

**KOMPARASI ALGORITMA RANDOM FOREST, NAÏVE BAYES, DAN
SVM PADA SENTIMEN KEBIJAKAN PPN 12%**

***COMPARISON OF RANDOM FOREST, NAÏVE BAYES, AND SVM
ALGORITHMS IN SENTIMENT ANALYSIS
OF THE 12% VAT POLICY***

Skripsi



Disusun oleh

Dandi Purnomo

D0221078

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS SULAWESI BARAT
2025**

LEMBAR PERSETUJUAN

Skripsi

**KOMPARASI ALGORITMA RANDOM FOREST, NAÏVE BAYES, DAN
SVM PADA SENTIMEN KEBIJAKAN PPN 12%**

Telah dipersiapkan dan disusun oleh

Dandi Purnomo

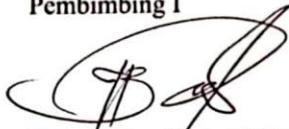
D0221078

Telah dipertahankan di depan Tim Penguji

Pada tanggal 28 Mei 2025

Susunan Tim Penguji

Pembimbing I



Nahya Nur, S.T., M.Kom
NIP. 199111052019032024

Penguji I



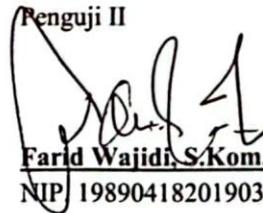
Dr. Eng. Sulfavanti, S.Si., M.T
NIP. 198903172020122011

Pembimbing II



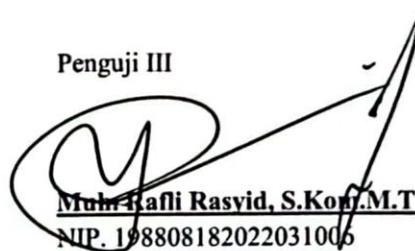
Wawan Firgiawan, S.T., M.Kom.
NIDK. 8948080023

Penguji II



Farid Wajidi, S.Kom., MT
NIP. 198904182019031018

Penguji III



Muh. Rafli Rasvid, S.Kom., M.T
NIP. 198808182022031006

HALAMAN PENGESAHAN

KOMPARASI ALGORITMA RANDOM FOREST, NAÏVE BAYES, DAN SVM PADA SENTIMEN KEBIJAKAN PPN 12%

SKRIPSI

Yang dipersiapkan dan disusun Oleh

DANDI PURNOMO
D0221078

Telah disetujui oleh dosen pembimbing Skripsi
Pada tanggal 28 Mei 2025

Dosen Pembimbing I,

Dosen Pembimbing II,



Nahya Nur, S.T., M.Kom
NIP. 199111052019032024



Wawan Firgiawan, S.T., M.Kom.
NIDK. 8948080023

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer (S. Kom)
Tanggal 28 Mei 2025
Mengetahui,

Dekan Fakultas Teknik,
Universitas Sulawesi Barat

Ketua Program Studi
Informatika



Prof. Dr. Ir. Hafsah Nirwana, M.T.
NIP. 196404051990032002



Muh. Rafli Rasyid, S.Kom.M.T.
NIP. 198808182022031006

HALAMAN PERNYATAAN

Dengan ini saya menyatakan bahwa skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar Strata 1 (S1) Perguruan Tinggi Universitas Sulawesi Barat dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Wonomulyo, 05 Mei 2025

Yang membuat pernyataan



Dandi Purnomo

D0221078

BAB I

PENDAHULUAN

A. Latar Belakang

Perkembangan internet telah mengubah cara orang berkomunikasi, dengan banyaknya platform media sosial yang memungkinkan komunikasi daring (Fachriza & Munawar, 2023). Pertumbuhan pesat media sosial dalam beberapa tahun terakhir menjadikan internet sebagai sarana yang efektif untuk menyebarkan pengetahuan. Penggunaan *big data* dari media sosial dapat dimanfaatkan untuk memprediksi kejadian di dunia nyata (Sekioka et al., 2023). Salah satu platform populer di Indonesia adalah Twitter, yang didirikan oleh Jack Dorsey dan digunakan untuk mengirim pesan singkat atau *tweet* (Krisdiyanto et al., 2021). Sebagai platform yang dinamis dengan jumlah pengguna yang cepat meningkat, Twitter menjadi sumber data penting untuk memahami perspektif dan emosi penggunanya (Azrul et al., 2024). Sementara itu, Pajak Pertambahan Nilai (PPN) adalah pajak yang dikenakan pada transaksi yang melibatkan barang dan jasa, baik domestik maupun internasional (Yani et al., 2024). Berdasarkan Undang-Undang Nomor 7 Tahun 2021 tentang Harmonisasi Peraturan Perpajakan, pemerintah berencana menaikkan tarif PPN dari 11% menjadi 12% mulai 1 Januari 2025. Kenaikan ini diharapkan dapat meningkatkan pendapatan negara guna mendukung pemulihan ekonomi dan pembangunan (Putri, 2024). Namun, rencana tersebut menuai berbagai reaksi publik, termasuk gelombang opini di media sosial seperti Twitter. Mengingat kebijakan ini mulai diberlakukan pada awal tahun 2025, penelitian ini dilaksanakan dalam kurun waktu yang berdekatan dengan penerapannya, sehingga opini publik yang dianalisis mencerminkan respons aktual yang masih relevan.

Guna memperoleh pemahaman mendalam tentang respon publik terhadap suatu kebijakan, analisis sentimen melalui platform media sosial menjadi krusial. Analisis sentimen adalah proses yang mengolah, mengekstrak, dan memahami teks yang tidak terstruktur dengan tujuan menemukan informasi tentang perasaan yang ada dalam kalimat opini (Elisa & Isnain, 2024). Dengan memanfaatkan teknik

Natural Language Processing (NLP), *data mining*, dan *machine learning*, analisis sentimen memungkinkan evaluasi opini masyarakat secara menyeluruh dan *real-time*, yang memberikan informasi berharga bagi para pembuat kebijakan. *Natural Language Processing* atau Pengolahan Bahasa Alami telah berkembang menjadi ranah yang memikat dalam dunia komputasi dan kecerdasan buatan, yang membolehkan mesin untuk menginterpretasikan dan menganalisis teks manusia dalam beragam bahasa (Amien, 2023). Sehingga teks dapat dimasukkan ke dalam kategori positif, netral, atau negatif melalui analisis sentimen. Oleh karena itu, penggunaan algoritma klasifikasi yang akurat penting untuk menjamin efektivitas analisis sentimen.

B. Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini adalah bagaimana perbandingan kinerja tiga algoritma *machine learning* yaitu, Random Forest, Naïve Bayes, dan Support Vector Machine (SVM), dalam mengklasifikasikan sentimen publik terhadap kebijakan kenaikan PPN 12% berdasarkan data dari media sosial Twitter?

C. Batasan Masalah

Batasan masalah dalam penelitian ini adalah:

1. Data yang digunakan dalam penelitian ini hanya berasal dari media sosial Twitter dan terbatas pada *tweet* yang secara spesifik membahas topik PPN 12%.
2. Penelitian ini hanya membandingkan tiga algoritma *machine learning*, yaitu Random Forest, Naïve Bayes, dan Support Vector Machine (SVM), tanpa mempertimbangkan algoritma lain.

D. Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Membandingkan kinerja tiga algoritma *machine learning*, yaitu Random

Forest, Naïve Bayes, dan Support Vector Machine (SVM), dalam melakukan analisis sentimen terhadap tanggapan publik mengenai kebijakan kenaikan PPN sebesar 12% di Indonesia

2. Menilai performa masing-masing algoritma berdasarkan metrik evaluasi seperti akurasi, *precision*, *recall*, dan *f1-score* untuk mengetahui algoritma yang paling optimal dalam mengklasifikasikan sentimen dari data media sosial Twitter.
3. Memberikan rekomendasi algoritma yang paling efektif dan efisien dalam pemrosesan bahasa alami untuk analisis sentimen, khususnya terhadap data berteks pendek seperti *tweet*.

E. Manfaat Penelitian

Penelitian ini diharapkan bermanfaat sebagai berikut:

1. Penelitian ini dapat memberikan kontribusi dalam pengembangan dan penerapan algoritma *machine learning*, khususnya di bidang pemrosesan bahasa alami (*Natural Language Processing*) untuk data teks pendek seperti *tweet*.
2. Hasil analisis sentimen ini dapat menjadi masukan dalam memahami persepsi dan respons publik terhadap kebijakan kenaikan PPN 12%, sehingga dapat dijadikan bahan pertimbangan dalam proses pengambilan keputusan yang lebih responsif terhadap opini masyarakat.

BAB V

PENUTUP

A. Kesimpulan

Penelitian ini bertujuan untuk membandingkan kinerja tiga algoritma klasifikasi yaitu Random Forest, Naïve Bayes, dan Support Vector Machine (SVM), dalam menilai persepsi masyarakat terhadap kebijakan PPN 12% menggunakan data yang diperoleh dari platform Twitter. *Dataset* yang digunakan mencakup 2747 *tweet* yang diperoleh melalui proses *crawling* menggunakan kata kunci “PPN 12%” pada rentang waktu April 2024 hingga Januari 2025. Setelah melalui tahapan *preprocessing* dan pelabelan, data dibagi menjadi 80% data latih dan 20% data uji secara acak untuk menjaga distribusi sentimen tetap seimbang. Jumlah data pelatihan sebanyak 2154, dan data uji sebanyak 539, dengan distribusi sentimen meliputi 689 data positif, 544 data netral, dan 1460 data negatif.

Hasil komparasi menunjukkan bahwa algoritma SVM memberikan performa paling tinggi dengan akurasi sebesar 76,25%, diikuti oleh Random Forest sebesar 72,35%, dan Naïve Bayes sebesar 60,66%. Karena *dataset* yang digunakan mengalami ketidakseimbangan kelas, penelitian ini juga mengukur kinerja dengan menggunakan *precision*, *recall*, dan *f1-score* untuk meningkatkan hasil evaluasi. Metrik ini diambil berdasarkan *weighted average* untuk mempertimbangkan distribusi masing-masing kelas secara proporsional.

Secara keseluruhan, SVM tidak hanya unggul dalam akurasi tetapi juga memiliki nilai *precision*, *recall*, dan *f1-score* tertinggi dibandingkan dua algoritma lainnya. Berdasarkan hasil perbandingan akurasi, *precision*, *recall*, dan *f1-score*, dapat disimpulkan bahwa SVM merupakan model dengan performa terbaik dalam analisis sentimen kebijakan PPN 12%. SVM unggul karena mampu mengklasifikasikan data teks yang kompleks dengan memisahkan kelas menggunakan batas (*hyperplane*) yang optimal. Dengan bantuan *kernel linear*, SVM dapat bekerja dengan baik pada data berdimensi tinggi tanpa perlu transformasi yang rumit. Selain itu, karena hanya memproses data yang dekat dengan batas kelas, SVM lebih tahan terhadap *overfitting* dan cocok untuk data

yang tidak seimbang, seperti pada analisis sentimen yang sering mengandung konteks yang samar. Kemampuan ini memungkinkan SVM menghasilkan prediksi yang lebih tepat dan seimbang dibandingkan kedua algoritma lainnya.

B. Saran

Namun demikian, penelitian ini juga memiliki keterbatasan, terutama dalam aspek pelabelan data yang memiliki potensi bias, termasuk dalam mengenali sarkasme. Selain itu, model yang digunakan belum dioptimasi melalui proses *hyperparameter tuning*, sehingga masih terdapat ruang untuk peningkatan performa.

Penelitian selanjutnya disarankan untuk menggunakan pendekatan pelabelan data yang lebih akurat untuk meminimalkan potensi bias, menyertakan teknik optimasi parameter, menggunakan teknik *oversampling* atau *undersampling*, serta mengeksplorasi algoritma *deep learning* yang berpotensi memberikan hasil klasifikasi yang lebih akurat dalam menangani data teks yang bersifat dinamis dan kontekstual seperti Twitter.

DAFTAR PUSTAKA

- Amien, M. (2023). *Sejarah dan Perkembangan Teknik Natural Language Processing (NLP) Bahasa Indonesia: Tinjauan tentang sejarah, perkembangan teknologi, dan aplikasi NLP dalam bahasa Indonesia*.
- Azrul, A., Purnamasari, A. I., & Ali, I. (2024). Analisis Sentimen Pengguna Twitter Terhadap Perkembangan Artificial Intelligence dengan Penerapan Algoritma Long Short-Term Memory (LSTM). In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 1).
- Zendrato, A. D., Berutu, S. S., Sumihar, Y. P., & Budiati, H. (2024). Pengembangan Model Klasifikasi Sentimen Dengan Pendekatan Vader dan Algoritma Naive Bayes Terhadap Ulasan Aplikasi Indodax. *Journal of Information System Research (JOSH)*, 5(3), 755–764. <https://doi.org/10.47065/josh.v5i3.5050>
- Eldo, H., Ayuliana, A., Suryadi, D., Chrisnawati, G., & Judijanto, L. (2024). Penggunaan Algoritma Support Vector Machine (SVM) Untuk Deteksi Penipuan pada Transaksi Online. *Jurnal Minfo Polgan*, 13(2), 1627–1632. <https://doi.org/10.33395/jmp.v13i2.14186>
- Elisa, P., & Isnain, A. R. (2024). Comparison Of Random Forest, Support Vector Machine and Naive Bayes Algorithms to Analyze Sentiment Towards Mental Health Stigma. *Jurnal Teknik Informatika (JUTIF)*, 5(1), 321–329. <https://doi.org/10.52436/1.jutif.2024.5.1.1817>
- Fachriza, M., & Munawar. (2023). *Analisis Sentimen Kalimat Depresi Pada Pengguna Twitter Dengan Naive Bayes, Support Vector Machine, Random Forest*. <http://studentjournal.umpo.ac.id/index.php/komputek>
- Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *SMATIKA JURNAL*, 10(02), 71–76. <https://doi.org/10.32664/smatika.v10i02.455>
- Gemasih, H., & Nizar, I. (2022). *Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid*.
- Irawan, I., Wardianto, Wathan, M. H., & Prayogi, M. B. (2024). *Studi Perbandingan: Algoritma Random Forest, Naive Bayes dan Support Vector Machine Dalam Analisis Sentimen Pada Aplikasi Capcut di Google Play Store*. Vol. 5.
- Krisdiyanto, T., Maricha, E., & Nurharyanto, O. (2021). Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers. *Jurnal CoreIT*, 7(1).
- Kusuma, M. R., Gata, W., Kurniawati, L., Rivian, A. F., & Kurniawan, T. (2023). Analisis sentimen masyarakat terhadap informasi penerapan PPN atas renovasi dan membangun rumah sendiri pada media sosial youtube dengan metode svm dan naive bayes. *Jurnal Digital Teknologi Informasi*, 6(2). <https://doi.org/10.32502/digital.v6i2.4758>
- Putri, I. M. (2024). *Kenaikan PPN 12% dan Dampaknya Terhadap Ekonomi*. 8(2).

- Lauwl, C. M., Husain, Nuzululnisa, B. N., & Wijaya, H. (2025). Komparasi Metode Random Forest Dan Support Vector Machine (SVM) Untuk Pemodelan Klasifikasi Serangan DDos. *Journal of Information System Research*, 6(2), 1471–1478. <https://doi.org/10.47065/josh.v6i2.6684>
- Miftahusalam, A., Febby Nuraini, A., Khoirunisa, A. A., & Pratiwi, H. (2022). *Perbandingan Algoritma Random Forest, Naïve Bayes, dan Support Vector Machine Pada Analisis Sentimen Twitter Mengenai Opini Masyarakat Terhadap Penghapusan Tenaga Honorer*.
- Misriati, T., & Aryanti, R. (2024). Optimalisasi Random Forest dan Support Vector Machine dengan Hyperparameter GridSearchCV untuk Analisis Sentimen Ulasan PrimaKu. *Journal of Information System Research (JOSH)*, 5(4), 1333–1341. <https://doi.org/10.47065/josh.v5i4.5347>
- Natalianus, G. K. V., & Feta, R. (2024). Perbandingan Akurasi Algoritma Naïve Bayes Dan Support Vector Machine Dalam Analisis Sentimen Pengguna Twitter Terhadap Aplikasi Sirekap. *Fountain of Informatics Journal*, 9(2), 2548–5113. <https://doi.org/10.21111/fij.v9i2.12717>
- Nurhidayat, R., & Dewi, K. E. (2023). *KOMPUTA : Jurnal Ilmiah Komputer dan Informatika Penerapan Algoritma K-Nearest Neighbor dan Fitur Ekstraksi N-Gram Dalam Analisis Sentimen Berbasis Aspek*. 12(1). <https://www.kaggle.com/datasets/hafidahmusthaanah/skincare-review?select=00.+Review.csv>.
- Rifaldi, R., Indra, J., Rizky Pratama, A., Ratna Juwita, A., Karawang, P., Jl Ronggo Waluyo, K. H., Timur, T., Barat, J., Kunci, K., Naïve Bayes, A., & Produk, P. (2024). Analisis Sentimen Pemboikotan Produk dengan Pendekatan Algoritma Naïve Bayes Media Sosial X. *Journal of Information System Research (JOSH)*, 5(4), 940–946. <https://doi.org/10.47065/josh.v5i4.5420>
- Sekioka, S., Hatano, R., & Nishiyama, H. (2023). Market prediction using machine learning based on social media specific features. *Artificial Life and Robotics*, 28(2), 410–417. <https://doi.org/10.1007/s10015-023-00857-z>
- Syahidin, A. A., Alwi, M. R., Subandi, & Hariyadi, M. R. (2023). Analisis Sentimen Data Twitter Terhadap Pelaksanaan Pembelajaran Online di Indonesia Pada Masa Pandemi Covid-19 Menggunakan Metode Natural Language Processing. *JURSISTEKNI (Jurnal Sistem Informasi Dan Teknologi Informasi)*.
- Tanggraeni, A. I., & Sitokdana, M. N. N. (2022). *Analisis Sentimen Aplikasi E-Government Pada Google Play Menggunakan Algoritma Naïve Bayes*. 9(2), 785–795.
- Torhino, R., & Andono, N. (2024). Penerapan Algoritma Random Forest dalam Prediksi Curah Hujan untuk Mendukung Analisis Cuaca. *Technology and Science (BITS)*, 6(3), 1688–1699. <https://doi.org/10.47065/bits.v6i3.6404>
- Widyanto, T., Ristiana, I., & Wibowo, A. (2023). *Komparasi Naïve Bayes dan SVM Analisis Sentimen RUU Kesehatan di Twitter*. <https://doi.org/10.31598>

Yani, R. E., Simandalahi, E., & Nasution, A. R. (2024). Pengaruh PPN (Pajak Pertambahan Nilai) terhadap Pendapatan Nasional. *Eksis: Jurnal Ilmiah Ekonomi Dan Bisnis*, 15(1), 30. <https://doi.org/10.33087/eksis.v15i1.424>

9. Surat Rekomendasi Penyetaraan Skripsi



**KEMENTERIAN PENDIDIKAN TINGGI, SAINS,
DAN TEKNOLOGI**
UNIVERSITAS SULAWESI BARAT
FAKULTAS TEKNIK
Alamat : Kampus Padhang-Padhang Banggae Timur Majene, Sulawesi Barat
Website://ft.unsulbar.ac.id. Instagram:ft-univeristas sulawesi barat.

SURAT REKOMENDASI PENYETARAAN SKRIPSI
Nomor: 1084/UN55.F4.5/KM/2025

Yang bertanda tangan di bawah ini:

Nama : **Muh Rafli Rasyid, S.Kom., M.T.**
NIP : 198808182022031006
Pangkat/Golongan : Penata Muda Tingkat I/III.b
Jabatan : Koordinator Program Studi Informatika

Dengan ini memberikan rekomendasi penyetaraan skripsi kepada:

Nama : **Dandi Purnomo**
NIM : D0221078
Judul : **Komparasi Algoritma Random Forest, Naive Bayes, Dan SVM pada sentimen kebijakan PPN 12%**
Dosen Pembimbing : 1. Nahya Nur, S.T.,M.Kom.
2. Wawan Firgiawan, S.T.,M.Kom.

Dengan ini menyetujui penyetaraan skripsi yang diajukan oleh mahasiswa tersebut melalui jalur "**Publikasi Nasional**" yang sesuai dengan prinsip dan tujuan publikasi tersebut dan bermanfaat bagi masyarakat, dunia industri, atau pihak-pihak lain yang terkait.

Demikian surat rekomendasi penyetaraan skripsi tersebut untuk dapat digunakan sebagaimana mestinya.

Majene, 28 April 2025

Koordinator Prodi,



Muh Rafli Rasyid, S.Kom., M.T.
NIP. 198808182022031006

Gambar Lampiran 7. Surat Rekomendasi Penyetaraan Skripsi