

Klasifikasi Opini Publik terhadap Kenaikan PPN 12% di Platform X menggunakan *Multinomial Naïve Bayes*

Hani Brilianti Rochmanto¹, Gangga Anuraga¹, Muhammad Athoillah¹, Harun Al Azies^{2,3}, Muhammad Naufal^{2,4}

¹Program Studi Statistka, Fakultas Sains dan Teknologi, Universitas PGRI Adi Buana Surabaya, rochmantohani@unipasby.ac.id

²Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang harun.alazies@dsn.dinus.ac.id

³*Research Center for Quantum Computing and Materials Informatics*, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang

⁴*Research Center for Intelligent Distributed Surveillance and Security*, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang

Abstract. The increase in Value-Added Tax to 12% in 2025 has sparked diverse public opinions on the social media platform X (Twitter). This study aims to classify public sentiment toward the policy using Multinomial Naïve Bayes with a Term Frequency-Inverse Document Frequency (TF-IDF) approach. Multinomial Naïve Bayes is a probabilistic classification algorithm that assumes feature independence. Data were collected through web crawling using the keyword "ppn 12%" and underwent pre-processing, including text normalization, stopword removal, and stemming. To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. The best-performing model was obtained by tuning the alpha hyperparameter to 0.01, achieving an average accuracy of 83.37%, precision of 83.32%, recall of 83.38%, and an F1-score of 82.99% using 10-fold cross-validation. The findings indicate that Multinomial Naïve Bayes, combined with SMOTE and hyperparameter tuning, effectively classifies public sentiment and provides insights into public responses regarding the Value-Added Tax policy.

Keywords: *Sentiment Analysis, Multinomial Naive Bayes, Value-Added Tax, SMOTE, TF-IDF*

Abstrak. Kenaikan Pajak Pertambahan Nilai (PPN) menjadi 12% pada tahun 2025 telah memicu beragam opini publik di platform media sosial X (Twitter). Penelitian ini bertujuan untuk mengklasifikasikan sentimen masyarakat terhadap kebijakan tersebut menggunakan *Multinomial Naïve Bayes* dengan pendekatan *Term Frequency-Inverse Document Frequency* (TF-IDF). *Multinomial Naïve Bayes* merupakan algoritma klasifikasi berbasis probabilistik yang bekerja dengan mengasumsikan independensi antar fitur. Data dikumpulkan melalui proses *web crawling* dengan kata kunci "ppn 12%", kemudian dilakukan *pre-processing* data termasuk normalisasi teks, penghapusan *stopwords*, serta *stemming*. Data yang tidak seimbang diatasi dengan *Synthetic Minority Over-sampling Technique* (SMOTE). Model terbaik diperoleh dengan *tuning hyperparameter alpha* pada nilai 0.01, menghasilkan akurasi rata-rata sebesar 83,37%, presisi 83,32%, recall 83,38%, dan *F1-score* 82.99% menggunakan *10-fold cross-validation*. Hasil penelitian ini menunjukkan bahwa *Multinomial Naïve Bayes* dengan *SMOTE* dan *tuning parameter* mampu mengklasifikasikan opini publik secara efektif dan memberikan wawasan terhadap respons masyarakat terkait kebijakan PPN.

Kata Kunci: *Analisis Sentimen, Multinomial Naive Bayes, PPN 12%, SMOTE, TF-IDF*

1 Pendahuluan

Kenaikan Pajak Pertambahan Nilai (PPN) menjadi 12% pada tahun 2025 telah menjadi topik perbincangan luas di masyarakat sejak diumumkan dalam Undang-Undang Harmonisasi Peraturan Perpajakan (UU HPP) pada tahun 2021 [1]. Kebijakan ini bertujuan untuk meningkatkan penerimaan negara dan menyesuaikan kebijakan perpajakan dengan dinamika ekonomi global serta kebutuhan pembangunan nasional [2], [3]. Namun, kebijakan ini menimbulkan beragam reaksi di masyarakat. Sebagian pihak mengkhawatirkan dampaknya terhadap daya beli dan beban ekonomi, terutama bagi kelompok berpenghasilan rendah serta pelaku Usaha Mikro, Kecil, dan Menengah (UMKM) [4]. Di sisi lain, pemerintah menegaskan bahwa kenaikan PPN ini lebih difokuskan pada barang dan jasa tertentu sehingga tidak akan berdampak langsung pada kebutuhan pokok masyarakat.

Dalam era digital, media sosial, khususnya X (Twitter), menjadi platform utama bagi masyarakat untuk mengekspresikan opini terhadap berbagai topik, termasuk kebijakan baru [5]. Oleh karena itu, analisis sentimen terhadap opini publik di media sosial menjadi penting untuk memahami bagaimana masyarakat merespons kebijakan kenaikan PPN. Analisis ini tidak hanya memberikan wawasan mengenai persepsi publik, tetapi juga dapat menjadi dasar bagi pemerintah dalam merancang strategi komunikasi kebijakan yang lebih efektif serta merespons kekhawatiran masyarakat secara lebih tepat.

Analisis sentimen merupakan salah satu metode dalam pemrosesan bahasa alami (*Natural Language Processing/NLP*) yang telah banyak digunakan untuk memahami opini publik terhadap isu-isu sosial dan ekonomi [6]. Studi terdahulu menunjukkan bahwa analisis sentimen telah diterapkan dalam berbagai kebijakan fiskal di Indonesia, termasuk studi mengenai kebijakan terkait Tabungan Perumahan Rakyat (Tapera) yang memicu perdebatan di kalangan masyarakat [7]. Dalam penelitian ini, analisis sentimen dilakukan menggunakan berbagai teknik *machine learning*, termasuk *Naïve Bayes Multinomial*, untuk menganalisis sentimen publik terhadap kebijakan Tapera. Hasil penelitian menunjukkan bahwa *Naïve Bayes Multinomial* memiliki performa lebih baik dibandingkan algoritma lain dalam klasifikasi sentimen negatif dan netral.

Salah satu tantangan dalam analisis sentimen adalah ketidakseimbangan kelas data, yang dapat mengakibatkan bias dalam hasil klasifikasi. Oleh karena itu, penelitian ini menerapkan teknik *Naïve Bayes Multinomial* dengan pendekatan *Synthetic Minority Over-sampling Technique (SMOTE)* untuk menyeimbangkan data guna meningkatkan kinerja model klasifikasi. Selain itu, evaluasi model dilakukan menggunakan berbagai metrik, seperti akurasi, presisi, recall, dan *F1-score*, untuk memastikan keandalan hasil klasifikasi. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan metode analisis sentimen terhadap kebijakan ekonomi menggunakan pendekatan NLP serta menyediakan model klasifikasi yang dapat digunakan oleh pembuat kebijakan untuk memahami opini publik. Penelitian ini diharapkan dapat berkontribusi dalam mengembangkan metode analisis sentimen yang lebih akurat dan adaptif terhadap distribusi data yang tidak seimbang.

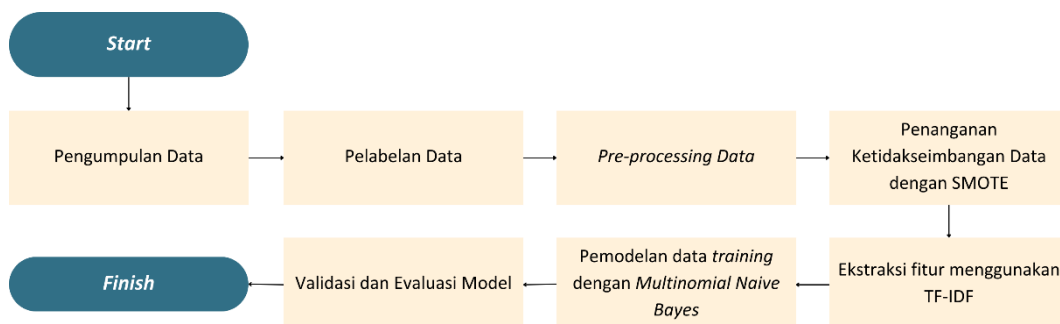
2 Metode Penelitian

2.1 Sumber Data

Data yang digunakan dalam penelitian ini diperoleh melalui proses crawling dari media sosial X (Twitter) dengan menggunakan kata kunci yang relevan, yaitu "ppn 12%", dalam rentang waktu 1 Oktober 2024 hingga 31 Oktober 2024. Data yang dikumpulkan berupa teks cuitan dalam bahasa Indonesia yang mencerminkan opini publik terhadap kebijakan kenaikan PPN. Untuk memastikan kualitas data, dilakukan *pre-processing* data, termasuk penghapusan duplikasi, penghapusan tanda baca dan karakter khusus, serta normalisasi teks guna meningkatkan akurasi analisis sentimen. Selanjutnya, data diberi label sentimen secara manual menurut persepsi peneliti dengan mengkategorikan opini publik ke dalam sentimen positif, negatif, atau netral berdasarkan makna dalam setiap cuitan.

2.2 Metode dan Langkah Analisis

Penelitian ini menggunakan pendekatan analisis sentimen berbasis machine learning dengan algoritma *Multinomial Naïve Bayes* dan teknik *Term Frequency-Inverse Document Frequency* (TF-IDF) sebagai representasi fitur. Proses analisis diawali dengan pengumpulan data melalui *web crawling*, diikuti dengan tahap *pre-processing* yang mencakup penghapusan karakter khusus, normalisasi teks, penghapusan stopwords, serta stemming untuk menyederhanakan bentuk kata. Setelah itu, data diberi label sentimen secara manual ke dalam kategori positif, negatif, atau netral. Untuk mengatasi ketidakseimbangan kelas dalam data, digunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) guna meningkatkan distribusi data dalam kelas minoritas. Model klasifikasi dilatih dengan *Multinomial Naïve Bayes* dengan menyesuaikan nilai *alpha* menggunakan *hyperparameter tuning* untuk mengoptimalkan kinerja model. Evaluasi model dilakukan menggunakan metrik akurasi, presisi, *recall*, dan *F1-score* dengan *cross-validation* sebanyak 10 kali lipat untuk memastikan keandalan hasil klasifikasi. Langkah-langkah tersebut digambarkan dalam diagram alir pada Gambar 1 berikut.



Gambar 1. Diagram Alir Penelitian

2.2.1 Pre-Processing Data

Pada tahap *pre-processing*, data hasil *crawling* dari media sosial X (Twitter) terlebih dahulu dibersihkan untuk menghilangkan elemen yang tidak relevan. Agar data teks dapat dianalisis secara optimal, sejumlah teknik *pre-processing* diterapkan untuk membersihkan dan menormalkan teks. *Pre-processing* data dilakukan melalui beberapa tahapan sebagai berikut [8]:

- a. Pembersihan teks (*cleansing*): Menghapus karakter khusus, angka, dan emoji yang tidak relevan dalam analisis.
- b. Normalisasi kata tidak baku: Mengonversi istilah slang atau kata tidak baku menjadi bentuk baku sesuai dengan kaidah bahasa.
- c. *Case folding*: Mengubah seluruh teks menjadi huruf kecil guna menjaga konsistensi dalam pemrosesan.
- d. Tokenisasi: Memecah teks menjadi unit kata-kata yang lebih kecil agar lebih mudah dianalisis.
- e. Penghapusan *stopword*: Mengeliminasi kata-kata umum yang tidak memiliki makna signifikan dalam analisis, seperti "dan", "di", atau "ke".
- f. *Stemming*: Mengubah kata berimbuhan menjadi bentuk dasar untuk mengurangi variasi kata yang memiliki makna serupa.

2.2.2 Synthetic Minority Over-sampling Technique (SMOTE)

Synthetic Minority Over-sampling Technique (SMOTE) merupakan sebuah metode *resampling* yang bertujuan untuk menyeimbangkan distribusi kelas dengan menghasilkan sampel sintesis pada kelas minoritas [9]. SMOTE bekerja dengan membuat data baru berdasarkan interpolasi dari titik-titik yang sudah ada dalam ruang fitur, sehingga tidak sekadar menduplikasi data minoritas tetapi menciptakan variasi yang lebih representatif [10]. Dengan menerapkan SMOTE, model dapat menghindari bias terhadap kelas mayoritas serta meningkatkan performa klasifikasi dalam mengenali pola sentimen publik secara lebih akurat.

2.2.3 Term Frequency-Inverse Document Frequency (TF-IDF)

Ekstraksi fitur ke dalam TF-IDF digunakan untuk mengukur seberapa sering suatu kata muncul dalam sebuah dokumen (*Term Frequency*) sekaligus menyesuaikan bobotnya berdasarkan seberapa umum kata tersebut muncul di seluruh dokumen (*Inverse Document Frequency*) [11]. Perhitungan TF-IDF dilakukan menggunakan rumus berikut.

$$w_{ij} = TF_{ij} \times IDF \quad (1)$$

$$IDF = \log\left(\frac{N}{df_{ij}}\right) \quad (2)$$

Di mana w_{ij} : adalah bobot ke- j dalam dokumen ke- i , TF_{ij} merepresentasikan jumlah kemunculan kata ke- j dalam dokumen ke- i , N adalah total jumlah dokumen dalam korpus, dan df_{ij} merupakan jumlah dokumen yang mengandung data ke- j [12].

2.2.4 Multinomial Naïve Bayes

Multinomial Naïve Bayes (MNB) adalah salah satu algoritma klasifikasi yang sering digunakan dalam analisis sentimen berbasis teks. Algoritma ini bekerja berdasarkan prinsip Teorema Bayes dengan asumsi bahwa setiap fitur atau kata dalam dokumen bersifat independen satu sama lain [13]. Model MNB efektif untuk data teks karena menghitung probabilitas suatu kategori berdasarkan distribusi kata dalam dokumen. Pada analisis sentimen, MNB mengklasifikasikan teks ke dalam kategori positif, negatif, atau netral berdasarkan distribusi kemunculan kata dalam tiap kelas. Probabilitas suatu teks termasuk dalam kelas C_{map} dihitung melalui persamaan berikut menggunakan rumus berikut:

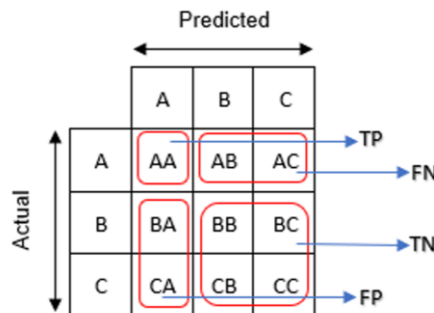
$$C_{map} = \arg \max_{c \in C} \hat{P}(c) \prod_{k=1}^n \hat{P}(t_k|c) \quad (3)$$

$\hat{P}(c)$ dan $\hat{P}(t|c)$ dihitung pada saat proses *training* di mana $\hat{P}(c)$ merupakan probabilitas prior suatu dokumen termasuk ke dalam kelas c dan $\hat{P}(t|c)$ adalah probabilitas frekuensi relatif *term* (t) pada kelas c [14].

2.2.5 Validasi dan Evaluasi

Validasi model dilakukan untuk memastikan keandalan hasil klasifikasi dan menghindari bias akibat pembagian data. *K-fold cross validation* adalah metode yang digunakan untuk membagi data menjadi k bagian (*folds*), dengan salah satu bagian digunakan sebagai data uji (*testing set*) dan sisanya sebagai data latih (*training set*). Proses ini dilakukan secara bergantian sebanyak k kali sehingga seluruh data mendapatkan kesempatan untuk menjadi data uji [15].

Langkah selanjutnya adalah mengevaluasi hasil klasifikasi. Pada tahap ini, nilai akurasi, presisi, *recall*, dan *F1-score* dihitung menggunakan *confusion matrix*. *Confusion matrix* adalah sebuah tabel yang menunjukkan hasil klasifikasi aktual dibandingkan dengan prediksi yang dihasilkan dari model [16]. *Confusion matrix* memiliki ukuran $L \times L$, di mana L adalah jumlah kelas dalam klasifikasi. Dalam penelitian ini, digunakan *confusion matrix* berukuran 3×3 karena terdapat tiga kelas label. Struktur dari *confusion matrix* ditunjukkan pada Gambar 2.



Gambar 2. Confusion Matrix Berukuran 3×3 [16]

Hasil dari pemberian label secara manual menunjukkan bahwa data terbagi menjadi tiga kategori sentimen, yaitu negatif (-1), netral (0), dan positif (1). *Wordcloud* pada Gambar 3 menunjukkan kata-kata yang paling sering muncul dalam setiap kategori sentimen terkait kenaikan PPN menjadi 12%. Pada sentimen negatif, kata-kata seperti *naik*, *harga*, *rakyat*, dan *ekonomi* mendominasi, mencerminkan kekhawatiran masyarakat terhadap dampak kenaikan pajak pada daya beli dan kondisi ekonomi. Sentimen netral lebih banyak berisi istilah teknis seperti *pajak*, *PPN*, *persen*, dan *faktur*, yang menunjukkan pembahasan seputar aturan tanpa opini emosional. Sementara sentimen positif mengindikasikan bahwa terdapat masyarakat yang memiliki pandangan mendukung terhadap kebijakan pajak yang diterapkan pemerintah terkait kenaikan PPN tersebut.

Namun, distribusi awal data menunjukkan adanya ketidakseimbangan kelas, di mana data dengan label sentimen positif jauh lebih sedikit dibandingkan dengan sentimen negatif dan netral. Ketidakseimbangan kelas pada data dapat mempengaruhi performa model, karena algoritma klasifikasi cenderung lebih akurat dalam memprediksi kelas dengan jumlah data yang lebih banyak, sementara kelas dengan jumlah data lebih sedikit memiliki akurasi lebih rendah [18]. Untuk mengatasi masalah tersebut, diterapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) untuk menambah jumlah data pada kelas minoritas (positif) hingga seimbang dengan kelas lainnya. Perbandingan distribusi data setelah penerapan SMOTE ditampilkan pada Tabel 2 berikut.

Tabel 2. Distribusi Data Setelah SMOTE

Label Sentimen	Sebelum <i>Preprocessing</i>	Setelah <i>Preprocessing</i>
Negatif (-1)	397	397
Netral (0)	394	397
Positif (1)	134	397

Setelah proses *preprocessing*, tahap selanjutnya adalah ekstraksi fitur menggunakan metode TF-IDF. Teknik ini digunakan untuk mengubah teks ulasan menjadi representasi numerik sehingga dapat diolah oleh model. Metode TF-IDF mengukur seberapa sering sebuah kata muncul dalam sebuah dokumen (TF) dan menyesuaikannya berdasarkan seberapa sering kata tersebut muncul di seluruh dokumen (IDF). Dengan demikian, kata-kata yang umum dan tidak terlalu bermakna mendapatkan bobot lebih rendah, sementara kata-kata yang lebih spesifik dalam topik tertentu mendapatkan bobot lebih tinggi. Setelah dilakukan ekstraksi fitur, setiap teks diubah menjadi vektor dengan nilai bobot yang merepresentasikan pentingnya kata-kata dalam dokumen tersebut.

Dalam penelitian ini, data yang telah dikonversi ke bentuk TF-IDF digunakan sebagai input untuk melatih model MNB. Untuk memastikan model dapat bekerja secara optimal, dilakukan pembagian data menggunakan *10-fold cross-validation*. Teknik ini membagi data menjadi 10 bagian (*folds*) di mana model dilatih pada 9 subset dan diuji pada 1 subset lainnya secara bergantian. Kemudian, evaluasi model dilakukan untuk mengukur kinerja klasifikasi berdasarkan akurasi, presisi, *recall*, dan *F1-score*. Metode evaluasi yang digunakan adalah *confusion matrix*, yang memberikan gambaran bagaimana prediksi model dibandingkan dengan label asli.

Untuk meningkatkan performa model, dilakukan tuning hyperparameter dengan mengoptimalkan nilai α dalam *Multinomial Naïve Bayes*. Parameter α berperan dalam menangani kata-kata yang jarang muncul dalam data pelatihan. Percobaan dilakukan dengan menguji beberapa nilai α , yaitu 0,01; 0,10; 0,50; dan 1,00 menggunakan *Grid Search*. Hasil evaluasi dari berbagai nilai α ditampilkan pada Tabel 3 berikut.

Tabel 3. Hasil Evaluasi Model *Naïve Bayes* Berdasarkan Nilai Alpha

Nilai Alpha	Rata-rata Akurasi	Rata-rata Presisi	Rata-rata Recall	Rata-rata F1-Score
0,01	83,37%	83,32%	83,38%	82,99%
0,10	83,20%	83,25%	83,22%	82,76%
0,50	81,78%	81,84%	81,79%	81,15%
1,00	81,02%	81,15%	81,04%	80,33%

Dari hasil *tuning hyperparameter*, terlihat bahwa nilai α terbaik adalah 0,01 dengan nilai akurasi 83,37%, nilai presisi 83,32%, nilai *recall* 83,38% dan *F1-score* 82,99%. Hal ini menunjukkan bahwa model bekerja lebih optimal ketika nilai α rendah, karena memberikan pengaruh minimal terhadap *smoothing*, sehingga kata-kata dengan frekuensi rendah tetap berkontribusi dalam klasifikasi. Hasil klasifikasi menunjukkan bahwa model *Multinomial Naïve Bayes* mampu mengklasifikasikan sentimen terkait kenaikan PPN dengan akurasi yang cukup tinggi, terutama setelah dilakukan SMOTE untuk menangani ketidakseimbangan kelas.

4 Kesimpulan

Berdasarkan hasil penelitian, dapat disimpulkan bahwa model *Multinomial Naïve Bayes* (MNB) dengan α 0,01 mampu mengklasifikasikan sentimen masyarakat terhadap kebijakan kenaikan PPN 12% dengan tingkat akurasi yang cukup tinggi, yaitu 83,37%. Proses *pre-processing* data berhasil meningkatkan kualitas teks dengan menghilangkan elemen yang tidak relevan dan menormalisasi kata, sehingga data lebih siap untuk dianalisis. Selain itu, teknik *Synthetic Minority Over-sampling Technique* (SMOTE) terbukti efektif dalam menangani ketidakseimbangan kelas, yang awalnya dapat menyebabkan bias dalam klasifikasi. Untuk penelitian lanjutan, pengembangan model dapat ditingkatkan dengan mengeksplorasi teknik lain seperti *deep learning* atau kombinasi metode klasifikasi untuk memperoleh hasil yang lebih optimal.

5 Daftar Pustaka

- [1] Lala Febiola and Hasnawati, "ANALISA PERSEPSI WAJIB PAJAK ORANG PRIBADI TERHADAP KEBIJAKAN KENAIKAN TARIF PAJAK PERTAMBAHAN NILAI MENJADI 12% DI TAHUN 2025," *Journal of Social and Economics Research*, vol. 6, no. 1, pp. 1863–1876, Aug. 2024, doi: 10.54783/jsr.v6i1.543.

- [2] K. Fatha Rabani *et al.*, “Analisis Statistik Pengaruh Kenaikan PPN Tahun 2025 Terhadap Harga Permintaan, Kondisi Pasar dan Sosial Ekonomi Indonesia,” *JISMA: Jurnal Ilmu Sosial, Manajemen, dan Akuntansi*, vol. 3, no. 2, pp. 1315–1322, Jun. 2024. Available: <https://melatijournal.com/index.php/jisma/article/view/546>
- [3] H. Hajatina and U. Hasanah, “Analisis Dampak Kenaikan Tarif Pajak Pertambahan Nilai (PPN) Terhadap Perilaku Konsumsi Dan Kesejahteraan Masyarakat Di Indonesia,” *JURNAL MANAJEMEN DAN BISNIS*, vol. 3, no. 2, pp. 36–51, Dec. 2024, doi: 10.36490/JMDB.V3I2.1452.
- [4] P. Kenaikan Pajak Pertambahan Nilai Terhadap Pendapatan Wajib Pajak Badan Usaha di Kota Metro Era Yudistira and T. Nisa, “Pengaruh Kenaikan Pajak Pertambahan Nilai (PPN) Terhadap Pendapatan Wajib Pajak Badan Usaha di Kota Metro,” *Innovative: Journal Of Social Science Research*, vol. 4, no. 3, pp. 8090–8104, May 2024, doi: 10.31004/INNOVATIVE.V4I3.10872.
- [5] 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, Feb. 2024, doi: 10.36040/JATI.V8I1.8378.
- [6] A. Mukti, A. D. Hadiyanti, A. Nurlaela, and J. Panjaitan, “Sistem Analisa Sentiment Bakal Calon Presiden 2024 Menggunakan Metode NLP Berbasis Web,” *SOSCIED*, vol. 6, no. 1, pp. 128–140, Jun. 2023, doi: 10.32531/JSOSCIED.V6I1.621.
- [7] H. Leidiyana, T. Misriati, and R. Aryanti, “Klasifikasi Sentimen Terhadap Kebijakan Tapera Menggunakan Komparasi Machine Learning dan SMOTE,” *Jurnal Komtika (Komputasi dan Informatika)*, vol. 8, no. 2, pp. 125–135, Nov. 2024, doi: 10.31603/KOMTIKA.V8I2.12595.
- [8] S. Shevira, I. Made, A. D. Suarjaya, and P. Wira Buana, “Pengaruh Kombinasi dan Urutan Pre-Processing pada Tweets Bahasa Indonesia,” 2022.
- [9] C. Cahyaningtyas, Y. Nataliani, and I. R. Widiyari, “Analisis Sentimen Pada Rating Aplikasi Shopee Menggunakan Metode Decision Tree Berbasis SMOTE,” *AITI*, vol. 18, no. 2, pp. 173–184, Nov. 2021, doi: 10.24246/aiti.v18i2.173-184.
- [10] A. A. ZD, Kusriani, and A. D. Hartanto, “ANALISIS SENTIMEN PEMINDAHAN IBU KOTA NEGARA (IKN) MENGGUNAKAN METODE OVERSAMPLING SYNTHETIC MINORITY (SMOTE),” *Jurnal INSTEK (Informatika Sains dan Teknologi)*, vol. 9, no. 2, pp. 324–335, Dec. 2024, doi: 10.24252/INSTEK.V9I2.50944.
- [11] A. E. Widjaja, A. Fransisko, C. A. Haryani, and Hery, “Text Mining Application with K-Means Clustering to Identify Sentiments and Popular Topics: a Case Study of the three Largest Online Marketplaces in Indonesia,” *Journal of Applied Data Sciences*, vol. 4, no. 4, pp. 441–453, Dec. 2023, doi: 10.47738/jads.v4i4.134.
- [12] H. Al Azies, H. B. Rochmanto, C. A. Pravesti, and F. Fitriani, “Integrating counseling with technology: An evaluation of the Bicarakan.id application through user review analysis with machine learning,” *KONSELI: Jurnal Bimbingan dan Konseling (E-Journal)*, vol. 11, no. 2, pp. 205–218, Nov.

2024. Available:
<https://ejournal.radenintan.ac.id/index.php/konseli/article/view/24357>
- [13] S. Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti, N. Hidayah, S. Sahibu, M. Sistem Komputer, and S. Handayani Makassar, “Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 820–826, Aug. 2021, doi: 10.29207/RESTI.V5I4.3146.
- [14] T. Ernayanti, A. Rusgiyono, A. Rachman Hakim, D. Statistika, F. Sains dan Matematika, and U. Diponegoro, “PENGUNAAN SELEKSI FITUR CHI-SQUARE DAN ALGORITMA MULTINOMIAL NAÏVE BAYES UNTUK ANALISIS SENTIMEN PELANGGGAN TOKOPEDIA,” *Jurnal Gaussian*, vol. 11, no. 4, pp. 562–571, Feb. 2023, doi: 10.14710/J.GAUSS.11.4.562-571.
- [15] T. Ridwansyah, “Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier,” *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 2, no. 5, pp. 178–185, Apr. 2022, doi: 10.30865/KLIK.V2I5.362.
- [16] K. Wabang, O. D. Nurhayati, and Farikhin, “Application of The Naïve Bayes Classifier Algorithm to Classify Community Complaints,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 5, pp. 872–876, Nov. 2022, doi: 10.29207/RESTI.V6I5.4498.
- [17] A. P. Randika Setiawan, B. Setya Rintyama, and T. Adi Cahyanto, “Preferensi Konsumen Terhadap Produk By.U Dan MPWR Dengan Analisis Sentimen Berbasis Multinomial Naïve Bayes Consumer Preferences On By.U And MPWR Products Using Sentimen Analysis Based On Multinomial Naïve Bayes,” 2021. Available: <http://jurnal.unmuhjember.ac.id/index.php/JST>
- [18] M. P. Pulungan, A. Purnomo, and A. Kurniasih, “Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Kepribadian MBTI Menggunakan Naive Bayes Classifier,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 5, pp. 1033–1042, Oct. 2024, doi: 10.25126/jtiik.1077989.