Klasifikasi Sentimen Publik di *X* Menggunakan *Multinomial Naive Bayes*

Yanuar Pramana Samsudin *, Abdul Kudus

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

pramanayanuar30@gmail.com, abdul.kudus@unisba.ac.id

Abstract. This study aims to classify public sentiment towards an issue using the Multinomial Naive Bayes method. The data used are Indonesian-language tweets that have undergone preprocessing stages, such as case folding, punctuation removal, stopword removal, and stemming. Feature representation is performed using CountVectorizer to convert the text into numeric form. The data is then divided into training data and test data with a ratio of 80:20. To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) method is used. Performance evaluation is carried out by calculating accuracy, precision, recall, f1-score, and constructing a confusion matrix. The results show that the model without SMOTE produces an accuracy of 63%, with performance tending to be biased towards the majority class. After applying SMOTE, the accuracy increases to 65%, and the distribution of predictions between classes becomes more balanced. These findings indicate that Multinomial Naive Bayes is quite effective for short text classification tasks, and the application of SMOTE contributes to improving performance on imbalanced data. This research can be a basis for the development of more accurate and applicable sentiment analysis systems in the future.

Keywords: *Classification, Demographic Bonus, Multinomial Naive Bayes*.

Abstrak. Penelitian ini bertujuan untuk mengklasifikasikan sentimen publik terhadap suatu isu menggunakan metode *Multinomial Naive Bayes*. Data yang digunakan berupa tweet berbahasa Indonesia yang telah melalui tahap pre-processing, seperti *case folding*, penghapusan tanda baca, *stopwords removal*, dan *stemming*. Representasi fitur dilakukan menggunakan *CountVectorizer* untuk mengubah teks menjadi bentuk numerik. Data kemudian dibagi menjadi data latih dan data uji dengan perbandingan 80:20. Untuk menangani ketidakseimbangan kelas, digunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE). Evaluasi performa dilakukan dengan menghitung akurasi, *precision*, *recall*, *f1-score*, serta menyusun *confusion matrix*. Hasil penelitian menunjukkan bahwa model tanpa SMOTE menghasilkan akurasi sebesar 63%, dengan kinerja yang cenderung bias terhadap kelas mayoritas. Setelah penerapan SMOTE, akurasi meningkat menjadi 65%, dan distribusi prediksi antar kelas menjadi lebih seimbang. Temuan ini menunjukkan bahwa *Multinomial Naive Bayes* cukup efektif untuk tugas klasifikasi teks pendek, dan penerapan SMOTE berkontribusi dalam meningkatkan performa pada data tidak seimbang. Penelitian ini dapat menjadi dasar bagi pengembangan sistem analisis sentimen yang lebih akurat dan aplikatif di masa depan.

Kata Kunci: *Klasifikasi, Bonus Demografi, Multinomial Naïve Bayes.*

*pramanayanuar30@gmail.com

A. Pendahuluan

Bonus demografi terjadi ketika proporsi penduduk usia produktif lebih besar dibandingkan usia non-produktif dan menjadi peluang strategis bagi pertumbuhan ekonomi. Di Indonesia, fenomena ini diperkirakan mencapai puncaknya pada 2020–2035 (Retno Widiarti et al., 2025). Namun, muncul kekhawatiran terkait kesiapan sumber daya manusia, ketimpangan pendidikan, keterbatasan lapangan kerja, dan kebijakan pemerintah yang belum optimal. Isu ini banyak dibahas di media sosial, khususnya *X*, dengan peran signifikan dari Generasi Z dalam menyuarakan opini (Rafli Tanjung et al., 2024).

Penelitian ini bertujuan menganalisis persepsi publik terhadap bonus demografi melalui analisis sentimen tweet menggunakan algoritma *Multinomial Naive Bayes* (MNB) (Sabrani et al., n.d.). Algoritma ini dipilih karena efisien, sederhana, dan cocok untuk data berukuran kecil namun tetap memberikan hasil yang akurat (anakblogger, 2025). Analisis sentimen diharapkan dapat mengungkap kecenderungan opini publik (positif, negatif, netral) dan isu yang menjadi perhatian, sehingga hasilnya dapat menjadi masukan dalam perumusan kebijakan serta memahami dinamika diskusi masyarakat di era digital.

Berdasarkan permasalahan yang terdapat pada latar belakang, rumusan masalah dalam penelitian ini adalah bagaimana membuat label pada cuitan *X* sebagai sentimen positif, netral, atau negatif, bagaimana persepsi publik terhadap isu bonus demografi di Indonesia berdasarkan analisis sentimen pada *X*, serta bagaimana performa algoritma Multinomial Naive Bayes dalam mengklasifikasikan sentimen pada tweet terkait bonus demografi. Penulis membatasi permasalahan dalam penelitian ini pada penggunaan metode klasifikasi *Multinomial Naïve Bayes* serta pemanfaatan data primer hasil *scraping* dari *X* dengan kata kunci "Bonus Demografi" yang diambil pada periode 23 April 2025 hingga 6 Mei 2025.

B. Metode

Proses penelitian ini dilakukan secara sistematis, dimulai dari pengambilan data hingga penyajian hasil. Data dikumpulkan dari *X* menggunakan kata kunci "Bonus Demografi" dengan bantuan tools *Tweet Harvest*, mencakup teks *tweet*, waktu, *username*, dan metadata lainnya. Selanjutnya, *tweet* yang telah dibersihkan diberi label sentimen (positif, negatif, netral) secara manual oleh dua anotator, kemudian divalidasi menggunakan model IndoBERTweet untuk mengurangi subjektivitas, dengan penentuan label akhir melalui mekanisme voting. Data kemudian melalui tahap pra-pemrosesan yang meliputi pembersihan teks (menghapus URL, tanda baca, angka, dan simbol), *case folding* (mengubah seluruh teks menjadi huruf kecil) (Yuyun et al., 2021), penghapusan *stopwords* (Handayani et al., n.d.), *stemming* untuk mengubah kata menjadi bentuk dasar, serta tokenisasi (Saputra et al., 2015). Setelah itu dibuat tabel frekuensi berdasarkan label sentimen, kemudian dataset dibagi menjadi data latih dan data uji dengan rasio 80:20 (Azzahra Nasution et al., 2019). Fitur teks diekstraksi menggunakan *CountVectorizer* sehingga setiap kata direpresentasikan dalam bentuk numerik. Model klasifikasi dilatih menggunakan algoritma Multinomial Naïve Bayes dengan *library scikit-learn*, di mana model mempelajari distribusi kata pada tiap kelas. Tahap selanjutnya adalah pengujian model menggunakan data uji untuk melihat kemampuan klasifikasi sentimen, kemudian mengevaluasi kinerjanya dengan menghitung metrik akurasi, presisi, *recall*, dan *F1-Score*.

Tabel 1. Operasional Variabel

No.	Variabel	Keterangan
1.	X	*Tweet*
2.	Y	Label (Sentimen)

Tabel 2. Data Hasil *Data Cleaning*

No.	*full_text*
1	Kejeniusan beliau dalam menyeting…
2	Wakil Presiden Gibran Rakabuming Raka ...
3	Salah satu bonus demografi
4	udah ditahap pemerintah gabisa mengatasi lonjakan …
5	Dan para pengkhianat memanfaatkan generation gap ...
:	:
1012	Asmara1701 Definisi Indonesia Besar Indonesia …
1013	Ih dia suka bagi bagi susu dan skincare gak masuk …
1014	Nihh tweet utk anda pahami Anda bicara tentang …
1015	ya kan budaya kita bonus demografi jadi di sekolah …
1016	menarik analisa amp sudut pandang pak Anies …

Sumber: Data dari hasil *X Data Scraping*, 2025.

C. Hasil Penelitian dan Pembahasan

Sebelum melalui tahap *pre-processing*, data *tweet* diberi label sentimen untuk mengidentifikasi sikap atau opini publik terkait isu bonus demografi di Indonesia. Pelabelan dilakukan secara manual agar hasilnya akurat, dengan tiga kategori utama, yaitu positif (dukungan atau pandangan optimis), negatif (kritik atau pandangan pesimis), dan netral (informasi tanpa ekspresi sentimen tertentu). Sebanyak 1.016 tweet dianalisis secara kontekstual untuk menangkap nuansa bahasa informal, sarkasme, dan ambiguitas yang sering muncul di media sosial. Hasil pelabelan ini digunakan sebagai variabel target pada pelatihan model klasifikasi sentimen dengan algoritma *Multinomial Naïve Bayes*, di mana kualitas pelabelan menjadi faktor penting untuk meningkatkan performa dan akurasi model.

Tabel 3. Hasil Analisis Sentimen Final dari 2 Anotator Manual dan IndoBERTweet

No.	*Tweet*	Sentimen final
1	Kejeniusan beliau dalam menyeting hingga insyaAllah Indonesia mencapai Indonesia Emas 2024 plus bonus demografi ada dalam pidato beliau Mengukur tingkat urbanisasi dan pemecahannya memberikan sertifikat kepada rakyat Semacam transmigrasi era kini lah	*positive*
2	Wakil Presiden Gibran Rakabuming Raka menyampaikan melalui video monolog yang diunggah melalui kanal YouTube pribadinya bahwa bonus demografi diprediksi akan terjadi pada tahun 2030 2045 di Indonesia	*neutral*
3	Salah satu bonus demografi	*positive*
4	udah ditahap pemerintah gabisa mengatasi lonjakan bonus demografi	*negative*

No.	*Tweet*	Sentimen final
5	Dan para pengkhianat memanfaatkan generation gap untuk jadi keungtungan mereka mendapatkan kapital Alih alih bonus demografi tapi gap demografi lah yang menjadi pemecahbelah bangsa ini dikomandoi oleh tetua tetua wahn Yang portonya selalumemilih kabur saat medan perang terbuka	*negative*
:		
1016	menarik analisa amp sudut pandang pak Anies dlm menyikapi amp memaknai bonus demografi sy yakin jika ngerti konsep amp paham mk potensi yg mungkin timbul paling tdk tau cara utk meminimalkan resiko sehingga apa yg kita sebut berkah demografi itu memang bisa diwujudkan	*positive*

Sumber: Data Penelitian yang Sudah Diolah, 2025.

Setelah melalui tahap preprocessing, data tweet diproses melalui serangkaian langkah yang dirancang untuk meningkatkan kualitas dan konsistensinya sebelum dilakukan analisis lanjutan. Proses ini diawali dengan pembersihan data untuk menghapus elemen yang tidak relevan seperti tanda baca, angka, simbol, dan karakter khusus yang dapat mengganggu analisis. Selain itu, data duplikasi yang dapat menyebabkan bias pada hasil analisis juga dihapus agar dataset menjadi unik dan representatif. Selanjutnya dilakukan *case folding* untuk menyamakan semua huruf menjadi format kecil (lowercase), sehingga kata yang sama dengan perbedaan huruf kapital tidak dianggap sebagai entitas yang berbeda. Tahap berikutnya adalah *stopwords removal*, yaitu penghapusan kata-kata umum yang tidak memiliki nilai informatif signifikan, seperti “dan”, “yang”, atau “atau”. Kemudian dilakukan *stemming* untuk mengubah kata menjadi bentuk dasarnya, sehingga variasi kata dengan makna serupa dapat disederhanakan ke satu bentuk yang sama. Langkah terakhir adalah *tokenisasi* yang memecah kalimat menjadi potongan kata-kata, sehingga mempermudah proses analisis berbasis teks.

Hasil dari keseluruhan tahap *preprocessing* ini adalah data yang bersih, terstruktur, dan siap digunakan untuk tahapan analisis selanjutnya. Dengan teks yang sudah distandardisasi dan berfokus pada kata-kata relevan, potensi gangguan dari kata yang tidak informatif dapat diminimalkan, sehingga kualitas fitur yang dihasilkan menjadi lebih baik. Hal ini sangat penting karena hasil preprocessing yang optimal akan berkontribusi langsung pada akurasi pelabelan data serta kinerja model analisis sentimen.

Tabel 4. Hasil Proses *Preprocessing Data*

No.	*Tweet*	Sentimen final
1	['jenius', 'beliau', 'atur', 'insyaallah', 'indonesia', 'capai', 'indonesia', 'emas', 'plus', 'pidato', 'beliau', 'ukur', 'tingkat', 'urbanisasi', 'pecah', 'sertifikat', 'rakyat', 'transmigrasi', 'era']	*positive*
2	['wakil', 'presiden', 'gibran', 'rakabuming', 'raka', 'video', 'monolog', 'unggah', 'kanal', 'youtube', 'pribadi', 'prediksi', 'indonesia']	*neutral*
3	['salah']	*positive*
4	['tahap', 'perintah', 'gabisa', 'atas', 'lonjak']	*negative*
5	['khianat', 'manfaat', 'generation', 'gap', 'keungtungan', 'kapital', 'alih', 'alih', 'gap', 'pemecahbelah', 'bangsa', 'komando', 'tetua', 'tetua', 'wahn', 'porto', 'selalumemilih', 'kabur', 'medan', 'perang', 'buka']	*negative*
:	:	
953	['tarik', 'analisa', 'sudut', 'pandang', 'anies', 'sikap', 'makna', 'ngerti', 'konsep', 'paham', 'mk', 'potensi', 'timbul', 'tau', 'minimal', 'resiko', 'berkah', 'wujud']	*positive*

Sumber: Data Penelitian yang Sudah Diolah, 2025.

Setelah data *tweet* diberi label sentimen dan melalui tahap pra-pemrosesan, dibuat tabel frekuensi untuk mengetahui jumlah *tweet* pada setiap kategori sentimen (positif, negatif, netral) menggunakan fungsi *value_counts()* dari pustaka *pandas*. Tabel ini memberikan gambaran awal persebaran opini publik terhadap isu bonus demografi serta menjadi dasar pembagian data latih dan uji, sekaligus membantu mengidentifikasi apakah data bersifat seimbang atau tidak yang dapat memengaruhi performa model klasifikasi.

Tabel 5. Frekuensi *Tweet* per Kelas

Sentimen	Jumlah *Tweet*
Positif	104
Netral	274
Negatif	575
Total	953

Sumber: Data Penelitian yang Sudah Diolah, 2025.

Setelah ekstraksi fitur menggunakan *CountVectorizer* selesai, langkah berikutnya adalah melatih model klasifikasi dengan metode *Multinomial Naïve Bayes*, yang dikenal efektif untuk data teks yang direpresentasikan dalam bentuk frekuensi kata. Algoritma ini menggunakan prinsip probabilistik dari teorema Bayes dengan asumsi bahwa setiap kata dianggap independen satu sama lain. Model mempelajari distribusi probabilitas setiap kata terhadap kelas sentimen (positif, netral, negatif) serta menghitung probabilitas prior untuk masing-masing kelas berdasarkan data pelatihan.

Langkah pertama dari perhitungan Multinomial Naïve Bayes adalah menghitung prior kelas, dimana rumusnya adalah sebagai berikut:

$$\text{P(Kelas)} = \frac{\text{jumlah dokumen dalam kelas}}{\text{total jumlah dokumen}} \quad \ldots (1)$$

Sebagai contoh, pada data latih terdapat 83 dokumen kelas positif dari total 762 dokumen, maka prior kelas positif adalah:

$$\text{P(Positif)} = \frac{83}{762} = 0{,}1089238845$$

Langkah selanjutnya adalah menghitung probabilitas tiap kata dalam setiap kelas. Probabilitas ini dihitung berdasarkan jumlah kemunculan kata dalam masing-masing kelas ditambah dengan konstanta satu untuk menghindari bias lalu dibagi dengan total kata dalam kelas tersebut dan seluruh kata dalam seluruh dokumen. Rumus dari langkah ini adalah sebagai berikut:

$$\mathrm{P}(\mathrm{w_i}|\mathrm{kelas}) = \frac{\text{jumlah kemunculan } \mathrm{w_i} \text{ pada kelas} + 1}{\text{total kata pada kelas} + |\mathrm{V}|} \quad \ldots (2)$$

Sebagai contoh, jika kata "produktif" muncul sebanyak 8 kali dalam dokumen kelas Positif, dan jumlah seluruh kata dalam kelas Positif adalah 1033 dan seluruh kata dari seluruh dokumen adalah 8435, maka:

$$\mathrm{P}(\mathrm{Produktif}|\mathrm{Positif}) = \frac{8+1}{1033+8435} = 0{,}0009505703422$$

Karena Model Naïve Bayes menggunakan perkalian terhadap banyak probabilitas yang sangat kecil, maka dilakukan transformasi ke logaritma untuk menghindari nilai terlalu kecil (underflow) dan menyederhanakan perhitungan (dari perkalian jadi penjumlahan):

$$\log_{10}(\mathrm{P}(\mathrm{w_i}|\mathrm{kelas})) = \log_{10}(0{,}0009505703422) = -3{,}02201574$$

Jika kata "produktif" muncul sebanyak 2 kali dalam dokumen uji, maka kontribusi terhadap skor kelas adalah:

$$2 \times (-3{,}02201574) = -6{,}04403148$$

Setelah mendapatkan nilai log dari masing-masing kata, total skor kelas dihitung dengan menjumlahkan log prior kelas dan seluruh kontribusi kata-kata dalam dokumen:

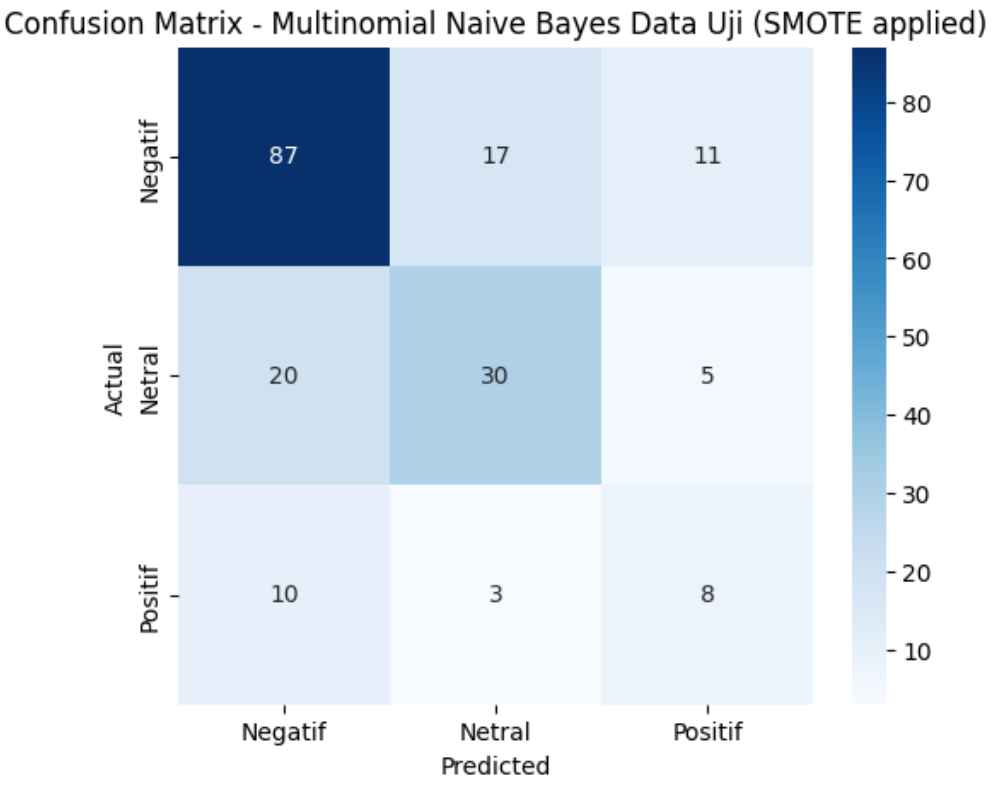
$$\mathrm{Skor}_{\mathrm{Kelas}} = \log(\mathrm{P}(\mathrm{Kelas}) + \sum_{\mathrm{i}=1}^{\mathrm{n}} f_i \cdot \log \mathrm{P}(\mathrm{w_i}|\mathrm{Kelas})$$

Misalnya:
Log prior kelas Positif = - 0,9629
Log kata “produktif” = - 3,02201574
Log kata “kerja” = - 3,022474195
Log kata “perintah” = - 3,675686709
Jika ketiga kata tersebut muncul masing-masing 1 kali, maka skor total:

$$\text{-0,9629} + (-3{,}0220) + (\text{- } 3{,}0225) + (\text{-}3{,}6757) = \text{-}10{,}6831$$

Proses ini dilakukan terhadap semua kelas, dan kelas dengan skor tertinggi dipilih sebagai hasil prediksi oleh model. Setelah pelatihan selesai, dilakukan prediksi terhadap data uji menggunakan metode predict(). Hasil prediksi ini kemudian dibandingkan dengan label sebenarnya untuk mengevaluasi performa model klasifikasi.

Model dilatih menggunakan 80% data latih yang telah dipisahkan sebelumnya, di mana frekuensi kata pada setiap kelas dihitung untuk memperkirakan bobot probabilitas kata terhadap kelas tersebut. Teknik *Laplace smoothing* diterapkan agar kata yang tidak muncul di suatu kelas tidak menghasilkan probabilitas nol. Dengan pendekatan ini, model dapat mengklasifikasikan *tweet* berdasarkan pola kemunculan kata dalam tiap kelas sentimen yang telah dipelajari selama proses pelatihan.



Gambar 1. Confusion Matrix Data Uji yang sudah diterapkan SMOTE

Setelah dilakukan pelatihan model menggunakan algoritma Multinomial Naive Bayes dengan data latih yang telah diseimbangkan melalui metode SMOTE (Arifiyanti & Wahyuni, n.d.), tahap berikutnya adalah mengevaluasi performa model terhadap data uji yang tidak terpengaruh oleh proses oversampling. Evaluasi dilakukan menggunakan confusion matrix untuk melihat seberapa baik model dapat mengklasifikasikan setiap kelas sentimen (Negatif, Netral, dan Positif) pada data uji.

Model berhasil mengklasifikasikan 87 data Negatif dengan benar, meskipun masih terdapat 17 data Negatif yang diprediksi sebagai Netral dan 11 data sebagai Positif. Untuk kelas Netral, 30 data berhasil diklasifikasikan dengan benar, sementara 20 data salah diprediksi sebagai Negatif dan 5 data sebagai Positif. Pada kelas Positif, hanya 8 data yang diprediksi benar, dengan 10 data diprediksi sebagai Negatif dan 3 sebagai Netral. Hasil ini menunjukkan bahwa penerapan SMOTE membantu menyeimbangkan pembelajaran pada kelas minoritas, namun model masih mengalami kesulitan terutama dalam mengenali pola pada kelas Positif.

Berikut adalah rumus dari SMOTE:

$$x_{syn} = x_i + (x_{knn} - x_i) \times \delta \quad \ldots (3)$$

Dimana x_{syn} adalah data sintetis yang akan diciptakan, x_i adalah data yang akan direplikasi, x_{knn} adalah data yang memiliki jarak terdekat dari data yang akan direplikasi, dan δ adalah nilai random antara 0 sampai 1.

Tabel 6. *Classification Report* pada Data Uji yang sudah diterapkan SMOTE

	Precision	***Recall***	***F1-Score***	***Support***
Positif	0,33	0,38	0,36	21
Netral	0,60	0,55	0,57	55
Negatif	0,74	0,76	0,75	115
Accuracy			0,65	191

	Precision	*Recall*	*F1-Score*	*Support*
Macro Avg.	0,56	0,56	0,56	191
Weighted Avg.	0,66	0,65	0,66	191

Sumber: Data Penelitian yang Sudah Diolah, 2025.

Setelah pelatihan model dengan data latih yang diseimbangkan menggunakan SMOTE, model diuji pada data uji untuk mengevaluasi performanya. Hasilnya menunjukkan 87 data Negatif berhasil diklasifikasikan dengan benar, sementara 17 data Negatif diprediksi sebagai Netral dan 11 sebagai Positif. Pada kelas Netral, 30 data diklasifikasikan dengan benar, tetapi 20 salah diprediksi sebagai Negatif dan 5 sebagai Positif. Untuk kelas Positif, hanya 8 data berhasil diklasifikasikan dengan benar, sedangkan 10 data diprediksi sebagai Negatif dan 3 sebagai Netral.

Evaluasi model menghasilkan akurasi sebesar 65,44% dengan *precision*, *recall*, dan *f1-score* yang bervariasi antar kelas. Kelas Negatif memiliki *precision* 0,74, *recall* 0,76, dan *f1-score* 0,75; kelas Netral memiliki *precision* 0,60, *recall* 0,55, dan *f1-score* 0,57; sedangkan kelas Positif menunjukkan nilai lebih rendah dengan *precision* 0,33, *recall* 0,38, dan *f1-score* 0,36. Nilai rata-rata *macro* untuk *precision*, *recall*, dan *f1-score* adalah 0,56, sementara *weighted average* berada di sekitar 0,66. Hasil ini menunjukkan bahwa meskipun SMOTE membantu menyeimbangkan data latih, model masih kesulitan dalam mengenali pola pada kelas minoritas, terutama kelas Positif, yang dipengaruhi oleh keterbatasan fitur dalam membedakan pola antar kelas pada data uji yang lebih variatif.

D. Kesimpulan

Penelitian ini menyimpulkan bahwa pelabelan sentimen tweet dilakukan oleh tiga anotator, yaitu dua manual dan satu menggunakan model IndoBERTweet, dengan penentuan label akhir berdasarkan suara mayoritas. Hasil analisis sentimen menunjukkan bahwa persepsi publik terhadap isu bonus demografi di Indonesia cenderung negatif, terlihat dari dominasi sentimen negatif dibandingkan netral maupun positif, yang mencerminkan adanya kekhawatiran atau pandangan pesimis masyarakat. Model terbaik setelah penerapan SMOTE menghasilkan akurasi 65,44% dengan *precision* dan *recall* yang cukup seimbang, menunjukkan kemampuan model mengenali pola sentimen secara moderat, meskipun masih perlu peningkatan terutama pada kelas minoritas.

Ucapan Terimakasih

Penulis mengucapkan puji syukur kepada Allah SWT. atas rahmat dan karunia-Nya sehingga artikel berjudul "Klasifikasi Sentimen Publik terhadap Isu Bonus Demografi di X Menggunakan Multinomial Naive Bayes" dapat diselesaikan sebagai salah satu syarat memperoleh gelar Sarjana Statistika di Universitas Islam Bandung. Penyelesaian skripsi ini tidak lepas dari dukungan berbagai pihak, di antaranya orang tua yang selalu memberikan doa dan motivasi, pimpinan fakultas dan program studi, dosen pembimbing, seluruh dosen Program Studi Statistika, keluarga, serta teman-teman dari Himpunan Mahasiswa Statistika, teman-teman Top Management & Ketua Bidang di Himpunan Mahasiswa Statistika periode 2023-2024, teman-teman bidang Pembinaan Aparatur Organisasi Himasta Unisba, dan sahabat-sahabat SMA yang selalu memberikan semangat dan menjaga silaturahmi.

Daftar Pustaka

Ajeng Mega Pratiwi, & Mutaqin, A. K. (2021). Penerapan Algoritma Naïve Bayes Classifier dalam Memprediksi Status Keberlanjutan Polis Nasabah Asuransi PT.X. *Jurnal Riset Statistika*, *1*(2), 117–126. https://doi.org/10.29313/jrs.v1i2.435

anakblogger. (2025, May 30). *Kelebihan & Kekurangan Algoritma Naive Bayes*. Anakblogger.Com. https://www.anakblogger.com/2020/01/kelebihan-kekurangan-algoritma-naive.html?utm_source=chatgpt.com

Arifiyanti, A. A., & Wahyuni, E. D. (n.d.). *SMOTE: METODE PENYEIMBANG KELAS PADA KLASIFIKASI DATA MINING*. https://www.cs.

Azzahra Nasution, D., Khotimah, H. H., & Chamidah, N. (2019). PERBANDINGAN NORMALISASI DATA UNTUK KLASIFIKASI WINE MENGGUNAKAN ALGORITMA K-NN. *CESS (Journal of Computer Engineering System and Science)* , *4*(1), 2502–7131.

Fauziah, G., & Sunendiari, S. (2021). Estimasi Pseudo Poisson Maximum Likelihood untuk Mengatasi Masalah dalam Model Log-Linear pada Kasus Kusta di Jawa Barat Tahun 2018. *Jurnal Riset Statistika*, *1*(1), 57–62. https://doi.org/10.29313/jrs.v1i1.147

Handayani, F., Feddy, D., & Pribadi, S. (n.d.). *Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110*.

Luhung Mustika Budiharti, & Sunendiari, S. (2021). Pemodelan dan Pemetaan Jumlah Penderita Kusta di Jawa Barat dengan Regresi Binomial Negatif dan Flexibly Shaped Spatial Scan Statistic. *Jurnal Riset Statistika*, *1*(2), 99–106. https://doi.org/10.29313/jrs.v1i2.409

Putri, H. J., & Ramdani, Y. (2024). Prediksi Hujan Rencana Tahunan Di Stasiun Palolo, Sulawesi Tengah menggunakan Metode Gumbel, Log Normal, Dan Log Pearson III. *DataMath: Journal of Statistics and Mathematics*, *2*(1), 17–24.

Rafli Tanjung, M., Heryana, N., Sistem Informasi, S., Singaperbangsa Karawang Jl HSRonggo Waluyo, U., Timur, T., & Barat, J. (2024). ANALISIS SENTIMEN PENGGUNA PLATFORM X TERHADAP INDONESIA EMAS 2045 MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 4).

Retno Widiarti, H., Novita Sari, I., Nadzifah, atun, Anggraini, F., Taribu, K., Layla Syamsi, I., & Ainur Rokhim, D. (2025). *Pelatihan Kewirausahaan Siswa SMP Sebagai Bekal Menuju Bonus Demografi 2045* (Vol. 9, Issue 2). http://journal.unhas.ac.id/index.php/panritaabdi

Sabrani, A., Gede Putu Wirarama Wedashwara, I. W., & Bimantoro, F. (n.d.). *METODE MULTINOMIAL NAÏVE BAYES UNTUK KLASIFIKASI ARTIKEL ONLINE TENTANG GEMPA DI INDONESIA (Multinomial Naïve Bayes Method for Classification of Online Article About Earthquake in Indonesia)*. http://jtika.if.unram.ac.id/index.php/JTIKA/

Saputra, N., Adji, T. B., & Permanasari, A. E. (2015). ANALISIS SENTIMEN DATA PRESIDEN JOKOWI DENGAN PREPROCESSING NORMALISASI DAN STEMMING MENGGUNAKAN METODE NAIVE BAYES DAN SVM Oleh. In *Jurnal Dinamika Informatika* (Vol. 5, Issue 1).

Yuyun, Nurul Hidayah, & Supriadi Sahibu. (2021). Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(4), 820–826. https://doi.org/10.29207/resti.v5i4.3146

Zakiyatis Salmaini, & Suliadi. (2021). SPC (Statistical Process Control) Fase II Diagram Kendali Cusum (Cumulative Sum) Nonparametrik Berdasarkan Statistik Mann-Whitney Pada Data Harga Saham PT NO. *Jurnal Riset Statistika*, *1*(2), 83–91. https://doi.org/10.29313/jrs.v1i2.404