Jurnal Sains Dan Teknologi (JSIT)
Vol. 6 No.1 Januari-April 2026 Hal. 278 - 287
DOI : https://doi.org/10.47233/jsit.v6i1.4323
ISSN : **2807-7393**

Analisis Sentimen Publik Terkait CPNS dan PPPK Menggunakan Algoritma *Machine Learning* Pada Media Sosial

Sutirto Abas[a], Mohamad Ilyas Abas[b], Widya Eka Pranata[c], Rizal Lamusu[d]
[a]Ilmu Komputer, Sains dan Ilmu Komputer, Universitas Muhammadiyah Gorontalo, tirtoabas@gmail.com
[b]Ilmu Komputer, Sains dan Ilmu Komputer, Universitas Muhammadiyah Gorontalo, ilyasabas@umgo.ac.id
[c]Ilmu Komputer, Sains dan Ilmu Komputer, Universitas Muhammadiyah Gorontalo, widyapranata@umgo.ac.id
[d]Ilmu Komputer, Sains dan Ilmu Komputer, Universitas Muhammadiyah Gorontalo, rizal_lamusu@umgo.ac.id

Submitted: 19-01-2026, Reviewed: 22-01-2026, Accepted 09-03-2026

Abstract
The postponement of the appointment of Civil Service Candidates (CPNS) and Government Employees with Work Contracts (PPPK) has sparked a variety of public reactions, many of which have been expressed on the social media platform X. This study aims to analyze patterns of public sentiment toward this policy and to compare the performance of the Naive Bayes and Random Forest algorithms in classifying public sentiment. The study employs a quantitative approach using an experimental method based on the Cross Industry Standard Process for Data Mining (CRISP-DM) framework. Data were collected via web scraping on the social media platform X using keywords related to CPNS and PPPK during the period of March 5–16, 2025, with an initial total of 1,423 comments, which was reduced to 1,180 data points after duplicate removal. The preprocessing stages included cleaning, case folding, normalization, tokenization, and stopword removal. Sentiment labeling was performed using the InSet dictionary, while feature extraction used the TF-IDF method. The model was built using the Naive Bayes and Random Forest algorithms with a data split of 80% for training and 20% for testing. Model performance evaluation was performed using a confusion matrix and ROC curve. The results of the study indicate that public sentiment is dominated by negative sentiment, accounting for 61.27%. The Random Forest model achieved an accuracy of 78% with an AUC of 0.88, outperforming the Naive Bayes model, which had an accuracy of 75% and an AUC of 0.86. These results suggest that the Random Forest model is more effective at classifying sentiment in Indonesian-language social media text data.
Keywords: *Sentiment Analysis; Civil Service Candidate; PPPK; Naive Bayes; Random Forest*

Abstrak
Penundaan pengangkatan Calon Pegawai Negeri Sipil (CPNS) dan Pegawai Pemerintah dengan Perjanjian Kerja (PPPK) menimbulkan beragam respons publik yang banyak disampaikan melalui media sosial X. Penelitian ini bertujuan untuk menganalisis pola sentimen publik terhadap kebijakan tersebut serta membandingkan kinerja algoritma Naive Bayes dan Random Forest dalam mengklasifikasikan sentimen masyarakat. Penelitian menggunakan pendekatan kuantitatif dengan metode eksperimen berbasis kerangka kerja Cross Industry Standard Process for Data Mining (CRISP-DM). Data dikumpulkan melalui teknik web scraping pada media sosial X dengan kata kunci terkait CPNS dan PPPK selama periode 5–16 Maret 2025 dengan jumlah awal 1.423 komentar, yang setelah proses penghapusan duplikasi menjadi 1.180 data. Tahapan preprocessing meliputi cleaning, case folding, normalisasi, tokenisasi, dan stopword removal. Pelabelan sentimen dilakukan menggunakan kamus InSet, sedangkan ekstraksi fitur menggunakan metode TF-IDF. Model dibangun menggunakan algoritma Naive Bayes dan Random Forest dengan pembagian data 80% training dan 20% testing. Evaluasi kinerja model dilakukan menggunakan confusion matrix dan kurva ROC. Hasil penelitian menunjukkan bahwa sentimen publik didominasi oleh sentimen negatif sebesar 61,27%. Random Forest memperoleh akurasi sebesar 78% dengan nilai AUC 0,88, lebih tinggi dibandingkan Naive Bayes dengan akurasi 75% dan AUC 0,86. Hasil ini menunjukkan bahwa Random Forest lebih optimal dalam mengklasifikasikan sentimen pada data teks media sosial berbahasa Indonesia.
Kata Kunci: Analisis Sentimen; CPNS; PPPK; Naive Bayes; Random Forest

This work is licensed under Creative Commons Attribution License 4.0 CC-BY International license

PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah mendorong media sosial menjadi ruang komunikasi publik yang dominan dalam menyampaikan opini, sikap, dan respons masyarakat terhadap berbagai isu sosial dan kebijakan pemerintah. Melalui media sosial, masyarakat dapat mengekspresikan pandangan secara cepat, terbuka, dan massif, sehingga dinamika opini publik dapat terbentuk dalam waktu singkat [1]. Komentar yang muncul dimedia sosial tidak hanya bersifat informatif, tetapi juga mencerminkan emosi, ketidakpuasan, serta harapan masyarakat terhadap kebijakan yang berdampak langsung pada kehidupan mereka. Kondisi ini menjadikan media sosial sebagai sumber data yang relevan untuk memahami persepsi publik secara empiris dan berbasis data [2].

Tingginya tingkat adopsi media sosial di Indonesia semakin memperkuat peran platform digital dalam membentuk opini publik. Laporan Hootsuite (We Are Social) tahun 2025 menunjukkan bahwa jumlah pengguna aktif media sosial di Indonesia mencapai sekitar 143 juta orang, dengan platform X menjadi salah satu media

Jurnal Sains Dan Teknologi (JSIT)
Vol. 6 No.1 Januari-April 2026 Hal. 278 - 287
DOI : https://doi.org/10.47233/jsit.v6i1.4323
ISSN : **2807-7393**

yang banyak digunakan untuk memperoleh informasi secara real-time termasuk isu-isu terkait kebijakan publik [3]. Dalam konteks kebijakan publik, respons masyarakat yang terekam di media sosial memiliki implikasi penting karena dapat memengaruhi kepercayaan publik terhadap pemerintahan serta efektivitas komunikasi kebijakan. Kebijakan yang menyangkut aspek strategis, seperti kepastian status pekerjaan dan kesejahteraan, cenderung memicu respons publik yang lebih intens dan emosional. Oleh karena itu, pemahaman terhadap pola sentiment masyarakat menjadi kebutuhan agar pemerintah dapat mengevaluasi dan menyempurnakan strategi penyampaian kebijakan secara lebih transparan dan responsif.

Salah satu kebijakan yang memicu perhatian masyarakat di media sosial adalah penundaan pengangkatan Calon Pegawai Negeri Sipil (CPNS) dan Pegawai Pemerintah dengan Perjanjian Kerja (PPPK). Kebijakan ini diumumkan sebagai bagian dari penyesuaian manajemen aparatur sipil negara, kesiapan anggaran, serta penataan regulasi [4]. Namun, bagi masyarakat yang telah dinyatakan lulus seleksi, kebijakan tersebut menimbulkan ketidakpastian terkait status pekerjaan, penghasilan, dan perencanaan masa depan. Dampak langsung kebijakan ini mendorong munculnya beragam respons publik yang diekspresikan secara aktif melalui media sosial [5].

Platform X menjadi salah satu ruang utama diskusi publik terkait kebijakan penundaan pengangkatan CPNS dan PPPK. Karakteristik platform ini yang terbuka, berbasis teks singkat, serta responsive terhadap isu aktual memungkinkan masyarakat menyampaikan opini secara spontan. Komentar yang muncul sering kali menggunakan bahasa informal, singkatan, serta ungkapan emosional yang mencerminkan sikap subjektif pengguna. Di sisi lain, karateristik tersebut juga menghadirkan tantangan tersendiri dalam pengolahan data teks, khususnya dalam konteks bahasa Indonesia yang memiliki variasi linguistic tinggi [6].

Analisis sentiment merupakan bagian dari text mining yang bertujuan untuk mengidentifikasi kecenderungan sikap atau emosi dalam teks [7]. Pada data media sosial berbahasa Indonesia, analisis sentiment dihadapkan pada berbagai kendala, seperti penggunaan kata tidak baku, bahasa gaul, sarkasme, dan makna implisit yang tidak selalu dapat ditangkap secara langsung oleh model komputasional. Tantangan-tantangan tersebut menuntut penerapan metode analisis yang sistematis serta pemilihan algoritma yang mampu mengolah data teks berdimensi tinggi secara efektif dan akurat [8].

Sejumlah penelitian terdahulu telah dilakukan untuk menganalisis sentimen masyarakat terhadap isu CPNS dan PPPK. Penelitian oleh Hidayat dan Sugiyono menggunakan algoritma Naive Bayes dan Support Vector Machine dalam menganalisis sentimen terhadap perekrutan PPPK di media sosial Twitter, dengan hasil akurasi yang tinggi mencapai 96,14% untuk Naive Bayes dan 94,80% untuk SVM. Penelitian ini menunjukkan bahwa algoritma machine learning tradisional mampu memberikan performa yang baik dalam klasifikasi sentiment [9]. Selanjutnya, penelitian oleh Nurhidayat dan Hendrastuty menganalisis sentimen terhadap tes CPNS menggunakan algoritma Naive Bayes dengan jumlah data sebesar 5.599 komentar. Hasil penelitian menunjukkan bahwa model mampu mencapai akurasi hingga 95%, namun mengalami kendala pada ketidakseimbangan data yang menyebabkan dominasi klasifikasi pada sentimen negatif. Untuk mengatasi permasalahan tersebut, digunakan metode SMOTE yang terbukti mampu meningkatkan kemampuan model dalam mendeteksi sentimen positif [10].

Penelitian lain oleh Yudistira dan Isnain juga melakukan analisis sentimen terhadap seleksi CPNS berbasis media sosial X menggunakan algoritma Naive Bayes dengan jumlah data sebanyak 3.222 tweet. Hasil penelitian menunjukkan bahwa model mencapai akurasi sebesar 82,02%, dengan performa yang lebih baik dalam mendeteksi sentimen positif dibandingkan sentimen negative. Hal ini menunjukkan bahwa metode Naive Bayes cukup efektif dalam analisis sentimen, meskipun masih memiliki keterbatasan dalam menangani distribusi data yang tidak seimbang [11]. Selain pendekatan machine learning tradisional, penelitian terbaru juga mulai memanfaatkan metode deep learning. Penelitian oleh Asshiddiq dan Witanti menggunakan model IndoBERT untuk menganalisis sentimen terhadap penundaan pengangkatan CPNS 2025. Hasil penelitian menunjukkan bahwa model IndoBERT mampu mencapai akurasi sebesar 84,27% dengan kemampuan yang lebih baik dalam memahami konteks bahasa Indonesia secara mendalam [4]. Hal ini menunjukkan bahwa pendekatan berbasis transformer memiliki potensi yang besar dalam meningkatkan kualitas analisis sentiment [12].

Meskipun berbagai penelitian tersebut telah memberikan kontribusi yang signifikan, sebagian besar penelitian masih berfokus pada isu seleksi CPNS atau perekrutan PPPK, serta menggunakan satu pendekatan algoritma tanpa melakukan perbandingan kinerja secara komprehensif. Selain itu, penelitian yang secara khusus mengkaji sentimen publik terhadap kebijakan penundaan pengangkatan CPNS dan PPPK masih terbatas, terutama yang mengaitkan hasil analisis dengan implikasi terhadap evaluasi kebijakan publik.

Berdasarkan celah penelitian tersebut, diperlukan suatu kajian yang tidak hanya menganalisis sentimen publik terhadap kebijakan penundaan pengangkatan CPNS dan PPPK, tetapi juga membandingkan kinerja algoritma yang digunakan dalam proses klasifikasi. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen publik terhadap kebijakan penundaan pengangkatan CPNS dan PPPK menggunakan data dari media sosial X, serta membandingkan performa algoritma Naive Bayes dan Random Forest dalam mengklasifikasikan

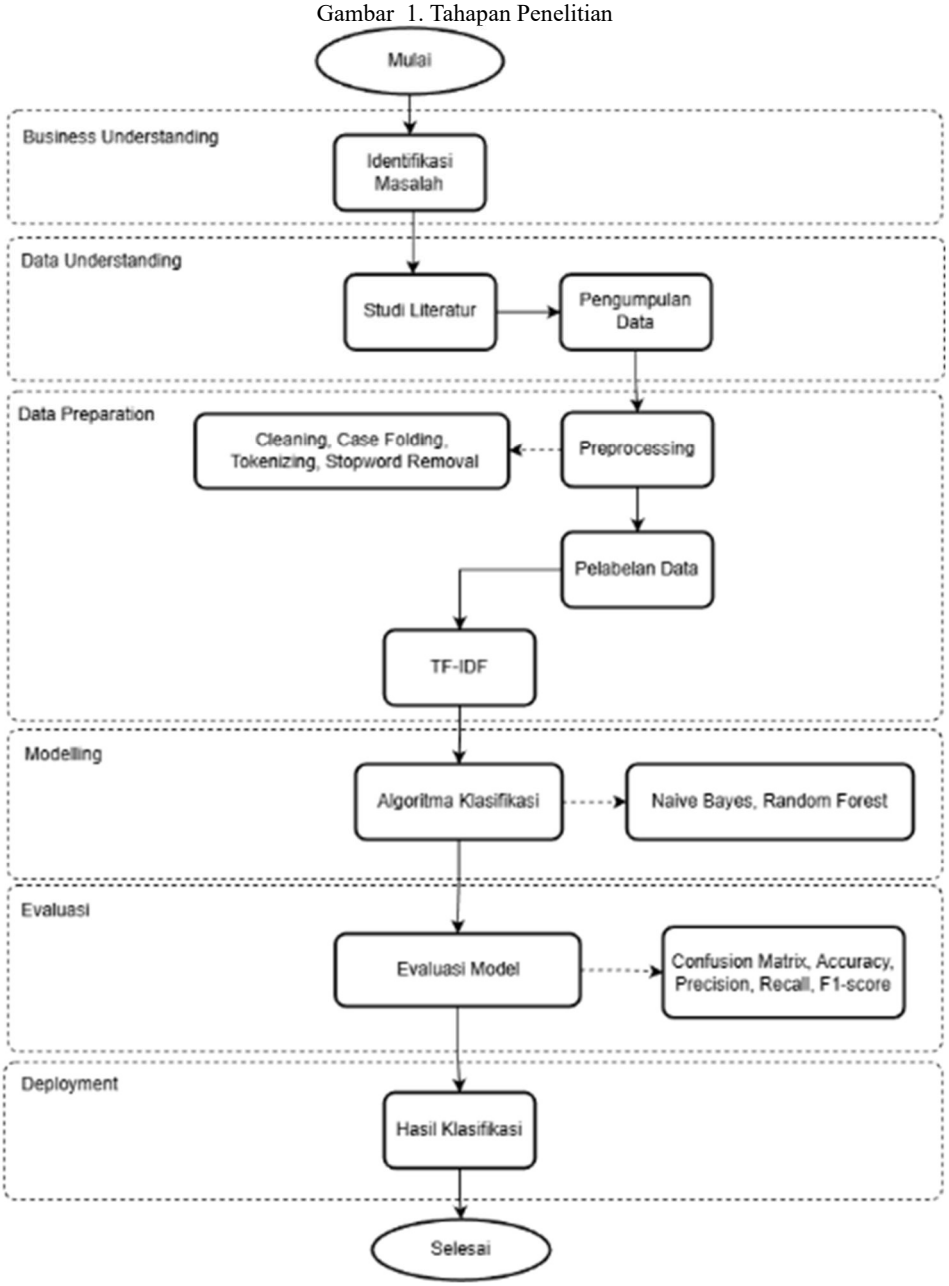
Jurnal Sains Dan Teknologi (JSIT)
Vol. 6 No.1 Januari-April 2026 Hal. 278 - 287
DOI : https://doi.org/10.47233/jsit.v6i1.4323
ISSN : **2807-7393**

sentimen masyarakat. Hasil penelitian ini diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai persepsi publik serta menjadi bahan pertimbangan dalam evaluasi dan penyusunan kebijakan pemerintah yang lebih responsif dan berbasis data.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimen untuk mengembangkan dan mengevaluasi model analisis sentiment terhadap kebijakan penundaan pengangkatan Calon Pegawai Negeri Sipil (CPNS) dan Pegawai Pemerintah dengan Perjanjian Kerja (PPPK) pada media sosial X. pendekatan eksperimen dipilih untuk memungkinkan pengujian performa algoritma secara terukur melalui indicator evaluasi kuantitatif, sekaligus membandingkan kinerja dua algoritma klasifikasi machine learning, yaitu naïve bayes dan random forest dalam mengklasifikasikan sentiment publik. Proses penelitian disusun berdasarkan kerangka kerja Cross Industry Standard Process for Data Mining (CRISP-DM) yang meliputi tahapan Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment [13]. karena kerangka ini menyediakan alur sistematis dalam pengembangan model berbasis data teks.

Gambar 1. Tahapan Penelitian



Sumber : Olahan Penelitian

Jurnal Sains Dan Teknologi (JSIT)
Vol. 6 No.1 Januari-April 2026 Hal. 278 - 287
DOI : https://doi.org/10.47233/jsit.v6i1.4323
ISSN : **2807-7393**

Berdasarkan gambar 1 tahap awal penelitian dimulai dengan pengumpulan data dari platform media sosial X menggunakan teknik web scraping melalui Google Colaboratory. Data diambil pada periode 5–16 Maret 2025, yaitu setelah munculnya kebijakan penyesuaian jadwal pengangkatan CPNS dan PPPK, sehingga data merepresentasikan respons publik pada fase awal munculnya polemik kebijakan. Proses pengambilan data menggunakan kata kunci "penundaan", "CPNS", dan "PPPK", dengan fokus pada komentar yang berkaitan dengan akun resmi Kementerian PANRB. Dataset awal yang diperoleh berjumlah 1.423 komentar dengan beberapa atribut metadata, namun penelitian ini hanya menggunakan atribut full_text sebagai variabel utama karena memuat opini pengguna seperti pada tabel 1. Data kemudian disimpan dalam format CSV untuk memudahkan proses analisis lanjutan.

Tabel 1. Dataset Awal

No.	full_text
1	Dalam Rapat Kerja Komisi 2 @DPR_RI dengan @kempanrb tidak ada kesepakatan penundaan. Justru DPR terlihat dari pendapat masing-masing fraksi meminta agar proses pengangkatan CPNS dan PPPK dipercepat.
2	Polemik Penundaan Pengangkatan CPNS dan PPPK Picu Unjuk Rasa
3	Penjelasan BKN Soal Penundaan Pengangkatan CPNS PPPK
....	.....
1420	Sabar dikit lagi! Yang penting CPNS tetap jalan bukan batal https://t.co/Xcj8wiZ8YT (10) #TOLAKKEBIJAKANTMTSERENTAK #CPNS
1421	@khasyahfr As I mentioned before yang pertama harus dibenerin adalah penggunaaan dana pajaknya yg ga dikorup&sebesar-besarnya utk kemakmuran rakyat. Naikin tarif PPN sebenernya juga key point (komparasi global) hasil tarif efektif ppn ga naik ya keknya salah satunya penundaan cpns&pppk.
1422	Ulasan Istana: Kisruh Penundaan Pengangkatan Bagaimana Solusi Nasib CPNS & PPPK?

Sumber : Hasil *Scrapping* pada media sosial X, 2025

Dataset yang digunakan dalam penelitian ini diperoleh dari pengumpulan data melalui hasil scrapping menggunakan google colaboratory pada platform media sosial X. Dataset awal yang diperoleh berjumlah 1.423 komentar, kemudian dilakukan proses penghapusan data duplikat sebanyak 243 komentar sehingga diperoleh dataset akhir sebanyak 1.180 komentar yang digunakan dalam penelitian. Tahapan preprocessing data dilakukan untuk meningkatkan kualitas dan konsistensi data teks sebelum digunakan dalam proses pemodelan. Proses ini diawali dengan tahap cleaning, yaitu menghapus elemen yang tidak relevan seperti URL, mention (@username), hashtag (#), angka, emoji, serta karakter non-alfabet menggunakan regular expression. Selanjutnya dilakukan case folding, yaitu mengubah seluruh teks menjadi huruf kecil (lowercase) untuk menghindari perbedaan representasi kata akibat penggunaan huruf kapital. Tahap berikutnya adalah normalisasi, di mana kata-kata tidak baku atau slang diubah menjadi bentuk baku menggunakan kamus slang Bahasa Indonesia agar makna kata menjadi lebih konsisten. Setelah itu dilakukan tokenizing, yaitu proses pemecahan teks menjadi unit kata (token) sehingga setiap kata dapat dianalisis secara individual. Kemudian dilakukan stopword removal untuk menghapus kata-kata umum yang tidak memiliki makna signifikan, seperti "dan", "di", "ke", dan "yang", menggunakan daftar stopword Bahasa Indonesia dari pustaka Natural Language Toolkit (NLTK). Terakhir, dilakukan stemming, yaitu proses mengubah kata ke bentuk dasar dengan menghilangkan imbuhan, sehingga kata-kata dengan makna yang sama dapat direpresentasikan dalam bentuk yang seragam. Tahapan preprocessing ini bertujuan untuk mengurangi noise pada data teks serta meningkatkan kualitas representasi fitur sehingga dapat meningkatkan performa model klasifikasi. Tahapan preprocessing ini bertujuan untuk mengurangi noise pada data teks serta meningkatkan kualitas representasi fitur sehingga dapat meningkatkan performa model kualifikasi [14].

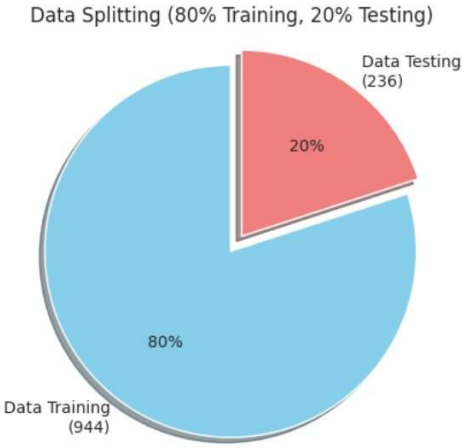
Pelabelan sentimen dilakukan secara otomatis menggunakan kamus leksikon InSet (Indonesian Sentiment Lexicon) untuk mengklasifikasikan komentar ke dalam kategori sentimen positif dan negatif. Pendekatan berbasis leksikon dipilih karena efisien dan tidak memerlukan pelabelan manual dalam jumlah besar, namun memiliki keterbatasan dalam mendeteksi sarkasme dan konteks implisit pada teks media sosial. Hasil pelabelan menunjukkan distribusi sentimen negatif sebanyak 723 data dan sentimen positif sebanyak 457 data. Setelah pelabelan, dilakukan ekstraksi fitur menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) untuk mengubah teks menjadi representasi numerik yang dapat diproses oleh algoritma machine learning.

Setelah data melalui tahapan preprocessing dan ekstrasi fitur menggunakan metode TF-IDF, dataset kemudian dibagi menjadi data latih dan data uji menggunakan fungsi train_test_split. Pembagian data dilakukan

Jurnal Sains Dan Teknologi (JSIT)
Vol. 6 No.1 Januari-April 2026 Hal. 278 - 287
DOI : https://doi.org/10.47233/jsit.v6i1.4323
ISSN : **2807-7393**

dengan proporsi 80% untuk data training dan 20% untuk data testing. Dari total 1.180 data, diperoleh 944 data training dan 236 data testing seperti pada gambar 2. Pembagian ini menggunakan parameter random_state=42 untuk menjaga konsistensi hasil eksperimen serta stratify=y agar distribusi kelas sentiment pada data training dan testing tetap seimbang. Pembagian data ini bertujuan agar model dapat dilatih menggunakan Sebagian besar data, kemudian diuji menggunakan data yang belum pernah dilihat sebelumnya. Dengan demikian, performa algoritma Naïve Bayes dan Random Forest dapat dievaluasi secara lebih objektif dalam mengklasifikasikan sentiment publik terhadap kebijakan penundaan pengangkatan CPNS dan PPPK.

Gambar 2. Pembagian Dataset



Sumber: Hasil analisis, 2025

Tahap pemodelan dilakukan dengan mengimplementasikan dua algoritma klasifikasi, yaitu Multinomial Naive Bayes dan Random Forest menggunakan pustaka scikit-learn dalam bahasa pemrograman Python. Algoritma Naive Bayes dipilih karena sederhana, efisien, dan banyak digunakan pada klasifikasi teks berbasis frekuensi kata [15]. Sementara itu, Random Forest dipilih karena merupakan metode ensemble berbasis banyak pohon keputusan yang mampu menangani hubungan non-linear antar fitur dan mengurangi risiko overfitting melalui mekanisme bootstrap aggregation [16]. Pada penelitian ini, model Random Forest diimplementasikan dengan parameter n_estimators=100 dan random_state=42 untuk menjaga konsistensi hasil eksperimen.

Evaluasi kinerja model dilakukan pada data uji menggunakan metrik accuracy, precision, recall, dan F1-score, serta analisis confusion matrix untuk mengamati pola kesalahan klasifikasi pada masing-masing kelas sentiment. Penggunaan confusion matrix dan metrik evaluasi tersebut bertujuan untuk memberikan gambaran yang lebih rinci mengenai performa prediksi serta pola kesalahan klasifikasi yang dihasilkan oleh model [17].Selain itu, dilakukan analisis kurva Receiver Operating Characteristic (ROC) dan perhitungan Area Under Curve (AUC) untuk menilai kemampuan model dalam membedakan kelas positif dan negatif berdasarkan probabilitas prediksi. Penggunaan berbagai metrik evaluasi tersebut bertujuan memberikan gambaran performa model secara komprehensif dan tidak hanya bergantung pada nilai akurasi semata.

Tahap akhir penelitian adalah deployment dalam bentuk penyajian hasil analisis melalui visualisasi distribusi sentimen, perbandingan confusion matrix, serta grafik ROC untuk kedua algoritma. Hasil analisis ini diharapkan dapat dimanfaatkan sebagai bahan evaluasi komunikasi kebijakan publik berbasis data, serta menjadi dasar pengembangan sistem monitoring sentimen otomatis yang terintegrasi dengan media sosial pemerintah untuk mendukung pengambilan keputusan yang lebih responsive.

HASIL PENELITIAN DAN PEMBAHASAN

Hasil Penelitian

Pengumpulan data dilakukan pada periode 5-16 Maret 2025, yaitu pada fase awal setelah diumumkannya kebijakan penyesuaian jadwal pengangkatan CPNS dan PPPK oleh pemerintah. Pemilihan periode ini didasarkan pada asumsi bahwa respons publik pada fase awal kebijakan cenderung lebih spontan dan mencerminkan persepsi awal masyarakat. Data yang diperoleh berjumlah 1.423 komentar dari media sosial X, kemudian dilakukan proses pembersihan data duplikat sehingga diperoleh dataset akhir sebanyak 1.180 komentar yang siap dianalisi. Distribusi dataset setelah melalui tahapan preprocessing ditunjukkan pada tabel 2 berikut :

Jurnal Sains Dan Teknologi (JSIT)
Vol. 6 No.1 Januari-April 2026 Hal. 278 - 287
DOI : https://doi.org/10.47233/jsit.v6i1.4323
ISSN : **2807-7393**

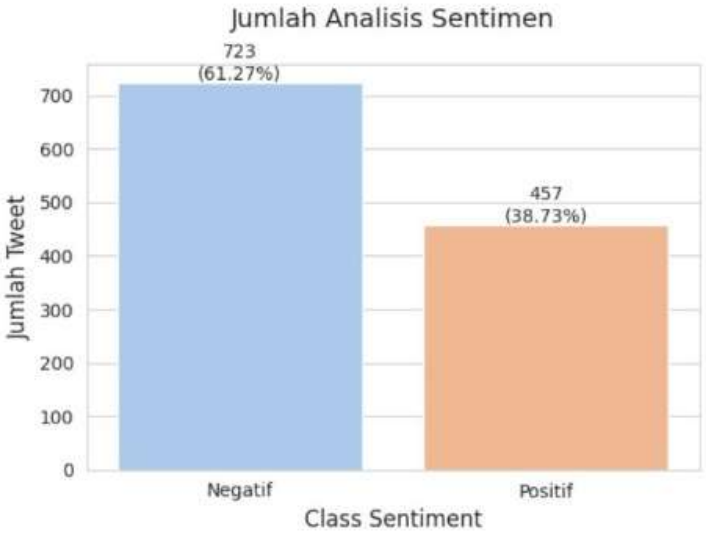
Tabel 2. Distribusi Dataset Penelitian

No.	Tahapan Data	Jumlah Data
1	Data awal hasil scraping	1.423
2	Data Duplikat	243
3	Data setelah cleaning	1.180
4	Sentimen Negatif	723
5	Sentimen Positif	457
6	Data Training (80%)	944
7	Data Testing (20%)	236

Sumber : Hasil analisis, 2025

Hasil pelabelan sentimen menggunakan kamus InSet menunjukkan bahwa dari total 1.180 komentar, sebanyak 723 data (61,27%) termasuk kategori sentimen negative dan 457 data (38%) termasuk kategori sentiment positif. Visualisasi distribusi hasil pelabelan sentimen ditunjukkan pada gambar 3, yang memperlihatkan dominasi sentimen negative dibandingkan sentimen positif.

Gambar 3. Hasil Pelabelan Sentimen



Sumber: Hasil Analisis, 2025

Dominasi sentimen negatif menunjukkan bahwa sebagian besar masyarakat merespons kebijakan penundaan pengangkatan CPNS dan PPPK dengan sikap ketidakpuasaan dan kekhawatiran. Proses preprocessing dilakukan menggunakan bahasa pemrograman Python dengan pustaka Pandas untuk manipulasi data, Regular Expression untuk cleaning teks, Natural Language Toolkit (NLTK) untuk stopword removal, serta Scikit-learn untuk ekstraksi fitur TF-IDF. Tahapan preprocessing meliputi cleaning, case folding, normalisasi kata tidak baku, tokenisasi, dan penghapusan stopword. Proses ini bertujuan meningkatkan kualitas representasi teks karena karakteristik komentar media sosial cenderung tidak terstruktur dan mengandung variasi bahasa informal. Hasil pelabelan sentimen menggunakan kamus InSet menunjukkan bahwa dari total 1.180 komentar, sebanyak 723 data (61,27%) termasuk kategori sentimen negatif dan 457 data (38,73%) termasuk sentimen positif. Dominasi sentimen negatif menunjukkan bahwa sebagian besar publik merespons kebijakan penundaan dengan sikap ketidakpuasan dan kekhawatiran. Dataset kemudian dibagi menggunakan metode hold-out dengan proporsi 80% data latih (944 data) dan 20% data uji (236 data). Pembagian dilakukan secara stratified untuk menjaga keseimbangan distribusi kelas.

Performa model klasifikasi menggunakan algoritma Naïve Bayes dan Random Forest disajikan dalam bentuk table dan visualisasi confusion matrix.

Jurnal Sains Dan Teknologi (JSIT)
Vol. 6 No.1 Januari-April 2026 Hal. 278 - 287
DOI : https://doi.org/10.47233/jsit.v6i1.4323
ISSN : **2807-7393**

Tabel 3. Hasil Evaluasi Model Klasifikasi

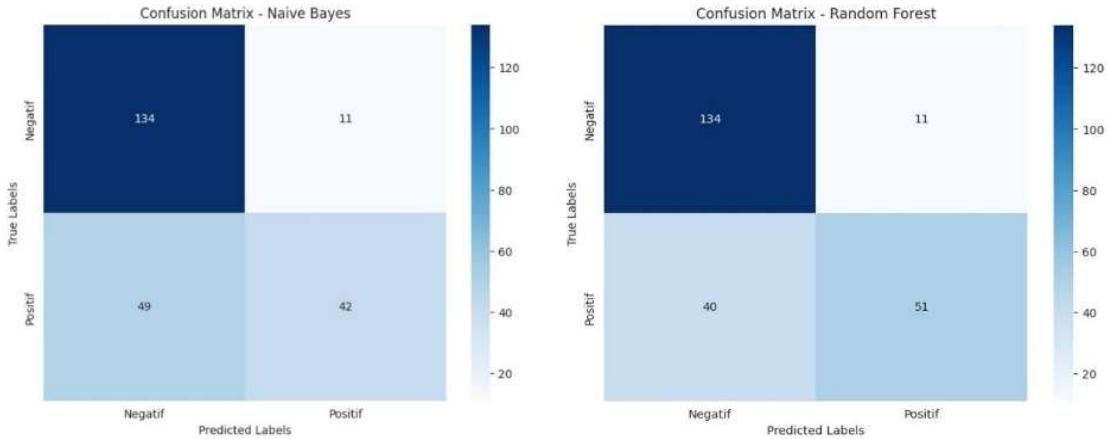
Algoritma	Kelas Sentimen	Precision	Recall	F1-Score	Support
Naive Bayes	Negatif	0.73	0.92	0.82	145
	Positif	0.79	0.46	0.58	91
	Akurasi			0.75	236
Random Forest	Negatif	0.77	0.92	0.84	145
	Positif	0.82	0.56	0.67	91
	Akurasi			0.78	236

Sumber: Hasil Analisis, 2025

Berdasarkan tabel 3 algoritma random forest memiliki performa yang lebih baik dibandingkan naïve bayes pada hampir seluruh metrik evaluasi. Dari sisi akurasi random forest mencapai nilai sebesar 0,78, lebih tinggi dibandingkan Naive Bayes yang memperoleh nilai sebesar 0,75. Hal ini menunjukkan bahwa Random Forest memiliki tingkat ketepatan yang lebih baik dalam mengklasifikasikan keseluruhan data sentimen. Pada metrik precision, Random Forest juga menunjukkan nilai yang lebih tinggi baik pada kelas negatif (0,77) maupun kelas positif (0,82), dibandingkan Naive Bayes yang masing-masing sebesar 0,73 dan 0,79, sehingga dapat dikatakan bahwa Random Forest lebih akurat dalam memprediksi label sentimen. Dari sisi recall, kedua algoritma memiliki nilai yang sama tinggi pada kelas negatif yaitu 0,92, yang menunjukkan bahwa keduanya mampu dengan baik mendeteksi sentimen negatif sebagai kelas mayoritas. Namun demikian, pada kelas positif, Random Forest memiliki nilai recall yang lebih tinggi yaitu sebesar 0,56 dibandingkan Naive Bayes sebesar 0,46, yang menunjukkan bahwa Random Forest lebih mampu mengenali data positif yang sebelumnya sulit terdeteksi. Selain itu, nilai F1-score yang menggambarkan keseimbangan antara precision dan recall juga menunjukkan bahwa Random Forest memiliki performa yang lebih stabil, dengan nilai sebesar 0,84 pada kelas negatif dan 0,67 pada kelas positif, dibandingkan Naive Bayes yang masing-masing sebesar 0,82 dan 0,58. Dengan demikian, berdasarkan keseluruhan metrik evaluasi pada Tabel 3, dapat disimpulkan bahwa Random Forest merupakan algoritma yang lebih optimal dalam mengklasifikasikan sentimen publik pada data media sosial dibandingkan Naive Bayes.

Visualisasi perbandingan confusion matrix dari kedua algoritma ditunjukkan pada gambar 4, yang menggambarkan distribusi prediksi benar dan salah pada masing-masing kelas sentiment.

Gambar 4. Hasil Perbandingan Confusion Matrix



Sumber : Hasil analisis, 2025

Berdasarkan Gambar 4. Hasil Perbandingan Confusion Matrix, terlihat perbedaan performa antara algoritma Naive Bayes dan Random Forest dalam mengklasifikasikan sentimen publik. Pada algoritma Naive Bayes, jumlah data yang berhasil diklasifikasikan dengan benar untuk kelas negatif (True Negative) sebanyak 134 data, sedangkan kesalahan klasifikasi negatif sebagai positif (False Positive) sebanyak 11 data. Untuk kelas positif, terdapat 42 data yang berhasil diprediksi dengan benar (True Positive), namun masih terdapat 49 data yang salah diklasifikasikan sebagai negatif (False Negative). Hal ini menunjukkan bahwa Naive Bayes memiliki kelemahan dalam mendeteksi sentimen positif, yang terlihat dari jumlah kesalahan yang cukup tinggi pada kelas tersebut. Sementara itu, pada algoritma Random Forest, jumlah data yang diklasifikasikan dengan benar untuk kelas negatif tetap sama yaitu 134 data, dengan kesalahan False Positive sebanyak 11 data. Namun, pada kelas

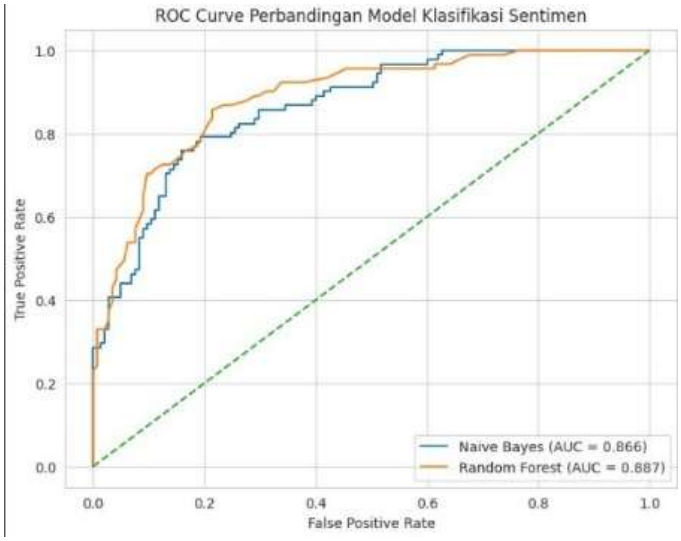
Jurnal Sains Dan Teknologi (JSIT)
Vol. 6 No.1 Januari-April 2026 Hal. 278 - 287
DOI : https://doi.org/10.47233/jsit.v6i1.4323
ISSN : **2807-7393**

positif, Random Forest menunjukkan peningkatan performa dengan jumlah True Positive sebanyak 51 data dan False Negative yang lebih rendah yaitu 40 data. Hal ini menunjukkan bahwa Random Forest lebih mampu mengenali sentimen positif dibandingkan Naive Bayes.

Secara keseluruhan, kedua algoritma memiliki kemampuan yang sama baik dalam mengklasifikasikan sentimen negatif, namun Random Forest lebih unggul dalam mendeteksi sentimen positif, yang ditunjukkan dengan peningkatan jumlah prediksi benar dan penurunan kesalahan klasifikasi pada kelas tersebut. Dengan demikian, confusion matrix ini memperkuat hasil evaluasi sebelumnya bahwa Random Forest memiliki performa yang lebih baik dan lebih seimbang dalam menangani klasifikasi sentimen pada data media sosial.

Selain evaluasi berbasis confusion matrix, dilakukan juga analisis menggunakan kurva ROC untuk menilai kemampuan model dalam membedakan kelas sentimen. Hasil visualisasi ROC ditunjukkan pada gambar 5.

Gambar 5. ROC Curve



Sumber: Hasil analisis, 2025

Berdasarkan analisis tersebut, algoritma Naive Bayes memperoleh nilai AUC sebesar 0,86, sedangkan Random Forest memperoleh nilai AUC sebesar 0,88. Nilai AUC yang lebih tinggi pada Random Forest menunjukkan bahwa model ini memiliki kemampuan diskriminasi yang lebih baik dalam membedakan antara sentimen positif dan negatif.

Pembahasan

Dominasi sentimen negatif dalam penelitian ini dapat dijelaskan oleh karakteristik kebijakan penundaan pengangkatan CPNS dan PPPK yang berdampak langsung terhadap kondisi ekonomi dan psikologis masyarakat, khususnya bagi peserta yang telah dinyatakan lulus seleksi. Kebijakan ini menimbulkan ketidakpastian terkait status pekerjaan, penghasilan, serta perencanaan masa depan, sehingga memicu respons emosional berupa kekecewaan, kekhawatiran, dan ketidakpercayaan terhadap kebijakan pemerintah. Selain itu, karakteristik media sosial X yang bersifat terbuka dan real-time turut memperkuat ekspresi emosi pengguna, di mana individu dengan pengalaman negatif cenderung lebih aktif menyampaikan opini dibandingkan pengguna dengan respons netral atau positif.

Jika dikaitkan dengan kondisi sosial masyarakat, dominasi sentimen negatif ini mencerminkan tingginya sensitivitas publik terhadap kebijakan yang berkaitan dengan stabilitas pekerjaan dan kesejahteraan. Dalam konteks sosial ekonomi Indonesia, status sebagai Aparatur Sipil Negara (ASN) masih dianggap sebagai pekerjaan yang menjanjikan keamanan finansial dan kepastian karier. Oleh karena itu, penundaan pengangkatan tidak hanya dipandang sebagai penyesuaian administratif, tetapi juga sebagai ancaman terhadap kepastian hidup masyarakat. Kondisi ini semakin diperkuat oleh informasi di media yang menunjukkan bahwa sebagian peserta telah mengundurkan diri dari pekerjaan sebelumnya setelah dinyatakan lulus seleksi, sehingga memperbesar dampak psikologis dan ekonomi yang dirasakan.

Hasil penelitian ini sejalan dengan penelitian terdahulu yang menunjukkan bahwa isu kebijakan publik yang berdampak langsung terhadap masyarakat cenderung memunculkan dominasi sentimen negatif di media sosial. Penelitian sebelumnya terkait analisis sentimen CPNS dan PPPK juga menemukan bahwa sebagian besar komentar masyarakat mengandung respons negatif, terutama pada isu yang berkaitan dengan ketidakpastian atau ketidakpuasan terhadap kebijakan [9]–[11]. Namun demikian, penelitian ini memberikan kontribusi

Jurnal Sains Dan Teknologi (JSIT)
Vol. 6 No.1 Januari-April 2026 Hal. 278 - 287
DOI : https://doi.org/10.47233/jsit.v6i1.4323
ISSN : **2807-7393**

tambahan dengan secara khusus mengkaji kebijakan penundaan pengangkatan CPNS dan PPPK, yang belum banyak dibahas pada penelitian sebelumnya. Selain itu, penelitian ini tidak hanya menganalisis sentimen, tetapi juga membandingkan performa dua algoritma klasifikasi, yaitu Naive Bayes dan Random Forest.

hasil penelitian menunjukkan bahwa Random Forest memiliki performa yang lebih baik dibandingkan Naive Bayes, terutama dalam meningkatkan recall dan F1-score pada kelas positif. Hal ini menunjukkan bahwa Random Forest lebih mampu menangkap variasi dan kompleksitas bahasa dalam data media sosial yang bersifat tidak terstruktur. Dengan demikian, temuan ini tidak hanya memperkuat hasil penelitian sebelumnya, tetapi juga memberikan pendekatan yang lebih komprehensif dalam analisis sentimen berbasis machine learning.

KESIMPULAN

Berdasarkan hasil analisis sentimen publik terhadap kebijakan penundaan pengangkatan CPNS dan PPPK pada media sosial X, dapat disimpulkan bahwa respons masyarakat didominasi oleh sentimen negatif dengan proporsi sebesar 61,27%. Dominasi sentimen negatif ini tidak hanya menunjukkan ketidakpuasan, tetapi juga mencerminkan adanya ketidakpastian psikologis dan kekhawatiran terhadap dampak ekonomi serta status kepegawaian peserta yang telah dinyatakan lulus seleksi. Karakter kebijakan yang menyangkut kepastian kerja dan stabilitas penghasilan menjadi faktor utama yang memicu respons emosional kuat di ruang publik digital.

Dari sisi metodologis, perbandingan algoritma menunjukkan bahwa Random Forest memperoleh performa lebih baik dibandingkan Naive Bayes, dengan akurasi sebesar 78% dan nilai AUC sebesar 88%, lebih tinggi dibandingkan Naive Bayes dengan akurasi 75% dan AUC 86%. Keunggulan Random Forest terletak pada kemampuannya menangani hubungan non-linear antar fitur serta mengurangi variansi melalui pendekatan ensemble learning. Sebaliknya, Naive Bayes yang mengasumsikan independensi antar fitur cenderung kurang optimal dalam menangkap keterkaitan kontekstual antar kata dalam teks media sosial yang bersifat informal dan kompleks. Hal ini terlihat dari nilai recall dan F1-score kelas positif yang lebih rendah pada Naive Bayes.

Implikasi praktis dari penelitian ini menunjukkan bahwa analisis sentimen berbasis machine learning dapat dimanfaatkan sebagai instrumen monitoring opini publik secara kuantitatif dalam evaluasi kebijakan pemerintah. Pemerintah dapat menggunakan sistem analisis sentimen untuk mengidentifikasi lonjakan persepsi negatif secara real-time, memperbaiki strategi komunikasi publik, serta meningkatkan transparansi informasi terkait kebijakan strategis yang berdampak langsung pada masyarakat. Dengan pendekatan berbasis data, proses pengambilan keputusan dapat menjadi lebih responsif dan adaptif terhadap dinamika opini publik.

Namun demikian, penelitian ini memiliki beberapa keterbatasan. Pertama, pelabelan sentimen menggunakan pendekatan leksikon (InSet) memiliki keterbatasan dalam mendeteksi sarkasme, ironi, dan konteks implisit yang sering muncul dalam komentar media sosial. Kedua, distribusi kelas yang tidak sepenuhnya seimbang berpotensi memengaruhi performa model dalam mendeteksi kelas minoritas (positif). Ketiga, data yang digunakan hanya berasal dari satu platform media sosial dalam periode waktu terbatas, sehingga generalisasi hasil ke populasi masyarakat secara luas perlu dilakukan dengan hati-hati. Keempat, penelitian ini hanya membandingkan dua algoritma klasifikasi tradisional tanpa mengeksplorasi pendekatan berbasis deep learning yang berpotensi memberikan performa lebih tinggi.

Berdasarkan keterbatasan tersebut, penelitian selanjutnya disarankan untuk: Menggunakan metode pelabelan semi-supervised atau manual annotation sebagai pembanding pendekatan leksikon, Mengimplementasikan algoritma berbasis deep learning seperti LSTM, Bidirectional LSTM, atau Transformer untuk menangkap konteks semantik yang lebih kompleks, Menerapkan teknik penyeimbangan data seperti SMOTE atau class weighting untuk mengurangi bias kelas, serta Memperluas sumber data dari berbagai platform media sosial dan periode waktu yang lebih panjang guna meningkatkan validitas eksternal penelitian.

Dengan demikian, penelitian ini memberikan kontribusi empiris dalam memahami pola sentimen publik terhadap kebijakan penundaan pengangkatan CPNS dan PPPK serta menunjukkan bahwa Random Forest merupakan model yang lebih stabil dan efektif untuk klasifikasi sentimen pada konteks data teks media sosial berbasis kebijakan publik.

UCAPAN TERIMAKASIH

Penulis menyampaikan penghargaan dan rasa terima kasih kepada Universitas Muhammadiyah Gorontalo atas dukungan yang telah diberikan. Penulis menyadari bahwa penyusunan jurnal ini dapat terlaksana dengan baik berkat kontribusi dan bantuan dari berbagai pihak. Oleh karena itu, penulis juga mengucapkan terima kasih yang sebesar-besarnya kepada Dosen Pembimbing, rekan-rekan, serta seluruh pihak yang telah memberikan dukungan dan bantuan, baik secara langsung maupun tidak langsung, sehingga jurnal ini dapat diselesaikan dengan baik.

Jurnal Sains Dan Teknologi (JSIT)
Vol. 6 No.1 Januari-April 2026 Hal. 278 - 287
DOI : https://doi.org/10.47233/jsit.v6i1.4323
ISSN : **2807-7393**

DAFTAR PUSTAKA

S. Suhendra and F. Selly Pratiwi, "Peran Komunikasi Digital dalam Pembentukan Opini Publik: Studi Kasus Media Sosial," Iapa Proc. Conf., p. 293, 2024, doi: 10.30589/proceedings.2024.1059.

Juleha, J. Yuniar, and N. R. Marsuki, "Peran Media Sosial Dalam Dinamika Opini Publik dan Partisipasi Politik Era Digital," Concept J. Soc. Humanit. Educ., vol. 3, no. 1, pp. 38–45, 2024, [Online]. Available: https://doi.org/10.55606/concept.v3i1.951

Andi, "Hootsuite (We are Social): Data Digital Indonesia 2025 - Dosen, Praktisi, Konsultan, Pembicara/Fasilitator Digital Marketing, Internet marketing, SEO, Technopreneur dan Bisnis Digital." Accessed: Sep. 21, 2025. [Online]. Available: https://andi.link/hootsuite-we-are-social-data-digital-indonesia-2025/

K. PANRB, "Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi - Pengangkatan Serentak Butuh Waktu, Harus Cermat dan Hati-Hati." Accessed: Sep. 21, 2025. [Online]. Available: https://menpan.go.id/site/berita-terkini/pengangkatan-serentak-butuh-waktu-harus-cermat-dan-hati-hati

BBC, "Pemerintah undur jadwal pengangkatan CPNS, puluhan ribu warga protes – 'Banyak yang sudah resign' - BBC News Indonesia." Accessed: Sep. 21, 2025. [Online]. Available: https://www.bbc.com/indonesia/articles/c39vrdplmmro

K. Pramudya, "Media Sosial X dalam Membentuk Identitas Sosial Digital Generasi Z," J. Ilm. Penelit. Mhs., vol. 3, no. 2, pp. 178–183, 2025, [Online]. Available: https://doi.org/10.61722/jipm.v3i2.807

L. O. Lukmana, D. R. L, and P. Elisya, "Analisis Sentimen Terhadap Ulasan Pengguna Aplikasi Threads Instagram di Playtore Menggunakan Algoritma Naive Bayes," vol. 13, no. 2, 2025.

R. talita Trista and E. tri Asmoro, "Analisis Sentimen dalam Data Big Data," J. Multidiscip. Inq. Sci. Technol. Educ. Res., vol. 2, no. 1, pp. 700–706, 2025.

F. N. Hidayat and S. Sugiyono, "Analisis Sentimen Masyarakat Terhadap Perekrutan Pppk Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine," J. Sains dan Teknol., vol. 5, no. 2, pp. 665–672, 2023, doi: 10.55338/saintek.v5i2.1359.

R. Nurhidayat and N. Hendrastuty, "Analisis Sentimen Komentar Media Sosial Twitter Terhadap Tes CPNS dengan Algoritma Naive Bayes," Build. Informatics, Technol. Sci., vol. 6, no. 3, pp. 1477–1489, 2024, doi: 10.47065/bits.v6i3.6148.

F. Yudistira and A. Rahman Isnain, "Analisis Sentimen Terhadap Seleksi CPNS Tahun 2024 Berbasis Media Sosial X Menggunakan Algoritma Naïve Bayes," J. Pendidik. dan Teknol. Indones., vol. 5, no. 3, pp. 887–897, 2025.

A. W. M. Asshiddiq, "Analisis Sentimen tentang Penundaan Pengangkatan CPNS 2025 pada Platform X Menggunakan Metode IndoBERT," vol. 17, no. 2, 2025, doi: 10.30736/jt.v17i2.1448.

F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, and M. Jajuli, "Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM," J. Teknol. dan Inf., vol. 12, no. 1, pp. 64–77, 2022, doi: 10.34010/jati.v12i1.6674.

A. M. Rizqiyah, I. Kadek, and D. Nuryana, "Analisis Sentimen Masyarakat terhadap Kebijakan Iuran Tabungan Perumahan Rakyat (Tapera) pada Platform X Menggunakan Algoritma Naïve Bayes Classifier dan Support Vector Machine," J. Emerg. Inf. Syst. Bus. Intell., vol. 5, no. 3, pp. 298–306, 2024.

A. N. M. P, A. A. Y. L, A. I. J, Y. A. S, Sumanto, and A. D. H, "Komparasi Naive Bayes dan SVM untuk Analisis Sentimen Pada E-Commerce Seller Center," J. Sains Dan Teknol., vol. 5, no. 3, pp. 199–208, 2025, doi: 10.47233/jsit.v5i2.3211.

S. Amaliah, M. Nusrang, and Aswi, "Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng," VARIANSI J. Stat. Its Appl. Teach. Res., vol. 4, no. 3, pp. 121–127, 2022, doi: 10.35580/variansiunm31.

M. I. Abas, S. Syarif, I. Nurtanio, and Z. Tahir, "Comparison of Convolutional Neural Network Methods For the Classification of Maize Plant Diseases," Regist. J. Ilm. Teknol. Sist. Inf., vol. 10, no. 1, pp. 46–59, 2024, doi: 10.26594/register.v10i1.3656.