

## Analisis Sentimen Publik Program PPPK di Media Sosial X menggunakan Naïve Bayes dan SVM

Lisyo Hileria Sarumpaet <sup>1</sup>, Ryan Randy Suryono <sup>2,\*</sup>

<sup>1</sup> Program Studi Sistem Informasi, Universitas Teknokrat Indonesia, Indonesia

<sup>2</sup> Program Studi Ilmu Komputer, Universitas Teknokrat Indonesia, Indonesia

\* Correspondence: ryan@teknokrat.ac.id

**Copyright:** © 2025 by the authors

Received: 21 April 2025 | Revised: 2 Mei 2025 | Accepted: 18 Mei 2025 | Published: 10 Agustus 2025

### Abstrak

Analisis sentimen terhadap Program Pegawai Pemerintah dengan Perjanjian Kerja (PPPK) penting dilakukan untuk memahami persepsi publik dan sebagai dasar evaluasi kebijakan. Penelitian ini bertujuan menganalisis sentimen publik terhadap kebijakan PPPK serta mengevaluasi kinerja algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes* dalam mengklasifikasikan opini masyarakat di media sosial X. Penelitian ini merupakan studi kuantitatif dengan pendekatan data mining. Tahapan diawali dengan pengumpulan data *collection* sebanyak 7.508 tweet dan diproses melalui tahapan *preprocessing*, pelabelan, ekstraksi fitur menggunakan TF-IDF, serta klasifikasi dengan SVM dan Naïve Bayes. Penyeimbangan data dilakukan menggunakan *Synthetic Minority Oversampling Technique* (SMOTE). Hasil temuan kami menunjukkan bahwa SVM menghasilkan akurasi tertinggi sebesar 95%, sementara Naïve Bayes mencapai 87%. Penerapan SMOTE terbukti meningkatkan performa kedua model, khususnya dalam mengenali sentimen negatif. Keunggulan SVM terletak pada kemampuannya memisahkan kelas secara optimal melalui margin maksimum, yang efektif untuk data teks berdimensi tinggi. Sementara itu, SMOTE berperan penting dalam menyeimbangkan distribusi kelas, sehingga meningkatkan *accuracy*, *precision*, dan *recall*. Temuan ini menjadi dasar penting bagi pengambil kebijakan untuk merespons opini publik secara lebih tepat berbasis data yang valid dan representatif.

**Kata kunci:** analisis sentimen; media sosial; *naive bayes*; *pppk*; *support vector machine*

### Abstract

*Sentiment analysis of the Government Employee Program with Work Agreement (PPPK) is important to understand public perception and as a basis for policy evaluation. This study aims to analyze public sentiment towards the PPPK policy and evaluate the performance of the Support Vector Machine (SVM) and Naïve Bayes algorithms in classifying public opinion on social media X. This study is a quantitative study with a data mining approach. The stages begin with collecting data collection of 7,508 tweets and processed through the stages of preprocessing, labeling, feature extraction using TF-IDF, and classification with SVM and Naïve Bayes. Data balancing is done using the Synthetic Minority Oversampling Technique (SMOTE). Our findings show that SVM produces the highest accuracy of 95%, while Naïve Bayes reaches 87%. The application of SMOTE has been shown to improve the performance of both models, especially in recognizing negative sentiment. The advantage of SVM lies in its ability to optimally separate classes through maximum margin, which is effective for high-dimensional text data. Meanwhile, SMOTE plays an important role in balancing class distribution, thereby increasing accuracy, precision, and recall. These findings provide an important basis for policy makers to respond to public opinion more appropriately based on valid and representative data.*

**Keywords:** *sentiment analysis*; *social media*; *naive bayes*; *pppk*; *support vector machine*



## PENDAHULUAN

Perkembangan teknologi digital telah memudahkan masyarakat dalam menyampaikan opini dan tanggapan terhadap berbagai isu publik melalui media sosial (Fitriani, 2023). Salah satu isu yang banyak diperbincangkan adalah kebijakan Program Pegawai Pemerintah dengan Perjanjian Kerja (PPPK) yang direncanakan oleh pemerintah Indonesia (Khairunissabina & Kurniawan, 2024). Melalui media sosial X (sebelumnya Twitter), masyarakat menyampaikan berbagai respon, baik berupa dukungan maupun kritik, yang mencerminkan persepsi publik terhadap pelaksanaan program tersebut (Hidayat & Sugiyono, 2023).

Pelaksanaan Program PPPK telah menimbulkan beragam respon di kalangan masyarakat, khususnya yang terekam melalui media sosial (Hermawan et al., 2025). Persepsi publik terhadap kebijakan ini perlu dianalisis secara sistematis karena dapat menjadi dasar penting dalam evaluasi dan perbaikan kebijakan (Rizqiyah & Nuryana, 2024). Oleh karena itu, analisis sentimen diperlukan untuk menangkap opini masyarakat secara luas dan objektif, terutama karena data media sosial bersifat tidak terstruktur dan sangat dinamis (Anjani et al., 2023).

Penelitian ini menerapkan analisis sentimen berbasis *Natural Language Processing* (NLP) dengan menggunakan dua metode klasifikasi, yaitu *Naïve Bayes* dan *SVM* (Sanjaya et al., 2023). Kedua algoritma ini dipilih karena mampu mengelola karakteristik data media sosial secara efektif serta memberikan klasifikasi sentimen yang akurat (Fitriani et al., 2022). Selain itu, untuk mengatasi ketidakseimbangan kelas dalam data sentimen yang merupakan permasalahan umum pada data media sosial diterapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE)(Khairunnisa et al., 2024). Teknik ini berfungsi untuk meningkatkan representasi kelas minoritas pada data pelatihan, sehingga model dapat melakukan pembelajaran secara lebih merata dan menghasilkan performa klasifikasi yang lebih optimal (Sari & Kurniawan, 2024).

*Naïve Bayes* merupakan algoritma klasifikasi berbasis probabilistik yang efisien dan sederhana, serta cocok untuk memproses teks dalam jumlah besar (Pratama & Maharani, 2025) (Rahayu et al., 2022). Sementara itu, *SVM* memiliki kemampuan mengelompokkan data yang bersifat linier maupun non-linier, sehingga cocok untuk pola data media sosial yang bervariasi (Pamungkas & Cahyono, 2024). Dengan memanfaatkan kedua algoritma ini, diharapkan analisis sentimen dapat memberikan hasil klasifikasi yang akurat dan valid terhadap opini publik mengenai kebijakan PPPK (Sari & Suryono, 2024).

Penelitian sebelumnya yang dilakukan oleh (Karimah et al., 2024), Naive Bayes mencapai *accuracy* sebesar 70,69%, dengan *precision* 43,64% dan *recall* 39,48% dalam menganalisis sentimen masyarakat terhadap mobil listrik melalui konten YouTube. Analisis sentimen menunjukkan bahwa persepsi masyarakat masih cenderung negatif, yang dipengaruhi oleh kurangnya pemahaman terhadap teknologi mobil listrik. Penerapan teknik SMOTE terbukti efektif meningkatkan performa model, di mana akurasi naik dari sebelumnya 50,70% menjadi 70,69%, serta terjadi peningkatan pada nilai *true positive*, *true negative*, dan *true neutral*.

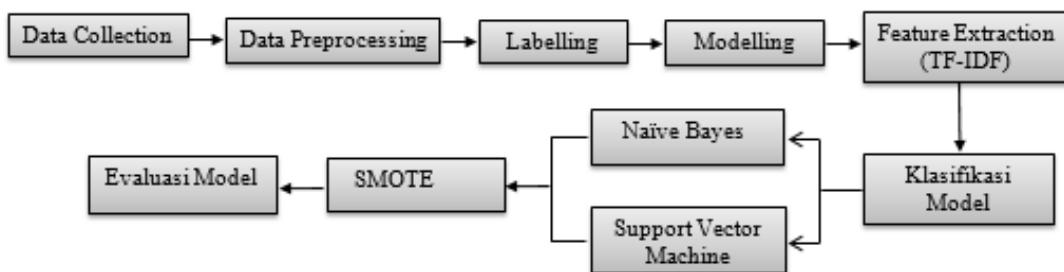
Pada penelitian lain, Algoritma *SVM* dalam analisis sentimen terhadap vaksinasi COVID-19 menunjukkan kinerja yang sangat baik, dengan *accuracy* sebesar 89%, *precision* 88%, dan *recall* 99%. Dari 4.708 tweet yang dianalisis, sebanyak 83,6% menerapkan sentimen positif terhadap program vaksinasi, sedangkan 16,4% menerapkan sentimen negatif. Hal ini mengindikasikan bahwa mayoritas masyarakat Indonesia di Twitter memberikan respons yang positif terhadap kebijakan vaksinasi COVID-19 yang diterapkan pemerintah (Syah & Witanti, 2022). Kedua penelitian tersebut menunjukkan potensi penggunaan algoritma pembelajaran mesin dalam analisis sentimen, tetapi masih memiliki keterbatasan. Penelitian sebelumnya umumnya hanya menggunakan satu algoritma tanpa pembandingan performa dan sebagian besar belum menerapkan teknik *balancing* data seperti SMOTE. Selain itu, topik yang diangkat masih berfokus pada isu-isu umum seperti mobil listrik atau vaksinasi, sementara analisis

sentimen publik terhadap program PPPK sebagai bagian dari kebijakan ketenaga kerjaan nasional masih jarang ditemukan (Nugroho, 2023).

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap program PPPK di media sosial X menggunakan dua algoritma klasifikasi, yaitu *Naïve Bayes* dan *SVM*. Penelitian ini juga menerapkan teknik SMOTE untuk mengatasi ketidakseimbangan data, sehingga model yang dihasilkan dapat memiliki perforsma klasifikasi yang lebih optimal. Diharapkan, penelitian ini dapat memberikan kontribusi nyata dalam memahami persepsi publik terhadap program PPPK serta memperkaya literatur analisis sentimen di Indonesia, khususnya dalam konteks kebijakan publik.

## METODE

Penelitian ini menggunakan pendekatan kuantitatif, dan data diperoleh dari *platform* media sosial X (*Twitter*) dengan kata kunci “PPPK”, “Seleksi PPPK”, dan “Guru PPPK”. Alur penelitian terdiri atas 8 tahapan, yaitu data *collection*, data *preprocessing*, *labelling*, *modelling*, *feature extraction* menggunakan metode TF-IDF, klasifikasi model dengan algoritma *Naïve Bayes* dan *SVM*, penanganan ketidakseimbangan data menggunakan teknik SMOTE, serta evaluasi model.



**Gambar 1.** Tahapan penelitian

Berdasarkan gambar 1, *data collection* dari media sosial X (*Twitter*) menggunakan *library Python Tweepy* dan API v2 selama periode 1 Januari 2024 hingga Februari 2025. Kata kunci yang digunakan meliputi “PPPK”, “seleksi PPPK”, dan “guru PPPK”. Hanya tweet berbahasa Indonesia yang disertakan, dan *filtering* bot dilakukan dengan pendekatan *heuristik* berdasarkan pola nama pengguna dan frekuensi tweet. Total data yang diperoleh sebanyak 7.508 tweet. Tahap *preprocessing* dilakukan untuk membersihkan data sebelum dianalisis. Proses ini meliputi penghapusan tanda baca, emotikon, dan tautan; *case folding*; *stopword removal*; serta *stemming* menggunakan *library NLP Sastrawi* (Setiawan & Suryono, 2024).

Selanjutnya, tweet diberi label sentimen menggunakan pendekatan *lexicon-based*. *Lexicon* yang digunakan menggunakan existing. Keakuratan lexicon merupakan gabungan dari indonesia *sentiment lexicon* (*Inset*). Skor sentimen dihitung berdasarkan penjumlahan nilai kata skor  $> 0$  menunjukkan sentimen positif, dan skor  $\leq 0$  menunjukkan sentimen negatif. SMOTE diterapkan setelah pemisahan data menjadi set data latih untuk menghindari terjadinya *data leakage* yang dapat mempengaruhi akurasi model (Hidayatullah et.al 2023). Model *Naïve Bayes* yang digunakan adalah *Multinomial Naïve Bayes* yang cocok untuk data teks, sementara untuk *SVM*, jenis kernel yang digunakan adalah *linear*. Kemudian, dilakukan transformasi data teks ke bentuk numerik menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF menekankan kata-kata unik dibandingkan kata umum, sehingga lebih efektif dalam klasifikasi teks (Natasha & Suryono, 2025).

Model klasifikasi yang digunakan adalah *Naïve Bayes* dan *SVM*. *Naïve Bayes* merupakan algoritma berbasis probabilitas yang sederhana namun efektif dalam klasifikasi teks. Sementara itu, *SVM* bekerja dengan mencari *hyperplane* terbaik untuk memisahkan data

ke dalam kelas yang berbeda dan dapat menangani data non-linier melalui fungsi kernel. Kedua model dipilih karena performa yang baik dalam analisis sentimen.

Selanjutnya pada tahapan evaluasi model bertujuan untuk mengukur kinerja algoritma klasifikasi dimana diterapkan dalam analisis ini yaitu *Naïve Bayes* dan *SVM* untuk mengkategorikan sentimen publik terhadap Program PPPK di Indonesia. Kinerja model ini dievaluasi menggunakan *Confusion Matrix* yang menghasilkan *Matrix* berupa *accuracy*, *precision*, dan *F1-score*.

## HASIL DAN PEMBAHASAN

### Hasil

Data penelitian, dataset diperoleh dari media sosial X (Twitter) melalui proses *Crawling* menggunakan *Library Tweet Harvest* dengan kata kunci “Program PPPK”, “Seleksi PPPK” dan “Rekrutmen PPPK”. Proses ini berhasil mengumpulkan 7508 *tweet*, yang kemudian disimpan dalam format CSV untuk memudahkan tahap analisis selanjutnya. Pengelolaan data dilakukan menggunakan Google Colab, melalui kode *autentikasi API* dari media sosial X, dan didukung oleh Bahasa pemrograman *Python*. Hasil Tabel 1 menunjukkan bahwa proses *preprocessing* meliputi *Cleaning data* menghapus simbol, tanda baca, dan URL, *case folding* mengubah huruf menjadi kecil, normalisasi kata tidak baku, *tokenizing* memecah kalimat menjadi kata, *stopword removal* menghapus kata tidak penting, dan *stemming* mengubah kata menjadi bentuk dasar. Tahapan ini bertujuan untuk menghasilkan dataset yang bersih dan terstruktur. Hasil akhir menunjukkan bahwa data menjadi lebih terorganisir dan relevan untuk proses analisis sentimen selanjutnya.

**Tabel 1.** Hasil *preprocessing*

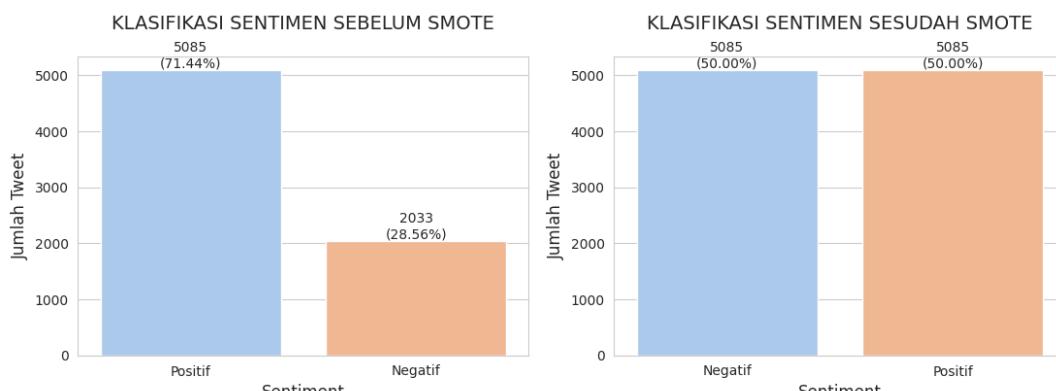
No	Tahapan	Hasil
1	<i>Cleaning data</i>	Abis liat tiktok mbuh iku sopo ngomong anggaran PPPK membebani negara KALIAN yg membebani negara Kalo mo pangkas gaji mending yg atas dulu dipangkas jgn yg bawah
2	<i>Case Folding</i>	abis liat tiktok mbuh iku sopo ngomong anggaran pppk membebani negara kalian yg membebani negara kalo mo pangkas gaji mending yg atas dulu dipangkas jgn yg bawah
3	<i>Normalisasi kata</i>	habis lihat tiktok embuh iku sopo mengomong anggaran pppk membebani negara kalian yang membebani negara kalau mau pangkas gaji mending yang atas dulu dipangkas jangan yang bawah
4	<i>Tokenizing</i>	[habis, lihat, tiktok, embuh, iku, sopo, mengomong, anggaran, pppk, membebani, negara, kalian, yang, membebani, negara, kalau, mau, pangkas, gaji, mending, yang, atas, dulu, dipangkas, jangan, yang, bawah]
5	<i>StopWord</i>	[habis, lihat, tiktok, embuh, iku, sopo, mengomong, anggaran, pppk, membebani, negara, membebani, negara, pangkas, gaji, mending, dipangkas]
6	<i>Stemming</i>	habis lihat tiktok embuh iku sopo omong anggar pppk beban negara beban negara pangkas gaji mending pangkas

Berdasarkan hasil pelabelan, *tweet* pertama yang berisi pembahasan tentang anggaran PPPK diberi label Positif. *Tweet* kedua yang mengkritik pengurangan anggaran dan kebijakan PPPK diberi label Negatif. Pelabelan ini dilakukan untuk menganalisis sentimen publik terhadap program PPPK. Hasil pelabelan ini memberikan gambaran mengenai persepsi masyarakat terhadap kebijakan tersebut yang dapat dilihat pada tabel 2.

**Tabel 2.** Hasil pelabelan

No	<i>Tweet</i>	Pelabelan Data
1	benar benar dagel potong anggar media belanja pegawai cakup gaji tunjang asn pns pppk gaji non asn honorer tenaga ahli dll masuk belanja barang jasa	Positif
2	habis lihat tiktok embuh iku sopo omong anggar pppk beban negara beban negara pangkas gaji mending pangkas	Negatif

Setelah seluruh tahapan penelitian dilaksanakan, jumlah data yang berhasil diperoleh mengalami penurunan menjadi 7118 *tweet*. Dari jumlah tersebut, sebanyak 5085 *tweet* dikategorikan sebagai positif dan 2033 *tweet* sebagai negatif. Ketidakseimbangan antara kedua kategori ini dapat menyebabkan model algoritma, seperti *Naïve Bayes* dan *svm*, lebih terlatih dalam menganalisis pola kata dari kategori sentimen yang lebih dominan. Kondisi ini berpotensi meningkatkan akurasi klasifikasi terhadap sentimen mayoritas. Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan teknik optimasi menggunakan SMOTE untuk menyeimbangkan jumlah data antara kategori sentimen positif dan negatif. Dengan teknik ini, jumlah data pada kategori yang mayoritas dapat seimbangkan, sehingga model tidak terlalu mendominasi satu jenis sentimen tertentu.

**Gambar 2.** Perbandingan klasifikasi sentimen

Berdasarkan hasil yang ditampilkan pada gambar 2, diperoleh perbandingan klasifikasi sentimen sebelum dan sesudah penerapan SMOTE. Sebelum SMOTE diterapkan, terjadi diseimbangkan menjadi 5085 *tweet*. Proses penyeimbangan ini berperan penting dalam mengatasi ketidakseimbangan jumlah data antara sentimen positif dan negatif. Melalui proses penyeimbangan tersebut, jumlah data masing-masing kategori berhasil disamakan 5085 *tweet*. Dengan jumlah data yang lebih seimbang, model klasifikasi dapat bekerja secara optimal dan menghasilkan prediksi secara akurat.

Pada tahap pengujian dalam penelitian ini, dilakukan perbandingan antara algoritma *Naïve Bayes* dan *svm* dengan perbandingan data sebesar 80% untuk pelatihan dan 20% untuk pengujian. Analisis dan evaluasi data dilakukan dalam dua skenario, yaitu sebelum dan setelah penerapan SMOTE, dengan tujuan untuk mengukur efektivitas teknik tersebut dalam meningkatkan kinerja model, pada tabel 3 dan tabel 4 menyajikan hasil klasifikasi *Naïve Bayes* dan *SVM* menggunakan SMOTE dan sebelum SMOTE.

Tabel 3 merupakan hasil perbandingan eksperimen algoritma *Naïve bayes* menunjukkan bahwa penerapan optimasi SMOTE mengalami peningkatan kinerja model, *accuracy* meningkat dari 74% menjadi 87%. *Recall* pada kelas negatif meningkat signifikan dari 8% menjadi 87%, yaitu kenaikan sebesar 79 poin persentase. Ini menunjukkan bahwa model menjadi jauh lebih mampu mengenali sentimen negatif. *Precision* untuk kelas positif juga naik

dari 74% menjadi 87%. Namun, *F1-score* untuk kelas negatif naik dari 14% menjadi 87%, mengindikasikan bahwa distribusi yang lebih seimbang membantu model menangani kelas minoritas secara jauh lebih efektif.

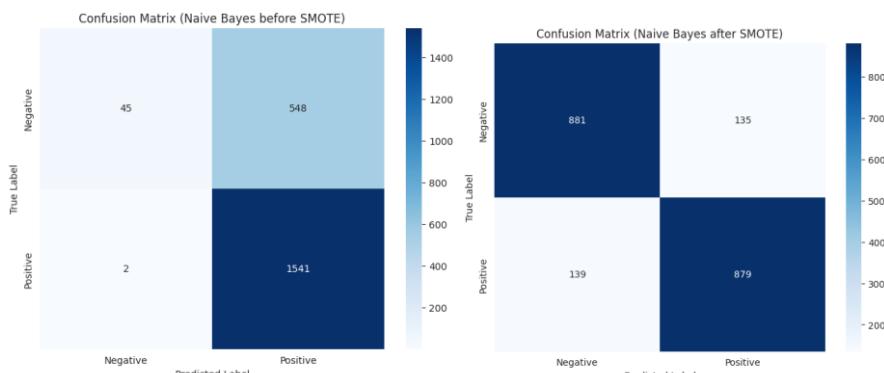
**Tabel 3.** Hasil klasifikasi *naïve bayes*

<b>Report</b>	<b>Sebelum SMOTE</b>		<b>Sesudah SMOTE</b>	
	Positif	Negatif	Positif	Negatif
<i>Accuracy</i>	74%			87%
<i>Precision</i>	74%	96%	87%	86%
<i>Recall</i>	1%	8%	86%	87%
<i>F1-Score</i>	85%	14%	87%	87%

**Tabel 4.** Hasil klasifikasi svm

<b>Report</b>	<b>Sebelum SMOTE</b>		<b>Sesudah SMOTE</b>	
	Positif	Negatif	Positif	Negatif
<i>Accuracy</i>	91%			95%
<i>Precision</i>	92%	89%	98%	93%
<i>Recall</i>	96%	78%	93%	98%
<i>F1-Score</i>	94%	83%	95%	95%

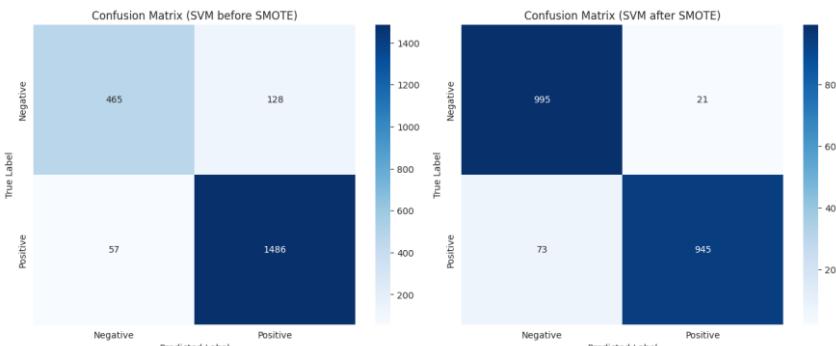
Tabel 4 memperlihatkan bahwa hasil eksperimen pada algoritma SVM mengalami peningkatan kinerja setelah penerapan teknik optimasi SMOTE. *Accuracy* naik dari 91% menjadi 95% setelah SMOTE diterapkan. *Recall* untuk kelas negatif meningkat dari 78% menjadi 98%, yang berarti ada peningkatan sebesar 20 poin persentase. Di sisi lain, *precision* untuk kelas positif meningkat dari 92% menjadi 98%, sedangkan *F1-score* untuk kelas negatif naik dari 83% menjadi 95%. Ini menunjukkan keseimbangan kinerja antar kelas yang lebih baik setelah penerapan SMOTE.



**Gambar 3.** Confusion matrix *naïve bayes* sebelum dan sesudah smote

Gambar 3 merupakan hasil perbandingan eksperimen algoritma *naïve bayes* sebelum dan sesudah penerapan SMOTE. Pada algoritma *naïve bayes*, penerapan SMOTE meningkatkan kemampuan model dalam mengenali sentimen negatif secara signifikan. Jumlah *True Negative* meningkat tajam dari 45 menjadi 881, sementara *False Positive* menurun dari 548 menjadi 135. Hal ini menunjukkan bahwa model menjadi lebih sensitif terhadap opini negatif yang sebelumnya terabaikan akibat ketidakseimbangan data. Namun, penurunan terjadi pada kemampuan mendeteksi opini positif, ditunjukkan oleh turunnya *True Positive* dari 1541 menjadi 879 dan naiknya *False Negative* dari 2 menjadi 139. Dalam konteks analisis sentimen

PPPK, peningkatan *False Negative* berarti ada lebih banyak opini positif yang diklasifikasikan sebagai negatif, yang dapat menimbulkan bias persepsi terhadap sentimen masyarakat.



**Gambar 4.** Confusion matrix svm sebelum dan sesudah smote

Hasil eksperimen performa algoritma SVM sebelum dan sesudah penerapan SMOTE yang disajikan pada gambar 4 merupakan efek SMOTE yang terlihat lebih stabil. *True Negative* meningkat dari 465 menjadi 995 dan *False Positive* menurun drastis dari 128 menjadi 21. Meski *True Positive* menurun dari 1486 menjadi 945 dan *False Negative* naik dari 57 menjadi 73, perubahan ini tidak sebesar pada Naïve Bayes. Hal ini menunjukkan bahwa SVM lebih mampu mempertahankan keseimbangan klasifikasi setelah proses oversampling. Sama seperti sebelumnya, SMOTE diterapkan hanya pada data pelatihan setelah proses split untuk menghindari data *leakage*. Namun, dalam penelitian ini belum dilakukan uji signifikansi seperti *paired t-test* atau *McNemar test*, sehingga efektivitas perubahan performa antar model belum dapat dipastikan secara statistik. Hal ini menjadi salah satu keterbatasan yang perlu diperhatikan dalam interpretasi hasil.

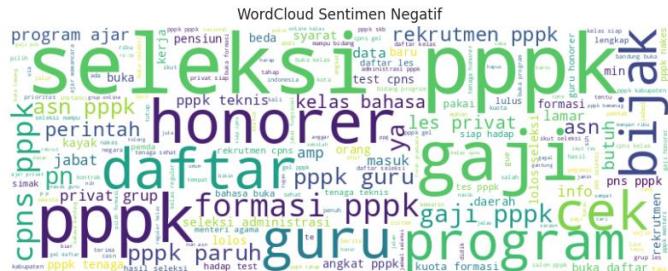


**Gambar 5.** Visualisasi wordcloud

Visualisasi *Wordcloud* digunakan untuk menampilkan kata-kata yang paling sering muncul pada dataset, sehingga memudahkan proses analisis teks. Berdasarkan gambar 5, terlihat bahwa kata-kata seperti "PPPK", "Seleksi", "Guru" Memiliki ukuran yang lebih besar dibandingkan kata lainnya, yang menunjukkan tingginya frekuensi kemunculannya kata-kata tersebut. Hal ini mengidentifikasi keterkaitan yang erat dengan topik yang dianalisis, yaitu mengenai pelaksanaan Program PPPK di Indonesia.

Hasil yang disajikan pada gambar 6 adalah *WordCloud* sentimen negatif yang menunjukkan kata-kata yang paling sering muncul dalam tweet bernuansa negatif terkait program PPPK. Kata-kata seperti "pppk", "honorer", "gaji", "seleksi", dan "daftar" mendominasi, mencerminkan isu-isu utama yang menjadi sorotan publik. Kata "honorer" dan "gaji" muncul dalam ukuran besar, yang menunjukkan kekhawatiran terkait status kepegawaian dan kompensasi yang diterima oleh tenaga honorer. Kata "seleksi", "formasi", dan "program" juga sering muncul, menandakan keluhan terhadap proses rekrutmen yang dirasa tidak adil atau

tidak transparan. Selain itu, kata seperti "tidak", "cek", "lolos", dan "formasi" mengisyaratkan frustrasi terhadap proses verifikasi dan ketidakpastian hasil seleksi.



**Gambar 6.** Visualisasi *wordcloud* sentimen negatif

## Pembahasan

Penelitian ini SVM menunjukkan performa yang lebih unggul dibandingkan *Naïve Bayes* terutama setelah mengklasifikasikan sentimen publik terhadap kebijakan Program PPPK di media sosial X, terutama setelah diterapkannya teknik SMOTE. Teknik ini berhasil mengatasi ketidakseimbangan data sentimen dengan mensintesis data baru pada kelas minoritas, sehingga meningkatkan kemampuan model dalam mengenali sentimen negatif yang sebelumnya sering terklasifikasi secara keliru. Dengan penerapan SMOTE, SVM mencapai akurasi 95%, sedangkan *Naïve Bayes* hanya 87%. Keunggulan SVM terletak pada kemampuannya membentuk *hyperplane* optimal dalam ruang fitur berdimensi tinggi, menjadikannya lebih efektif untuk data teks serta lebih tahan terhadap noise dan ketidakseimbangan. Sebaliknya, performa *Naïve Bayes* menurun karena asumsi independensi antar fitur yang tidak sesuai dengan karakteristik data teks, sehingga menghasilkan *accuracy* dan presisi yang lebih rendah.

Analisis visual menggunakan *wordcloud*, untuk mengidentifikasi kata-kata dominan seperti “Guru”, “Seleksi”, dan “PPPK”. Kata “Guru” sering muncul dalam konteks positif, mencerminkan harapan masyarakat terhadap peningkatan status dan kesejahteraan tenaga honorer. Sedangkan kata “Seleksi” muncul dalam konteks negatif, menunjukkan kekhawatiran masyarakat terhadap transparansi dan keadilan dalam proses rekrutmen. Analisis ini memberikan pemahaman yang lebih mendalam tentang isu utama yang memengaruhi persepsi publik terhadap kebijakan PPPK.

Hasil penelitian kami jika dibandingkan dengan temuan sebelumnya, penelitian yang dilakukan oleh (Karimah et al., 2024) telah menerapkan *Naïve Bayes* dengan dukungan SMOTE, namun hasil klasifikasi masih menunjukkan kecenderungan bias terhadap kelas mayoritas, yang mengintegrasikan keterbatasan algoritma *Naïve Bayes* dalam memanfaatkan data yang telah diseimbangkan. Sedangkan penelitian (Syah & Witanti, 2022) menggunakan algoritma SVM tanpa penerapan SMOTE, sehingga model yang digunakan belum sepenuhnya optimal dalam mengatasi ketidakseimbangan kelas.

Sementara itu, penelitian kami mengintegrasikan pendekatan klasifikasi SVM dengan penyeimbangan data melalui SMOTE, yang terbukti memberikan hasil yang lebih akurat, seimbang, dan informatif. Hal ini tidak hanya memperkuat hasil empiris, tetapi juga memberikan kontribusi metodologis yang penting dalam analisis sentimen berbasis media sosial terhadap kebijakan publik. Dengan demikian, dapat disimpulkan bahwa penerapan SVM yang didukung oleh SMOTE merupakan pendekatan yang lebih efektif dan andal dalam mengklasifikasikan sentimen yang tidak seimbang, sekaligus memberikan wawasan yang lebih kaya bagi pengambil kebijakan dalam merespons persepsi publik terhadap Program PPPK.

## SIMPULAN

Temuan kami menunjukkan bahwa penerapan algoritma klasifikasi *Naïve Bayes* dan SVM pada 7.508 data tweet mengenai program PPPK menghasilkan kinerja yang cukup optimal, terutama setelah penanganan ketidakseimbangan data menggunakan teknik SMOTE. Model klasifikasi dengan algoritma SVM menunjukkan rata-rata akurasi yang lebih tinggi, yaitu sebesar 95%, dibandingkan *Naïve Bayes* yang mencapai 87%. Selain itu, terjadi peningkatan signifikan pada metrik *presisi*, *recall*, dan *F1-score* setelah penerapan SMOTE, terutama dalam mendekripsi sentimen negatif. Visualisasi *wordcloud* memperlihatkan bahwa meskipun banyak kritik terhadap program PPPK, opini publik secara umum masih memperlihatkan keterlibatan aktif masyarakat dalam menanggapi isu ini. Dengan demikian, analisis ini dapat menjadi masukan berbasis data bagi pengambil kebijakan dalam menyempurnakan pelaksanaan program PPPK.

## REFERENSI

- Anjani, A. F., Anggraeni, D., & Tirta, I. M. (2023). Implementasi Random Forest Menggunakan SMOTE untuk Analisis Sentimen Ulasan Aplikasi Sister for Students UNEJ. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 9(2), 163–172. <https://doi.org/10.25077/TEKNOSI.v9i2.2023>
- Fitriani, F., Utami, E., & Hartanto, AD. (2022). Analisis Sentimen Masyarakat Terhadap Pelaksanaan P3K Guru Dengan Algoritma Naïve Bayes Dan Decision Tree. *TEKNIMEDIA: Teknologi Informasi Dan Multimedia*, 3(1), 23–30. <https://doi.org/10.46764/teknimedia.v3i1.53>
- Fitriani, F. (2023). Perbandingan Algoritma Naïve Bayes Dan Knn Dalam Analisis Sentimen Masyarakat Terhadap Pelaksanaan Pppk Guru. *Bulletin of Network Engineer and Informatics* 1(1), 18–24. <https://doi.org/10.59688/bufnets.v1i1.6>
- Hermawan, M. A., Faqih, A., Dwilestari, G. (2025). Implementasi Akurasi Model Naive Bayes Menggunakan Smote Dalam Analisis Sentimen Pengguna Aplikasi Brimo. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(1), 855-862. <http://dx.doi.org/10.23960/jitet.v13i1.5748>
- Hidayat, F. N., & Sugiyono, S. (2023). Analisis Sentimen Masyarakat Terhadap Perekutan Pppk Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine. *Jurnal Sains Dan Teknologi*, 5(2), 665–672. <https://doi.org/10.55338/saintek.v5i2.1359>
- Hidayatullah, H., Purwantoro, P., & Umaidah, Y. (2023). Penerapan Naive Bayes Dengan Optimasi Information Gain Dan Smote Untuk Analisis Sentimen Pengguna Aplikasi Chatgpt. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(3), 1546-1553. <https://doi.org/10.36040/jati.v7i3.6887>
- Karimah, A., Dwilestari, G., & Mulyawan, M. (2024). Analisis Sentimen Komentar Video Mobil Listrik Di Platform Youtube Dengan Metode Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 767–773. <https://doi.org/10.36040/jati.v8i1.8373>
- Khairunnisa, K., Dewi, S. K., Rahmawati, D. D., & Sari, A. P. (2024). Analisis Sentimen Komentar pada Postingan Instagram " StandWithUs " Menggunakan Klasifikasi Naive Bayes. *JURNAL ILMIAH INFORMATIKA*, 12(02), 191-199. <https://doi.org/10.33884/jif.v12i02.9263>
- Khairunissabina, K., & Kurniawan, R. (2024). Analisis Sentimen Mengenai Penerimaan Calon Pegawai Negeri Sipil Tahun 2024 Menggunakan Support Vector Machine. *Kemajuan: Jurnal Ilmiah Komputer*, 21 (1), 132–143. <http://dx.doi.org/10.35889/progresif.v21i1.2583>
- Natasha & R R Suryono. (2025). Analisis Sentimen Pengaruh Korean Wave Di Indonesia Dengan Metode Naive Bayes Dan Support Vector Machine. *INOVTEK Polbeng - Seri Informatika*, 10(1), 308–319. <https://doi.org/10.35314/85x4wd90>

- Nugroho, A., & Rilvani, E (2023). Penerapan Metode Oversampling SMOTE Pada Algoritma Random Forest Untuk Prediksi Kebangkrutan Perusahaan. *Techno.COM*, 22(1), 207–214. <https://doi.org/10.33633/tc.v22i1.7527>
- Pamungkas, A. S., & Cahyono, N. (2024). Analisis Sentimen Review ChatGPT di Play Store menggunakan Support Vector Machine dan K-Nearest Neighbor. *Edumatic : Jurnal Pendidikan Informatika*, 8(1), 1–10. <https://doi.org/10.29408/edumatic.v8i1.24114>
- Pratama, R. F. P., & Maharani, W. (2025). Comparative Analysis of Naive Bayes and SVM for Improved Emotion Classification on Social Media. *Edumatic: Jurnal Pendidikan Informatika*, 9(1), 11-20.
- Rahayu, S., Mz, Y., Bororing, J. E., & Hadiyat, R (2022). Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP. *Edumatic : Jurnal Pendidikan Informatika*, 6(1), 98–106. <https://doi.org/10.29408/edumatic.v6i1.5433>
- Rizqiyah, A. M., & Nuryana, I. K. D. (2024). Analisis Sentimen Masyarakat terhadap Kebijakan Iuran Tabungan Perumahan Rakyat ( Tapera ) pada Platform X Menggunakan Algoritma Naïve Bayes Classifier dan Support Vector Machine. *JEISBI: (Journal of Emerging Information Systems and Business Intelligence)*, 05(03), 298–306. <https://doi.org/10.26740/jeisbi.v5i3.64074>
- Sanjaya, T. P. R., Fauzi, A., & Masruriyah, A. F. N. (2023). Analisis sentimen ulasan pada e-commerce shopee menggunakan algoritma naive bayes dan support vector machine. *INFOTECH:Jurnal Informatika & Teknologi*, 4(1), 16–26. <https://doi.org/10.37373/infotech.v4i1.422>
- Sari, P. K., & Suryono, R. R. (2024). Komparasi Algoritma Support Vector Machine Dan Random Forest Untuk Analisis Sentimen Metaverse. *Jurnal MNEMONIC*, 7(1), 31–39. <https://doi.org/10.36040/mnemonic.v7i1.8977>
- Sari, D., & Kurniawan, R. (2024). Analisis Sentimen Terhadap Kinerja Program Walikota Medan pada Media Sosial X Menggunakan Support Vector Machine *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(4), 1539-1548. <https://doi.org/10.57152/malcom.v4i4.1685>
- Setiawan, A., & Suryono, R. R. (2024). Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes. *Edumatic: Jurnal Pendidikan Informatika*, 8(1), 183–192. <https://doi.org/10.29408/edumatic.v8i1.25667>
- Syah, H., & Witanti, A. (2022). Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine ( SVM ). *Jurnal Sistem Informasi Dan Informatika (Simika)*, 5(1). 59-67. <https://doi.org/10.47080/simika.v5i1.1411>