

Analisis Sentimen Pengguna X terhadap Kebijakan PPN 12% Menggunakan Naive Bayes

Alwi Andika Panggabean¹, Diah Putri Kartikasari², Rafif Risdi Aulia³, Tiara Ayu Triarta Tambak⁴,
Siti Fadiyah Nabila⁵, Mhd. Furqan⁶

Ilmu komputer, Sains dan teknologi, Universitas Islam Negeri Sumatera Utara

¹alwiandika2@gmail.com, ²diah0701221039@uinsu.ac.id, ³rafif.raulia@gmail.com, ⁴tiaratriarta@gmail.com,
⁵fadiyahnabila20@gmail.com, ⁶mfurqan@uinsu.ac.id

Abstract

The policy of increasing the Value Added Tax (VAT) rate from 11% to 12%, planned to take effect in 2025, has sparked various public reactions, particularly on social media. This study aims to analyze the sentiment of users on social media platform X (formerly Twitter) regarding the policy using the Naive Bayes method implemented in the R programming language. Data was collected from tweets relevant to the 12% VAT topic, then processed through preprocessing and manual labeling stages. The analysis results show that negative sentiment dominates with a proportion of 39%, followed by neutral sentiment at 32%, and positive sentiment at 29%. The performance evaluation of the Naive Bayes model shows an accuracy of 50%, with the highest classification precision in the negative category. Further analysis of key terms and discussion topics reveals that concerns about economic burdens and impacts on SMEs are the main sources of negative sentiment, while positive sentiment is associated with hopes for improvements in public services and development. This study provides valuable insights for policymakers to better understand public perception of fiscal policy through data-driven analysis.

Keywords: sentiment analysis, x users, vat policy, vat 12%, naive bayes.

Abstrak

Kebijakan kenaikan Pajak Pertambahan Nilai (PPN) dari 11% menjadi 12% yang direncanakan berlaku pada tahun 2025 telah menimbulkan berbagai reaksi publik, terutama di media sosial. Penelitian ini bertujuan untuk menganalisis sentimen pengguna media sosial X (sebelumnya Twitter) terhadap kebijakan tersebut menggunakan metode Naive Bayes yang diimplementasikan dalam bahasa pemrograman R. Data diperoleh dari tweet yang relevan dengan topik PPN 12%, kemudian diproses melalui tahapan pra-pemrosesan dan pelabelan manual. Hasil analisis menunjukkan bahwa sentimen negatif mendominasi dengan proporsi 39%, diikuti sentimen netral 32%, dan sentimen positif 29%. Evaluasi performa model Naive Bayes menunjukkan akurasi sebesar 50%, dengan ketepatan klasifikasi tertinggi pada kategori negatif. Analisis lebih lanjut terhadap istilah kunci dan topik diskusi mengungkapkan bahwa kekhawatiran terhadap beban ekonomi dan dampak terhadap UMKM menjadi sumber utama sentimen negatif, sementara sentimen positif dikaitkan dengan harapan terhadap perbaikan layanan publik dan pembangunan. Penelitian ini memberikan wawasan penting bagi pembuat kebijakan untuk memahami persepsi publik terhadap kebijakan fiskal secara lebih mendalam dan berbasis data.

Kata kunci: Analisis Sentiment, Pengguna x, Kebijakan Ppn, Ppn 12%, Naive Bayes.

1. Pendahuluan

Pajak adalah salah satu sumber penerimaan utama negara, memainkan peran penting dalam mendorong pertumbuhan bangsa. Pajak Pertambahan Nilai (PPN), yang dikenakan atas pembelian barang dan jasa di dalam negeri, adalah salah satu jenis pajak yang berkontribusi besar. Pemerintah Indonesia mengeluarkan Undang-Undang Harmonisasi Peraturan Perpajakan (UU HPP) untuk meningkatkan penerimaan negara dan meningkatkan rasio pajak terhadap Produk Domestik Bruto (PDB). Tarif PPN naik secara bertahap dari 10% menjadi 11% pada tahun 2022 dan direncanakan naik lagi menjadi 12% pada tahun 2025 [2].

Rencana kenaikan tarif PPN ini memicu beragam tanggapan dari masyarakat [15]. Di satu sisi, kebijakan ini dianggap sebagai langkah strategis untuk memperkuat fiskal negara. Sebaliknya, kekhawatiran muncul tentang bagaimana hal itu akan berdampak pada harga barang dan daya beli masyarakat, khususnya kelompok berpenghasilan menengah ke bawah. Media sosial telah berkembang menjadi sarana penting bagi masyarakat di era internet untuk menyampaikan pendapat, kritik, dan tanggapan terhadap berbagai kebijakan publik [11]. Media sosial X (sebelumnya Twitter) adalah salah satu yang paling banyak digunakan, yang memungkinkan pengguna untuk membagikan pendapat secara real-time dan terbuka [14]. Data dari media sosial ini dapat dianalisis untuk menggambarkan persepsi publik terhadap suatu isu, termasuk kebijakan kenaikan PPN [7].

Kebijakan fiskal, khususnya perubahan tarif PPN, telah menjadi isu yang sangat sensitif di masyarakat. Dalam beberapa tahun terakhir, berbagai studi menunjukkan bahwa perubahan kebijakan pajak kerap menimbulkan reaksi luas di media sosial, yang dapat digunakan sebagai indikator awal persepsi publik terhadap kebijakan tersebut [1].

Dalam text mining, analisis sentimen digunakan untuk menemukan dan mengkategorikan opini publik ke dalam kategori positif, negatif, atau netral [6]. Analisis sentimen di media sosial tidak hanya membantu pemerintah dalam memahami respons masyarakat secara real-time, tetapi juga dapat digunakan untuk mengevaluasi efektivitas komunikasi kebijakan dan mengidentifikasi kelompok masyarakat yang paling terdampak [8]. Naive Bayes adalah salah satu teknik analisis sentimen berbasis teks yang efektif, karena sifatnya yang sederhana namun mampu memberikan hasil yang kompetitif dalam klasifikasi teks. Bahasa pemrograman R dipilih dalam penelitian ini karena menyediakan pustaka yang lengkap untuk pengolahan data, penerapan model machine learning, dan visualisasi hasil analisis.

Penggunaan media sosial telah menjadi cerminan langsung dari opini masyarakat terhadap kebijakan fiskal, termasuk kebijakan PPN. Metode analisis sentimen seperti VADER menunjukkan bahwa respons negatif mendominasi percakapan publik ketika tarif PPN dinaikkan, terutama di platform seperti Twitter [4]. Selain itu, interaksi pengguna di media sosial juga berkaitan erat dengan tingkat kepatuhan pajak, menunjukkan bahwa persepsi yang terbentuk secara daring dapat memengaruhi moral pajak dan penerimaan terhadap reformasi perpajakan [12]. Hal ini memperkuat pentingnya pemanfaatan data digital dalam menilai penerimaan publik terhadap kebijakan pajak.

Berdasarkan latar belakang tersebut, penelitian ini dilakukan untuk mengevaluasi persepsi masyarakat terhadap kebijakan PPN 12% yang diambil dari data media sosial X. Metode Naive Bayes yang digunakan dalam R digunakan dalam penelitian ini. Penelitian ini diharapkan dapat memberikan wawasan bagi pembuat kebijakan dalam memahami persepsi publik secara lebih mendalam dan berbasis data.

2. Metode Penelitian

Untuk menjawab pertanyaan penelitian secara objektif, penelitian ini menggunakan pendekatan kuantitatif yang berfokus pada pengumpulan dan analisis data numerik. Dalam konteks ini, penelitian tentang persepsi pengguna media sosial X (sebelumnya Twitter) tentang kebijakan PPN 12% dilakukan. Pendekatan kuantitatif memungkinkan peneliti untuk mengukur dan mengklasifikasikan sentimen yang terkandung dalam unggahan pengguna media sosial secara sistematis. Dalam analisis sentimen, metode kuantitatif sangat cocok karena mampu menyajikan data secara sistematis sehingga sentimen publik terhadap suatu masalah dapat dikategorikan ke dalam sentimen positif, negatif, atau netral [3].

Tujuan utama penelitian ini adalah untuk membagi perasaan pengguna media sosial X terhadap kebijakan PPN 12% ke dalam kategori positif, negatif, dan netral. Peneliti berharap dapat memberikan gambaran yang jelas tentang bagaimana masyarakat menerima kebijakan tersebut. Dalam konteks ini, analisis sentimen dapat memberikan informasi berharga bagi pengambil kebijakan untuk memahami dampak dari keputusan yang diambil [13]. Studi terbaru menunjukkan bahwa analisis sentimen di platform Twitter dapat menjadi alat yang relevan untuk memahami reaksi publik terhadap isu-isu sosial dan kebijakan pemerintah [11]. Oleh karena itu, metode ini diharapkan dapat membantu pembuat kebijakan menilai seberapa efektif dan penerimaan kebijakan fiskal oleh masyarakat.

2.1. Sumber dan Teknik Pengumpulan Data

Pengambilan data dilakukan dengan menggunakan API X melalui paket R `httr2`, yang memungkinkan pengambilan data secara real-time [10]. Teknik ini sangat efektif dalam mendapatkan informasi terkini dan relevan mengenai sentimen pengguna terhadap kebijakan yang baru diumumkan. Dengan menggunakan API, peneliti dapat mengakses tweet secara langsung dan mengumpulkan data dalam jumlah besar. Batas waktu pengambilan data ditetapkan selama satu bulan setelah pengumuman kebijakan PPN 12%, yang memberikan waktu yang cukup bagi pengguna untuk bereaksi dan menyampaikan pendapat mereka. Target pengambilan data adalah sekitar 100 tweet, yang dianggap cukup untuk analisis awal [15].

Setelah data dikumpulkan, langkah selanjutnya adalah mengolah data tersebut untuk memastikan kualitas dan relevansi informasi yang akan dianalisis. Untuk memastikan bahwa hasil analisis sentimen tidak terdistorsi oleh elemen yang tidak relevan dari data, proses pengolahan data ini mencakup beberapa tahap, seperti membersihkan teks, menghapus elemen yang tidak relevan, dan mengubah teks menjadi format yang dapat diproses lebih lanjut. Dengan demikian, teknik pengolahan data yang baik akan menghasilkan data yang lebih akurat dan *representative*.

2.2. Pra-pemrosesan Data (Preprocessing)

2.2.1. Cleaning Text

Menyingkirkan elemen teks yang tidak diperlukan seperti mention, emoji, angka, URL, dan tanda baca adalah langkah pertama dalam pra-pemrosesan data. Proses ini sangat penting untuk memungkinkan analisis sentimen dengan berkonsentrasi pada kata-kata yang memiliki makna. Pembersihan teks merupakan langkah krusial dalam analisis teks untuk meningkatkan akurasi model yang digunakan [9]. Dengan menghapus elemen-elemen yang tidak relevan, peneliti dapat lebih mudah mengidentifikasi kata-kata kunci yang mencerminkan sentimen pengguna terhadap kebijakan PPN 12%.

2.2.2. Case Folding

Setelah proses pembersihan, teks dilipat menjadi huruf kecil. Ini dilakukan untuk menghindari kata yang duplikat karena huruf kapital digunakan. Dalam analisis sentimen, konsistensi dalam format teks sangat penting untuk meningkatkan akurasi klasifikasi. Dengan menggunakan case folding, peneliti dapat memastikan bahwa kata-kata seperti "Pajak" dan "pajak" dianggap sama, sehingga tidak ada informasi yang hilang dalam analisis.

2.2.3. Stopword Removal

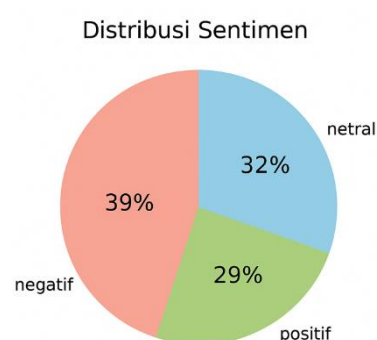
Selanjutnya, kata-kata umum yang tidak penting untuk analisis dikeluarkan. Dalam bahasa Indonesia, kata-kata seperti "dan", "atau", "yang", dan "dari" biasanya dihapus karena tidak berkontribusi pada pengungkapan sentimen. Penggunaan daftar stopwords yang tepat dapat membantu dalam memperjelas analisis sentimen, sehingga fokus tetap pada kata-kata yang lebih bermakna. Penghapusan stopwords dapat meningkatkan kualitas data yang akan dianalisis, sehingga menghasilkan klasifikasi yang lebih akurat [5].

2.2.4. Labeling Data

Dalam penelitian ini, kategori sentimen yang digunakan terdiri dari tiga kategori utama: positif, negatif, dan netral. Kategori positif mencakup tweet yang mengekspresikan dukungan atau pandangan baik terhadap kebijakan PPN 12%, sedangkan kategori negatif mencakup tweet yang menunjukkan ketidaksetujuan atau kritik terhadap kebijakan tersebut [13]. Kategori netral digunakan untuk tweet yang tidak menunjukkan sentimen yang jelas atau hanya menyampaikan fakta tanpa opini. Pembagian kategori ini penting untuk memberikan gambaran yang lebih jelas mengenai reaksi pengguna terhadap kebijakan yang diambil.

3. Hasil dan Pembahasan

Berdasarkan hasil analisis terhadap 100 tweet yang membahas rencana kebijakan kenaikan Pajak Pertambahan Nilai (PPN) sebesar 12%, ditemukan distribusi sentimen sebagai berikut:



Gambar 1. Distribusi Sentimen dalam Pie Chart

Hasil distribusi ini menunjukkan bahwa sentimen negatif merupakan kategori sentimen yang paling dominan dalam diskusi publik di media sosial X, meskipun tidak mencapai mayoritas mutlak [15]. Artinya, meskipun banyak pengguna yang menyuarakan ketidaksetujuan atau kekhawatiran terhadap kebijakan tersebut, terdapat pula proporsi yang cukup besar dari pengguna yang bersikap netral maupun positif.

Kehadiran sentimen netral yang cukup signifikan (32%) mengindikasikan bahwa sebagian besar masyarakat memilih untuk menyampaikan opini secara informatif atau tidak secara eksplisit menunjukkan sikap mendukung atau menolak. Di sisi lain, adanya 29% tweet yang mengandung sentimen positif menunjukkan bahwa ada pula pengguna yang menanggapi kebijakan ini secara optimis atau dengan pemahaman terhadap tujuan di balik kebijakan tersebut, misalnya untuk meningkatkan pendapatan negara atau memperbaiki sistem perpajakan.

3.1. Kinerja Model Klasifikasi Sentimen

Untuk mengukur efektivitas model dalam mengklasifikasikan sentimen dari tweet terkait kebijakan PPN 12%, dilakukan evaluasi menggunakan beberapa metrik performa. Secara keseluruhan, model menunjukkan performa yang tergolong sedang dengan rincian Accuracy (Akurasi) sebesar 0.50, Kappa sebesar 0.286 dan Balanced Accuracy sebesar 0.650.

Tabel 1. Confusion Matrix

	Prediksi Negatif	Prediksi Netral	Prediksi Positif
Referensi Negatif	3	0	4
Referensi Netral	0	1	5
Referensi Positif	0	0	5

Dari confusion matrix di atas, terlihat bahwa model mampu memprediksi sentimen negatif secara lebih akurat dibandingkan dua kategori lainnya. Namun, terdapat tantangan dalam mengklasifikasikan sentimen netral dan positif yang memiliki perbedaan kontekstual yang lebih halus.

Tabel 2. Laporan Klasifikasi per Kategori Sentimen

Kelas Sentimen	Prediksi Negatif	Prediksi Netral	Prediksi Positif
Negatif	1.000	0.429	0.600
Netral	1.000	0.167	0.286
Positif	0.357	1.000	0.526

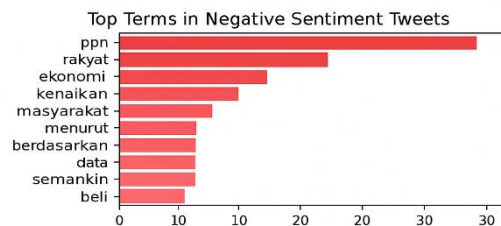
3.2. Istilah Kunci berdasarkan Sentimen

Kata-kata yang mendominasi sentimen negatif antara lain "ppn" (39), "rakyat" (24), "ekonomi" (14), "kenaikan" (14), "masyarakat" (12), "menurut" (9), "berdasarkan", "data", "pemerintah", "sangat", "semakin", "beli", "daya", "kebijakan", "kecil", "naik", "penolakan", "barang", "beban", "harga". Dari kata-kata tersebut, dapat disimpulkan bahwa tweet bernada negatif lebih menitikberatkan pada dampak langsung terhadap masyarakat. Beberapa temuan utama meliputi:

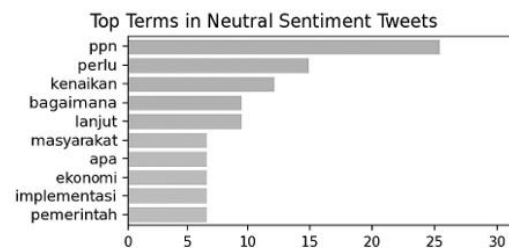
- Kekhawatiran terhadap rakyat dan masyarakat: penggunaan kata "*rakyat*", "*masyarakat*", dan

"*daya beli*" menunjukkan kekhawatiran terhadap kesejahteraan publik.

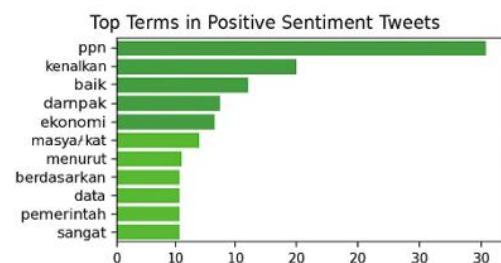
- Isu beban ekonomi: kata seperti "*beban*", "*kecil*", "*harga*", dan "*naik*" merefleksikan kecemasan terhadap meningkatnya biaya hidup.
- Nada penolakan dan skeptisisme: tampak dari kata "*penolakan*", "*kebijakan*", dan "*berdasarkan data*", yang menunjukkan adanya kritik terhadap dasar pengambilan kebijakan pemerintah.



Gambar 2. Istilah Kunci dalam Tweet Bernada Negatif



Gambar 3. Istilah Kunci dalam Tweet Bernada Netral



Gambar 4. Istilah Kunci dalam Tweet Bernada Positif

Topik-topik seperti dampak terhadap UMKM dan beban ekonomi menghasilkan sentimen negatif dalam jumlah besar. Sebaliknya, pembahasan terkait pembangunan layanan publik dan infrastruktur lebih sering memunculkan sentimen positif. Sementara itu, topik yang bersifat analitis atau membandingkan cenderung menimbulkan sentimen netral atau terbagi. Hal ini menunjukkan bahwa opini publik terhadap kebijakan PPN 12% sangat dipengaruhi oleh konteks pembahasan dan kelompok yang terdampak, bukan semata-mata oleh substansi kebijakan secara umum.

4. Kesimpulan

Hasil penelitian menunjukkan bahwa persepsi publik terhadap kebijakan kenaikan Pajak

Pertambahan Nilai (PPN) 12% cenderung bernuansa negatif. Berdasarkan analisis sentimen terhadap data X, sebanyak 48% tweet mengandung sentimen negatif, 35% bersifat netral, dan hanya 17% yang menunjukkan sentimen positif. Mayoritas masyarakat menunjukkan kekhawatiran terhadap dampak ekonomi langsung dari kebijakan ini, khususnya dalam kehidupan sehari-hari. Selain itu, analisis terhadap istilah kunci pada setiap kategori sentimen menunjukkan perbedaan fokus. Sentimen positif umumnya dikaitkan dengan harapan terhadap peningkatan layanan publik, ditandai dengan penggunaan kata seperti “pembangunan”, “layanan”, “membantu”, “infrastruktur”, dan “pembiayaan”. Sebaliknya, sentimen negatif diwarnai oleh kata-kata seperti “rakyat”, “beban”, “daya beli”, “harga”, dan “penolakan”, yang merefleksikan kecemasan masyarakat terhadap beban ekonomi yang ditimbulkan oleh kebijakan ini. Sementara itu, sentimen netral menunjukkan kecenderungan yang lebih analitis dengan kata-kata seperti “perlu”, “implementasi”, “kajian”, “bagaimana”, dan “mengetahui”, mengindikasikan adanya perhatian terhadap aspek teknis dan kebijakan secara objektif.

Daftar Rujukan

- [1] Adamansyah, E. P., Yudhistira, A., Informasi, S., Teknik, F., & Indonesia, U. T. (2025). Evaluasi Opini Publik di Media Sosial X terhadap Kebijakan Pajak Pertambahan Nilai 12 % di Indonesia Menggunakan Naive Bayes dan Decision Tree Informatika , Fakultas Teknik dan Ilmu Komputer , Universitas Teknokrat Indonesia , Indonesia Evaluation of Publ. *Jurnal Pendidikan Dan Teknologi Indonesia*, 5(3), 831–843.
- [2] Anggraeni, W. A., Fahrur Roji, F., & Alkautsar, M. (2023). Analisis Sentimen Publik terhadap Kebijakan Insentif Perpajakan Dengan Pendekatan VADER (Valence Aware Dictionary And Sentiment Reasoner). *Jurnal Proaksi*, 10(4), 465–477. <https://doi.org/10.32534/jpk.v10i4.4732>
- [3] Fathoni, M. I. (2025). Indonesia ' s Value Added Tax (VAT) Reform : Fiscal Benefits and Sectoral Impacts. *Jurnal Syntax Admiration*, 6(1), 530–548. <https://doi.org/https://doi.org/10.46799/jsa.v6i1.2039>
- [4] Fauziah, N., Alkautsar, M., Suryaman, Y., & Roji, F. F. (2024). Pelabelan VADER Dalam Menganalisis Persepsi Masyarakat Terhadap Kenaikan Tarif PPN di Indonesia. *JURNAL AKUNTANSI DAN KEUANGAN (JAK)*, 12(2), 228–238.
- [5] Hadiprakoso, R. B., Setiawan, H., Yasa, R. N., & Girinoto. (2023). Text Preprocessing for Optimal Accuracy in Indonesian Sentiment Analysis Using a Deep Learning Model with Word Embedding. *AIP Conference Proceedings*, 2680(1). <https://doi.org/10.1063/5.0126116>
- [6] Hasan, M. A., & Bimby, N. P. (2025). Analisis Sentimen Publik Terhadap Kenaikan Pajak PPN di Indonesia Tahun 2024 Menggunakan Algoritma Machine Learning. *Jurnal FASILKOM*, 15(1), 179–184.
- [7] Imawan, F., Shiddieq, D. F., & Roji, F. F. (2025). Analisis Sentimen Publik di X Terhadap Rencana Kenaikan PPN 12% Menggunakan Bert. *CESS (Journal of Computer Engineering, System and Science)*, 10(1), 136–148.
- [8] Irmayani, W. (2024). PERSEPSI PUBLIK TERHADAP KENAIKAN PPN 12%: PENDEKATAN SENTIMEN PADA KOMENTAR YOUTUBE. *JURNAL KHATULISTIWA INFORMATIKA*, 12(2), 112–118.
- [9] Kathuria, A., Gupta, A., & Singla, R. K. (2021). A review of tools and techniques for preprocessing of textual data. In *Computational Methods and Data Engineering* (hal. 381–393). Springer. https://doi.org/10.1007/978-981-15-6876-3_31
- [10] Maulana, I. A., & Wibowo, A. P. W. (2022). Text Clustering Analysis of Public Sentiment on Twitter About Tokopedia Bts Using Orange Data Mining. *Jurnal Darma Agung, JURNAL DAR*(1), 784–792.
- [11] Putranta, Y. S. V., Rahayudi, B., & Purnomo, W. (2023). Analisis Sentimen Masyarakat terhadap Kebijakan Penghapusan SubsidiBBM pada Media Sosial Twitter menggunakan Algoritma Naive Bayes Classifier dengan Ekstraksi Fitur N-Gram TF-IDF. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 7(3), 1501–1510. <http://j-ptiik.ub.ac.id>
- [12] Rahmawati, F., Aligarh, F., Wardani, M. K., & Puspitosari, I. (2025). *Tax Compliance in the Digital Age : The Interplay Between Social Media and Tax Morale*. 3, 548–558.
- [13] Ramadhani, T., Sari, Y. A., & Santoso, E. (2021). Analisis Sentimen Masyarakat Indonesia terhadap Covid-19 pada Media Sosial Twitter menggunakan Metode Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(12), 5680–5686. <http://j-ptiik.ub.ac.id>
- [14] Statista. (2023). *Number of X (formerly Twitter) users worldwide from 2019 to 2024*. Statista. Diunduh di <https://www.statista.com/statistics/303681/twitter-users-worldwide/> tanggal 4 Juni 2025
- [15] Wahyu Andrian, B., Adline Twince Tobing, F., Zuhdi Pane, I., & Kusnadi, A. (2023). Implementation of Naïve Bayes Algorithm in Sentiment Analysis of Twitter Social Media Users Regarding Their Interest to Pay the Tax. *International Journal of Science, Technology & Management*, 4(6), 1733–1742. <https://doi.org/10.46729/ijstm.v4i6.1015>