini positif yang sebelumnya terabaikan. Selain itu, dominasi opini negatif dalam dataset diduga muncul karena isu bonus demografi sering diasosiasikan dengan kekhawatiran terhadap pengangguran dan tekanan ekonomi. Ke depan, disarankan pelabelan dilakukan oleh lebih dari satu anotator untuk meningkatkan objektivitas, serta pengumpulan data dari sumber dan waktu yang lebih beragam guna memperoleh representasi opini masyarakat yang lebih seimbang. Penelitian selanjutnya juga dapat mempertimbangkan penggunaan metode lain seperti *word embedding*, *transformer-based models*, atau *deep learning* untuk meningkatkan akurasi dan pemahaman konteks sentimen yang lebih kompleks.

4. DAFTAR PUSTAKA

[1] N. Satyahadewi, A. Amir, and E. Hendrianto, “Proyeksi Peningkatan Perekonomian melalui Pemanfaatan Bonus Demografi 2040,” *Kaganga:Jurnal Pendidikan Sejarah dan Riset Sosial Humaniora*, vol. 6, no. 2, pp. 715–725, Dec. 2023, doi: 10.31539/kaganga.v6i2.7943.

[2] D. Irma Aprianti and S. Choirudin, “Tantangan Bonus Demografi Bagi Pemerintah,” *Nusantara Innovation Journal*, vol. 1, no. 1,2022. doi: 10.70260/nij.v1i1.12.

[3] N. Nurrahmah, C. Zuriana, D. Iskandar, S. Sanusi, A. Armia, and M. Idham, "Opini Publik terhadap Debat Capres 2024: Analisis Sentimen dalam Komentar Live Youtube KPU RI," *Ranah: Jurnal Kajian Bahasa*, vol. 13, no. 2, p. 472, Dec. 2024, doi: 10.26499/rnh.v13i2.6063.

[4] B. Wicaksono and V. R. S. Nastiti, "Analisis Sentimen dalam Opini Publik di Chanel Youtube Indonesia Lawyers Club Tentang Isu Populer dengan Menggunakan Metode LSTM dan Bi-LSTM," *Jurnal Algoritma*, vol. 21, no. 2, pp. 241–251, Dec. 2024, doi: 10.33364/algoritma/v.21-2.1696.

[5] T. Wijaya, R. Indriati, and M. Muzaki, "Analisis Sentimen Opini Publik Tentang Undang-Undang Cipta Kerja Pada Twitter," *Jambura: Journal of Electrical and Electronics Engineering*, vol. 3, pp. 78–83, 2021, doi: 10.37905/jjeee.v3i2.10885.

[6] J. E. Br Sinulingga and H. C. K. Sitorus, "Analisis Sentimen Opini Masyarakat terhadap Film Horor Indonesia Menggunakan Metode SVM dan TF-IDF," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 14, no. 1, pp. 42–53, Feb. 2024, doi: 10.34010/jamika.v14i1.11946.

[7] R. Wahyu Pratama, P. A. Simanullang, P. T. Hutabarat, R. Dly, and A. Perdana, "Rancang Bangun Website Pengelolaan Buku Digital Berbasis Natural Language Processing (NLP)," *JuTIK: Jurnal Teknologi Informasi dan Komputer*, vol. 11, no. 1, pp. 102–113, 2025, doi: 10.36002/jutik.v11i1.3755.

[8] Y. Nurtikasari, Syariful Alam, and Teguh Iman Hermanto, "Analisis Sentimen Opini Masyarakat Terhadap Film Pada Platform Twitter Menggunakan Algoritma Naïve Bayes," *INSOLOGI: Jurnal Sains dan Teknologi*, vol. 1, no. 4, pp. 411–423, Aug. 2022, doi: 10.55123/insologi.v1i4.770.

[9] A. Syukron, E. Saputro, and P. Widodo, "Penerapan Metode Smote Untuk Mengatasi Ketidakseimbangan Kelas Pada Prediksi Gagal Jantung," *Jurnal Teknologi Informasi dan Terapan (J-TIT*, vol. 10, no. 1, pp. 2580–2291, 2023, doi: 10/25047/jtit.v10i1.312.

[10] N. Rahmah, P. Purnawansyah, and F. Umar, "Metode Support Vector Machine Untuk Klasifikasi Data Penyakit Hati Yang Imbalance," *Buletin Sistem Informasi dan Teknologi Islam*, vol. 5, no. 1, pp. 55–64, Apr. 2024, doi: 10.33096/busiti.v5i1.2189.

[11] F. D. Adhiatma and A. Qoiriah, "Penerapan Metode TF-IDF dan Deep Neural Network untuk Analisa Sentimen pada Data Ulasan Hotel," *Journal of Informatics and Computer Science*, 2022, doi: 10.26740/jinacs.v4n02.p183-193.

[12] E. Poerwandono and J. Perwitosari, "Penerapan Data Mining Untuk Penilaian Kinerja Karyawan Di PT. Riksa Dinar DJaya Menggunakan Metode Naïve Bayes Classification," *Jurnal Sains dan Teknologi*, vol. 5, no. 1, p. |pp, 2023, doi: 10.55338/saintek.v5i1.1416.

[13] A. E. Budiman and A. Widjaja, "Analisis Pengaruh Teks Preprocessing Terhadap Deteksi Plagiarisme Pada Dokumen Tugas Akhir," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 6, no. 3, Dec. 2020, doi: 10.28932/jutisi.v6i3.2892.

[14] A. B. Susanto, S. Wiharjo, and T. Liyana, "Analisis Data Minat Customer Produk Dengan Menggunakan Algoritma Naïve Bayes Classifier (Studi Kasus: PT Jellyfish Education Indonesia)," Scientia Sacra: Jurnal Sains, Teknologi dan Masyarakat. 2023. [Online]. Available: http://pijarpemikiran.com/index.php/Scientia

[15] A. Sentia, "Multinomial Naïve Bayes Classifier Untuk Analisis Sentimen Twitter," 2023. [Online]. Available: https://www.researchgate.net/publication/376730724

[16] B. F. S. Supriyanto and S. Rosalin, "Analisis Sentimen Program Merdeka Belajar dengan Text Analysis Wordcloud & Word Frequency," *Jurnal Minfo Polgan*, vol. 12, no. 1, pp. 25–32, Mar. 2023, doi: 10.33395/jmp.v12i1.12312.

[17] S. Clara, D. L. Prianto, R. A. Habsi, E. F. Lumbantobing, dan N. Chamidah, "Implementasi seleksi fitur pada algoritma klasifikasi machine learning untuk prediksi penghasilan pada Adult Income Dataset," dalam *Prosiding Seminar Nasional Teknologi dan Rekayasa (SENAMIKA)*, 2021. [Online]. Tersedia: https://conference.upnvj.ac.id/index.php/senamika/article/view/1417