

Analisis Sentimen Terhadap Seleksi CPNS Tahun 2024 Berbasis Media Sosial X Menggunakan Algoritma *Naïve Bayes*

Firnanda Yudistira^{*1}, Auliya Rahman Isnain²

^{1,2}Program Studi Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia,
Indonesia

Email: ¹firnanda_yudistira@teknokrat.ac.id, ²aulyarahman@teknokrat.ac.id

Abstrak

Analisis sentimen opini terhadap seleksi CPNS tahun 2024 berdasarkan media sosial X menerapkan algoritma *Naïve Bayes* telah dilakukan dalam penelitian ini. Penelitian bertujuan agar dapat melakukan identifikasi terhadap sentimen masyarakat serta mengevaluasi tingkat akurasi algoritma. Data yang dipakai meliputi 3222 *tweet* yang diambil dari media sosial X dengan kata kunci "cpns 2024". Hasil pengujian memperlihatkan bahwasanya model *Naïve Bayes* mencapai akurasi sebesar 82.02%, dengan nilai *F1-score* untuk kategori sentimen positif mencapai 89% dan untuk kategori negatif sebesar 49%. *Precision* dan *recall* untuk sentimen positif masing-masing mencapai 92% dan 86%, sementara untuk sentimen negatif adalah 42% dan 59%. Penelitian ini menyumbangkan wawasan mengenai persepsi masyarakat terhadap seleksi CPNS dan dapat membantu pemerintah dalam menyusun strategi komunikasi yang semakin efektif. Hasil penelitian diharapkan mampu berkontribusi signifikan dalam evaluasi proses seleksi CPNS tahun 2024 dan menjadi inspirasi bagi penelitian serupa di masa mendatang.

Kata kunci: Analisis Sentimen, Algoritma *Naïve Bayes*, Seleksi CPNS, Media Sosial, Opini Publik

SENTIMENT ANALYSIS OF THE 2024 CPNS SELECTION BASED ON SOCIAL MEDIA X USING THE NAÏVE BAYES ALGORITHM

Abstract

This research analyzes public sentiment towards the 2024 CPNS selection based on social media X using the *Naïve Bayes* algorithm. The study aims to identify public sentiment and evaluate the accuracy level of the algorithm. The data used consists of 3,222 tweets collected from social media X with the keyword "cpns 2024". The testing results show that the *Naïve Bayes* model achieves an accuracy of 82.02%, with an *F1-score* for the positive sentiment category reaching 89% and for the negative sentiment category at 49%. *Precision* and *recall* for positive sentiment are 92% and 86%, respectively, while for negative sentiment, they are 42% and 59%. This research provides insights into public perceptions regarding the CPNS selection and can assist the government in designing more effective communication strategies. The findings are expected to contribute significantly to the evaluation of the 2024 CPNS selection process and inspire similar research in the future.

Keywords: Sentiment Analysis, *Naïve Bayes* Algorithm, CPNS Selection, Social Media, Public Opinion

1. PENDAHULUAN

Penerimaan Calon Pegawai Negeri Sipil (CPNS) adalah suatu proses seleksi yang sangat dinanti di Indonesia setiap tahunnya. Proses ini menarik perhatian masyarakat secara luas dan sering kali menjadi bahan diskusi di berbagai platform media sosial [1]. Dalam konteks digital saat ini, opini masyarakat terhadap seleksi CPNS sering kali dapat ditemukan dalam bentuk unggahan, komentar, dan diskusi di media sosial seperti Media Sosial X [2].

Media sosial sudah berperan sebagai sarana yang sangat berpengaruh dalam membangun pandangan masyarakat [3]. Dengan meningkatnya jumlah orang yang aktif di media sosial, data dihasilkan dari platform ini memberikan wawasan penting mengenai sentimen masyarakat terhadap berbagai kebijakan pemerintah, termasuk seleksi CPNS [4]. Data tersebut mencakup berbagai pandangan, mulai dari dukungan, kritik, hingga masukan yang dapat digunakan sebagai bahan evaluasi [5]. Satu dari beberapa algoritma yang umum dipakai dalam analisis sentimen ialah *Naïve Bayes*. Algoritma tersebut dilandasi oleh probabilitas dan memiliki keunggulan dalam mengelola data teks dalam jumlah besar [6].

Penelitian sebelumnya dengan pembobotan *tf-idf* diiringi penambahan konversi ikon emosi agar dapat mengetahui kelas sentimen yang terdapat dalam tweet mengenai toko JD.id. Hasil penelitian memperlihatkan

metode *Naïve Bayes* dengan tidak adanya penambahan fitur dapat melakukan klasifikasi terhadap sentimen dengan nilai akurasi sebesar 96,44%, sedangkan yang diiringi konversi ikon emosi dapat memiliki nilai akurasi sebesar 98% [7]. Analisis sentimen dari algoritma ini menunjukkan bahwa algoritma *Naïve Bayes* menunjukkan performa yang kompetitif jika dibandingkan dengan metode lain, seperti Support Vector Machine (SVM), sehingga menjadikannya sebuah algoritma yang andal dalam analisis sentimen berbasis media sosial [8]. Selain itu, algoritma ini efektif dalam memetakan opini publik terhadap isu tertentu, seperti pengunduran diri ASN setelah lolos seleksi [9]. Algoritma *Naïve Bayes* ini juga dapat dikombinasikan dengan metode lain, seperti *Decision Tree*, untuk menghasilkan analisis yang lebih mendalam dalam konteks kebijakan publik. Selain itu, algoritma ini telah menunjukkan keandalannya dalam membandingkan performanya dengan metode KNN untuk mengklasifikasikan sentimen masyarakat secara akurat [10]. Metode ini tidak hanya relevan untuk analisis teks, tetapi juga dapat diterapkan pada data berbasis citra [11]. Pada penelitian lain, algoritma ini telah dimanfaatkan dalam deteksi pola berbasis citra untuk analisis tertentu [12]. Bahkan dalam sektor jasa, algoritma ini telah diterapkan untuk menganalisis sentimen berbasis layanan, seperti ojek online, dan memberikan hasil yang signifikan [13]. Dalam ranah politik, algoritma *Naïve Bayes* telah digunakan untuk menganalisis opini masyarakat terhadap kinerja lembaga legislatif di media sosial, menyediakan pemahaman yang mendalam mengenai persepsi publik [14]. Algoritma ini juga terbukti mampu mengolah data komersial, seperti *e-commerce*, untuk memahami pola sentimen pelanggan secara efektif [15].

Analisis sentimen berperan sebagai sebuah pendekatan yang relevan untuk memahami opini publik secara lebih mendalam. Teknik ini digunakan untuk mengelompokkan opini masyarakat ke dalam kategori tertentu, seperti positif maupun negatif, dengan menggunakan data teks yang diambil dari media sosial [16] [17].

Penelitian ini memiliki tujuan agar dapat melakukan analisis mengenai sentimen masyarakat terhadap seleksi CPNS tahun 2024 dengan memanfaatkan data dari Media Sosial X. Dengan menggunakan algoritma *Naïve Bayes*, studi ini tidak hanya bertujuan agar dapat mengidentifikasi sentimen masyarakat, tetapi juga mengevaluasi tingkat akurasi dari algoritma tersebut.

Hasil pengujian menunjukkan bahwa model *Naïve Bayes* yang diterapkan dalam studi ini. ini mampu mencapai akurasi sebesar 82.02%. Selain itu, nilai F1-score untuk kategori sentimen positif mencapai 89%, sedangkan untuk kategori negatif sebesar 49%. Precision dan recall untuk sentimen positif masing-masing mencapai 92% dan 86%, sementara untuk sentiment negative adalah 42% dan 59%. Pendekatan berbasis analisis sentimen ini tidak hanya memberikan wawasan tentang persepsi masyarakat secara keseluruhan, tetapi juga memungkinkan untuk memetakan isu-isu spesifik yang menjadi perhatian utama publik.

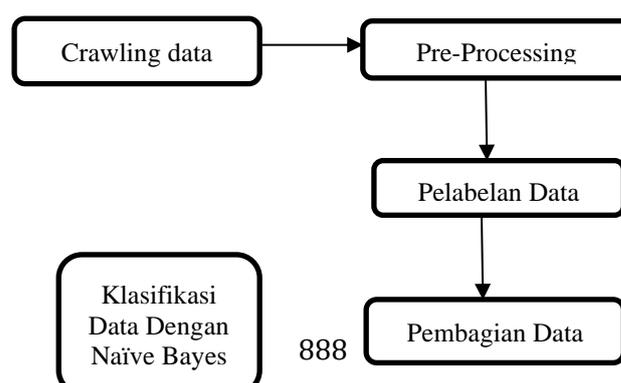
Dalam konteks seleksi CPNS, analisis ini dapat membantu pemerintah memahami tantangan dan peluang dalam memperbaiki proses seleksi. Selain itu, temuan dari penelitian ini mampu berperan sebagai dasar untuk pengembangan kebijakan yang lebih responsif terhadap kebutuhan masyarakat. Dengan memahami pola sentimen yang muncul, pemerintah mampu menyusun strategi komunikasi yang semakin efektif untuk memperbaiki transparansi dan membangun kepercayaan masyarakat terhadap proses seleksi CPNS.

Penelitian ini diharap bisa memberi dampak yang signifikan dalam mengevaluasi proses seleksi CPNS tahun 2024. Selain itu, hasilnya juga diharapkan dapat menjadi inspirasi bagi penelitian serupa di masa mendatang, baik dalam konteks seleksi pegawai pemerintah maupun analisis kebijakan publik lainnya.

2. METODE PENELITIAN

2.1 Tahapan Penelitian

Dalam hal ini, Peneliti menggunakan metode yang melibatkan opini dari X karena media sosial menyediakan data yang dapat dikumpulkan, dianalisis, dan diinterpretasikan untuk memahami pandangan masyarakat mengenai tes CPNS [2]. Proses ini mempermudah dalam mengidentifikasi sentimen, baik yang berkarakter positif maupun negatif. Tahapan penelitian mampuu diamati pada Gambar 1 berikut.





Gambar 1. Tahapan Penelitian

a. *Crawling Data*

Crawling data dilakukan dengan menggunakan python dengan Google Collaboratory. Penulis mendapatkan 3222 data tweet pengguna dengan kata kunci “cpns 2024” pada media sosial X . Hasil pencarian selanjutnya akan menunjukkan seluruh tweet yang berisi kata kunci “cpns 2024” seluruh tweet yang telah diperoleh selanjutnya melalui proses pengolahan menuju proses berikutnya dengan memanfaatkan Tools Python Libraries [18].

b. *Preprocessing Data*

Data *tweet* yang sudah diambil dari X masih berwujud data mentah sehingga dilaksanakan tahapan preprocessing guna memperoleh data bersih supaya bisa diproses ke tahapan berikutnya [19]. Tahap yang dilaksanakan mencakup:

1. *Cleansing*: Ddalam tahapan ini, manipulasi ataupun hapus data yang tidak lengkap dan tidak valid, ataupun hapus baris maupun kolom yang tidak memiliki relevansi.
2. *Case Folding*: Langkah ini memiliki tujuan agar dapat mengubah seluruh karakter pada teks menjadi huruf kecil.
3. *Tokenizing*: Langkah ini memiliki tujuan agar dapat memecah teks menjadi unsur-unsur kecil, bisa berwujud kata, kalimat, simbol, ataupun unsur teks lainnya yang bermakna khusus.
4. *Stop Removal*: Tahapan ini memiliki tujuan agar dapat menghapus data penghubung misalnya kata “dan”, ”atau”, ”ke”.
5. *Stemming*: Langkah ini membantu memperkecil perbedaan antara kata-kata yang berarti sama namun ditulis dalam wujud berbeda.

c. *Pelabelan Data*

Pelabelan dilakukan guna menjamin identitas masing-masing data komentar, dilaksanakan dengan cara manual sejalan dengan kebutuhan sistem selaku data latih, yang mana terdapat tiga kategori sentimen yakni positif, negatif maupun netral. Hasil pelabelan diberikan dalam bentuk .csv yang sudah berlabel [20].

d. *Pembagian Dataset*

Dalam Konteks ini, Peneliti membagi dataset yaitu data train (latih) dan data uji (tes) [21].

e. *Klasifikasi Naïve Bayes*

Naïve Bayes classifier dipahami sebagai model pembelajaran mesin probabilistik yang dipakai dalam mengklasifikasikan sesuatu berlandaskan *teorema bayes* [22].

2.2 *Naïve Bayes*

Naive Bayes ialah metode statistik yang digunakan dalam melakukan klasifikasi terhadap data dengan memperkirakan probabilitas keanggotaan suatu kelas. Sebagai satu dari sepuluh algoritma teratas dalam data mining, *Naive Bayes* memanfaatkan frekuensi klasifikasi dari data pelatihan agar dapat membuat prediksi dan menentukan klasifikasi [23].

Meskipun merupakan metode klasifikasi berbasis probabilitas yang sederhana dan didasarkan pada *teorema Bayes*, *Naive Bayes* memiliki tujuan agar dapat menentukan probabilitas paling tinggi dari klasifikasi yang akurat. Metode ini dapat melaksanakan proses klasifikasi dengan lebih cepat dibandingkan teknik lainnya yang lebih kompleks, namun tetap menghasilkan performa yang sebanding atau bahkan lebih unggul [24]. *Naive Bayes* berlandaskan pada teorema Bayes dengan rumus di bawah ini.

$$P(X|Y) = \frac{P(X|Y) \times P(X)}{P(Y)} \quad (1)$$

Penjelasan:

P(A): Probabilitas dari kejadian A.

P(B): Probabilitas dari kejadian B.
 P(A | B): Probabilitas dari kejadian A akan terjadi apabila B telah terjadi.
 P(B | A): Probabilitas dari kejadian B akan terjadi apabila A telah terjadi.

2.3 Confusion Matrix

Confusion Matrix ialah matriks yang dipakai dalam melakukan analisis sebaik apa classifier mengenali data dari kelas yang beragam. Tabel mengenai *Confusion Matrix* mampu diketahui dalam Tabel 1 Berikut .

Tabel 1. Rumus *Confusion Matrix*

Prediction Class	Actual Class	
	Pred.Positif	Pred.Negatif
Negatif	TP	TN
Positif	FP	FN

Dalam tabel 1 mampu dipahami pada pemodelan statistik, secara khusus pada klasifikasi, *Confusion Matrix* ialah tabel yang dipakai dalam melakukan evaluasi kinerja model prediksi. Umumnya dipakai dalam merepresentasikan kinerja model klasifikasi dalam set data uji yang sudah dilatih terlebih dahulu.

Confusion Matrix memiliki informasi mengenai sebanyak apa prediksi yang salah maupun benar yang disusun oleh model klasifikasi dalam beragam kelas target. Pada *Matrix Confusion*, terdapat empat entri utama, yakni:

- 1 *True Positive* (TP): Jumlah sampel yang secara benar diklasifikasikan sebagai positif oleh model.
- 2 *True Negative* (TN): Jumlah sampel yang secara benar diklasifikasikan sebagai negatif oleh model.
- 3 *False Positive* (FP): Jumlah sampel yang salah diklasifikasikan sebagai positif oleh model (false alarm).
- 4 *False Negative* (FN): Jumlah sampel yang salah diklasifikasikan sebagai negatif oleh model (miss).

Untuk mengevaluasi model, berbagai metrik misalnya akurasi, presisi, recall, maupun *F1-score* dapat dihitung. Evaluasi ini bertujuan agar dapat menentukan apakah model yang dikembangkan bekerja dengan baik ataupun tidak. Metode evaluasi yang dipakai dalam studi ini ialah *confusion matrix*, yaitu tabel yang menyajikan perbandingan antara hasil klasifikasi yang diprediksi oleh sistem dengan hasil klasifikasi sebenarnya. Istilah "Positif" dan "Negatif" dalam evaluasi ini merepresentasikan respons pengguna terhadap tes CPNS [25].

Dalam *Confusion Matrix*, ada empat istilah utama yang mewakili hasil klasifikasi, yakni *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), serta *False Negative* (FN). Keempat nilai ini dimanfaatkan dalam melakukan perhitungan metrik evaluasi misalnya presisi, *F1-score*, akurasi, maupun recall. Berdasarkan nilai-nilai yang diperoleh dari *confusion matrix*, metrik-metrik tersebut bisa dihitung menggunakan persamaan berikut (2).

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \tag{2}$$

$$Presisi = \frac{TP}{TP+FP} \tag{3}$$

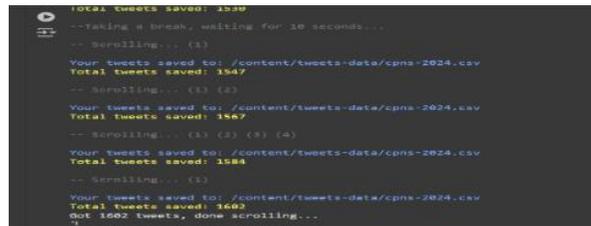
$$Recall = \frac{TP}{TP+FN} \tag{4}$$

$$f1 - score = \frac{2 \times Presisi \times Recall}{Presisi + Recall} \tag{5}$$

3. HASIL DAN PEMBAHASAN

3.1 Crawling Data

Proses *crawling* Menggunakan analisis yang dilaksanakan secara bertahap. Proses yang diikuti meliputi pengumpulan data *tweet* dari X dengan menerapkan bahasa pemrograman *Python* melalui *Google Collaboratory*. Proses pengumpulan data berhasil mengumpulkan 3222 *tweet* dari media sosial X. Sentimen yang dianalisis terdiri dari sentimen yang negatif maupun positif, yang bertujuan untuk memberikan peningkatan terhadap persepsi ataupun pandangan masyarakat terhadap Tes CPNS 2024. Data yang telah dikumpulkan akan disimpan dalam format *csv* [26].



Gambar 2. Hasil dan Proses *Crawling*

Pada gambar 2 terdapat hasil data crawling yang dimana peneliti melakukan 2 tahap scrolling untuk mendapatkan *tweet* data sampai 3500 dan menjadi 3222 data setelah dibersihkan dengan menghapus data-data duplikat.

3.2 PreProcessing

Preprocessing atau pra-pemrosesan, merupakan seperangkat tahapan yang diambil agar dapat menyiapkan data sebelum dianalisis atau diproses lebih lanjut. Dalam konteks analisis data, khususnya dalam pengolahan teks dan pembelajaran mesin, *preprocessing* sendiri dengan tujuan untuk membersihkan serta melakukan penyederhanaan data supaya semakin mudah dipahami dan diolah oleh algoritma [27].

a. Cleaning

Cleaning atau pembersihan data, merupakan proses yang bertujuan untuk mengidentifikasi serta memperbaiki atau menghilangkan data yang tidak tepat, tidak lengkap, ataupun tidak memiliki relevansi. dalam sebuah dataset. Tujuan dari pembersihan ini adalah untuk meningkatkan kualitas data, sehingga analisis yang dilakukan menjadi lebih tepat dan dapat diandalkan [28]. Mampu ditinjau dalam Tabel 2.

Tabel 2. Hasil *Preprocessing Cleaning*

Alur Preprocessing	Sebelum	Sesudah
<i>Cleaning</i>	Pengumuman Jadwal Ujian dan Cara Pilih Lokasi Tes SKB Non CAT Kemenag 2024 di https://t.co/lzEvJ300Yn https://t.co/GbRr5q8T7m lewat @tribunnews #skb #kemenag #cpns2024 #cpns #lowker #lowongankerja	Pengumuman Jadwal Ujian dan Cara Pilih Lokasi Tes SKB Non CAT Kemenag di lewat skb kemenag cpns cpns lowker lowongankerja

b. Case Folding

Case folding ialah langkah dalam pengolahan teks yang mencakup pengubahan seluruh huruf dalam teks menjadi huruf kecil (*lowercase*) ataupun huruf besar (*uppercase*). Tujuan dari *case folding* ialah agar dapat menghapus perbedaan antara huruf besar maupun kecil, oleh karenanya analisis dan pencocokan teks menjadi lebih mudah. Dengan menerapkan *case folding*, kata-kata yang sama dapat dikenali tanpa memperhatikan penggunaan huruf besar atau kecil, yang meningkatkan akurasi dalam analisis teks [29]. Hasil proses ini mampu ditinjau dalam Tabel 3.

Tabel 3. Hasil *Preprocessing Case Folding*

Alur Preprocessing	Sebelum	Sesudah
<i>Cleaning</i>	Pengumuman Jadwal Ujian dan Cara Pilih Lokasi Tes SKB Non CAT Kemenag 2024 di https://t.co/lzEvJ30OYn https://t.co/GbRr5q8T7m lewat @tribunnews #skb #kemenag #cpns2024 #cpns #lowker #lowongankerja	pengumuman jadwal ujian dan cara pilih lokasi tes skb non cat kemenag di lewat skb kemenag cpns cpns lowker lowongankerja

c. Tokenizing

Tokenizing ialah proses dalam pengolahan teks yang memiliki tujuan agar dapat memecah teks menjadi beberapa unit yang lebih kecil yang dikenal dengan sebutan token. Token ini bisa berwujud kata, frasa, atau simbol, bergantung pada tujuan analisis. Proses *tokenizing* sangat esensial dalam analisis teks maupun pemrosesan bahasa alami (NLP) sebab memberikan kemungkinan bagi komputer agar dapat memahami maupun melakukan pemrosesan teks dengan semakin efektif [30]. Hasil proses ini mampu ditinjau dalam Tabel 4.

Tabel 4. *Preprocessing Tokenizing*

Alur Preprocessing	Sebelum	Sesudah
<i>Cleaning</i>	Pengumuman Jadwal Ujian dan Cara Pilih Lokasi Tes SKB Non CAT Kemenag 2024 di https://t.co/lzEvJ30OYn https://t.co/GbRr5q8T7m lewat @tribunnews #skb #kemenag #cpns2024 #cpns #lowker #lowongankerja	['pengumuman', 'jadwal', 'ujian', 'dan', 'cara', 'pilih', 'lokasi', 'tes', 'skb', 'non', 'cat', 'kemenag', 'di', 'lewat', 'skb', 'kemenag', 'cpns', 'cpns', 'lowker', 'lowongankerja']

d. Stop Removal

Stop removal, atau penghapusan *stop words*, merupakan tahapan dalam pengolahan teks yang bertujuan untuk menghapus berbagai kata umum yang dinilai tidak bermakna signifikan dalam analisis. Berbagai kata tersebut, yang dikenal sebagai "*stop words*," kerap kali mencakup kata-kata semacam "dan," "atau," "adalah," "di," "ke," dan kata-kata fungsional lainnya yang kerap timbul dalam bahasa tetapi tidak memberikan informasi penting tentang konten atau makna teks [31]. Mampu ditinjau dalam Tabel 5 berikut.

Tabel 5. Hasil *Preprocessing Stopremoval*

Alur Preprocessing	Sebelum	Sesudah
Cleaning	Pengumuman Jadwal Ujian dan Cara Pilih Lokasi Tes SKB Non CAT Kemenag 2024 di https://t.co/lzEvJ30OYn https://t.co/GbRr5q8T7m lewat @tribunnews #skb #kemenag #cpns2024 #cpns #lowker #lowongankerja	['pengumuman', 'jadwal', 'ujian', 'cara', 'pilih', 'lokasi', 'tes', 'skb', 'non', 'cat', 'kemenag', 'lewat', 'skb', 'kemenag', 'cpns', 'cpns', 'lowker', 'lowongankerja']

e. *Stemming*

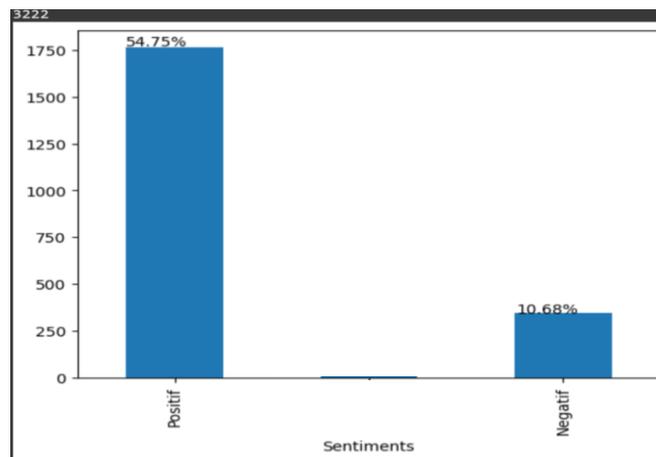
Stemming ialah proses dalam pengolahan bahasa yang alami (*Natural Language Processing*, NLP) yang memiliki tujuan agar dapat melakukan perubahan terhadap berbagai kata ke bentuk dasarnya ataupun akar kata (*stem*). Proses ini menghilangkan *afiks* (seperti awalan dan akhiran) dari istilah untuk memperoleh bentuk dasar yang umum, oleh karenanya berbagai kata yang bermakna sama bisa dikelompokkan bersama [32]. Hasil stemming mampu diamati dalam Tabel 6.

Tabel 6. Hasil *Preprocessing Stemming*

Alur Preprocessing	Sebelum	Sesudah
Cleaning	Pengumuman Jadwal Ujian dan Cara Pilih Lokasi Tes SKB Non CAT Kemenag 2024 di https://t.co/lzEvJ30OYn https://t.co/GbRr5q8T7m lewat @tribunnews #skb #kemenag #cpns2024 #cpns #lowker #lowongankerja	umum jadwal uji cara pilih lokasi tes skb non cat kemenag lewat skb kemenag cpns cpns lowker lowongankerja

3.3 Pelabelan Data

Labelling dalam sentimen analisis adalah proses memberi label atau kategori pada teks berdasarkan emosi atau pendapat yang terkandung di dalamnya. Pada tahap proses ini melibatkan pengumpulan data, pra pemrosesan teks, penentuan kategori sentimen seperti positif maupun negatif. Pada tahap ini, proses pengujian dilakukan menggunakan program *Python* di *Google Collaboratory*, di mana analisis sentimen menunjukkan hasil pelabelan data dengan persentase positif sebesar 54,75% dan negatif sebesar 10,68%. Di bawah ini ada diagram yang merepresentasikan hasil dari proses pengkategorian data ke dalam kategori positif maupun negatif.



Gambar 3. Data hasil pelabelan

Dalam Gambar 3, data yang sudah diberi label dari sejumlah dataset hasil *crawling* (3222 data) yang nanti hendak dipakai menjadi data latih untuk pengklasifikasian *Naïve Bayes*.

Tabel 7. Hasil Labelling Data

Tahapan	Tweet	Sentiment
Labelling Data	Soal & pembahasan SKB CPNS 2024 terupdate ada beberapa formasi jabatan; Auditor ahli pertama Auditor terampil Analisis anggaran Analisis hukum Analisis kebencanaan Analisis kebijakan Analisis keuangan pusat & daerah Analisis perkara peradilan Analisis SDM aparatur Ahli gizi Dokter Instruktur	Positif
	Boleh nanyain soal tes cpns gak ye Aku mah lebih suka upchar lah emang ada yang suka diving	Negatif

3.4 Pembagian Data

Tahapan berikutnya yakni tahapan mencari nilai akurasi pada data sentiment dengan pemisahan antara data pelatihan (training data) dengan data pengujian (testing data). Pembagian data ini mencakup 70% dipakai sebagai data pelatihan dan 30% untuk data pengujian. Prosedur ini sangat penting untuk mencegah terjadinya *overfitting*, yaitu kondisi pada saat model tertentu sesuai dengan data pelatihan sehingga tidak mampu diterapkan pada data baru.

3.5 Algoritma *Naïve Bayes*

Pada studi ini, Algoritma *Naïve Bayes* digunakan agar dapat mengklasifikasikan data sentimen pada aplikasi X. Proses ini melibatkan input berupa data latih maupun data uji. Data latih fungsinya menjadi masukan yang mencakup sentimen positif maupun negatif yang selanjutnya dilakukan klasifikasi dengan

4. KESIMPULAN

Penelitian ini berhasil menganalisis sentimen masyarakat terhadap seleksi CPNS tahun 2024 dengan memanfaatkan algoritma *Naïve Bayes*. Hasil analisis memperlihatkan bahwasanya model ini mampu mencapai akurasi sebesar 82.02%, dengan nilai F1-score untuk kategori sentimen positif mencapai 89% dan untuk kategori negatif sebesar 49%.

Walaupun model menunjukkan kinerja yang efektif dalam mengenali sentimen positif, terdapat tantangan dalam menangkap sentimen negatif secara menyeluruh, yang tergambar dari nilai recall yang semakin rendah. Hasil ini memberikan pemahaman yang signifikan bagi pemerintah untuk mengerti persepsi masyarakat dan merancang pendekatan komunikasi yang lebih efisien dalam proses seleksi CPNS.

Studi ini membawa harapan agar mampu berperan sebagai dasar dalam pengembangan kebijakan yang lebih responsif terhadap kebutuhan masyarakat dan memberikan kontribusi signifikan dalam evaluasi proses seleksi CPNS tahun 2024. Temuan studi ini juga mampu berperan sebagai sumber inspirasi bagi studi serupa di masa mendatang, baik dalam konteks seleksi pegawai pemerintah maupun analisis kebijakan publik lainnya.

Peneliti mengharapkan untuk penelitian selanjutnya dapat mengeksplorasi penggunaan metode lain seperti LSTM atau BERT untuk meningkatkan akurasi dalam analisis sentimen berbasis teks.

DAFTAR PUSTAKA

- [1] V. A. Sulistiani and M. Hamka, "Analisis Sentimen Pengguna Media Sosial Terhadap Identitas Kependudukan Digital Menggunakan Metode Support Vector Machine (SVM)," *J. Inf. Syst. ...*, vol. 9, no. 4, pp. 2185–2195, 2024, [Online]. Available: <http://ejurnal.seminar-id.com/index.php/josh/article/view/5614%0Ahttps://ejurnal.seminar-id.com/index.php/josh/article/download/5614/2906>
- [2] N. W. A. S. Aprilia and A. R. Isnain, "Analisis Sentimen Terhadap Media Sosial Twitter dengan Kasus Kampanye Anti-Korupsi di Indonesia Menggunakan Naive Bayes," *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 695, 2024, doi: 10.30865/mib.v8i2.7582.
- [3] S. Styawati, A. R. Isnain, N. Hendrastuty, and L. Andraini, "Comparison of Support Vector Machine and Naïve Bayes on Twitter Data Sentiment Analysis," *J. Inform. J. Pengemb. IT*, vol. 6, no. 1, pp. 56–60, 2021, doi: 10.30591/jpit.v6i1.3245.
- [4] U. Khaira, R. Aryani, and R. W. Hardian, "Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Kebijakan Kemdikbudristek Mengenai Kuota Internet Selama Covid-19," *J. Process.*, vol. 18, no. 2, pp. 272–285, 2023, doi: 10.33998/processor.2023.18.2.897.
- [5] F. A. Artanto, F. Teknik, U. Muhammadiyah, and P. Pekalongan, "Implementasi Algoritma Random Forest dan Model Bag of Words Dalam Analisis Sentimen Mengenai E-Materai," vol. 4, no. 2, pp. 139–145, 2024, doi: 10.54259/satesi.v4i2.3240.
- [6] F. S. Pamungkas and I. Kharisudin, "Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter," *Pros. Semin. Nas. Mat.*, vol. 4, pp. 1–7, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/prisma/article/view/45038>
- [7] Fuad Amirullah, Syariful Alam, and M. Imam Sulistyoso, "Analisis Sentimen Terhadap Kinerja KPU Menjelang Pemilu 2024 Berdasarkan Opini Twitter Menggunakan Naïve Bayes," *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 2, no. 3, pp. 69–76, 2023, doi: 10.55123/storage.v2i3.2293.
- [8] 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.
- [9] R. I. Lutfi, B. S. Rintyarna, and W. Suharso, "Mapping Opini Publik Terhadap Calon PNS Yang Mundur Setelah Lolos Dengan Teknik Berbasis Multinomial Naive Bayes," *J. Smart Teknol.*, vol. 5, no. 1, pp. 127–135, 2023.
- [10] Fitriani Fitriani, Ema Utami, and Anggit Dwi Hartanto, "Analisis Sentimen Masyarakat Terhadap Pelaksanaan P3K Guru Dengan Algoritma Naive Bayes Dan Decision Tree," *Tek. Teknol. Inf. dan Multimed.*, vol. 3, no. 1, pp. 23–30, 2022, doi: 10.46764/teknimedia.v3i1.53.
- [11] A. Septiarini, Rizqi Saputra, Andi Tejawati, and Masna Wati, "Deteksi Sarung Samarinda Menggunakan Metode Naive Bayes Berbasis Pengolahan Citra," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 5, pp. 927–935, 2021, doi: 10.29207/resti.v5i5.3435.
- [12] F. Fitriani, "Comparison of Naïve Bayes and Knn Algorithms in the Analysis of Public Sentiment Regarding the Implementation of Government Employees With Teacher Work Agreements," *Bull. Netw. Eng. Informatics*, vol. 1, no. 1, p. 18, 2023, doi: 10.59688/bufnets.v1i1.6.
- [13] S. Mandasari, B. H. Hayadi, and R. Gunawan, "Analisis Sentimen Pengguna Transportasi Online

- Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naive Bayes Classifier,” *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 5, no. 2, p. 118, 2022, doi: 10.53513/jsk.v5i2.5635.
- [14] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, “Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier,” *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.
- [15] T. Krisdiyanto, “Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naive Bayes Clasifiers,” *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 7, no. 1, p. 32, 2021, doi: 10.24014/coreit.v7i1.12945.
- [16] W. Astuti, E. Firasari, F. Lia Dwi Cahyanti, and F. Sarasati, “Analysis Sentiment on the Acceptance of Cpnas 2021 on Twitter Social Media Using Textblob,” *J. Comput. Inf. Technol. As an Accredit. J. Rank*, vol. 19, no. 1, pp. 15–21, 2022, [Online]. Available: <https://doi.org/10.33480/techno.v19i1.2980>
- [17] A. R. Isnain, A. I. Sakti, D. Alita, and N. S. Marga, “Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm,” *J. Data Min. dan Sist. Inf.*, vol. 2, no. 1, p. 31, 2021, doi: 10.33365/jdmsi.v2i1.1021.
- [18] S. Lestari and S. Saepudin, “Analisis Sentimen Vaksin Sinovac Pada Twitter Menggunakan Algoritma Naive Bayes,” *Semin. Nas. Sist. Inf. dan Manaj. Inform.*, pp. 163–170, 2021, [Online]. Available: <https://vaksin.kemkes.go.id/>
- [19] D. P. Artanti, A. Syukur, A. Prihandono, and D. R. I. M. Setiadi, “Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naive Bayes,” pp. 8–9, 2018.
- [20] K. Anwar, “Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone Menggunakan Naive Bayes,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 4, pp. 148–155, 2022, doi: 10.30865/klik.v2i4.315.
- [21] F. Ratnawati, “Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter,” *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, p. 50, 2018, doi: 10.35314/isi.v3i1.335.
- [22] A. Nugroho and Y. Religia, “Analisis Optimasi Algoritma Klasifikasi Naive Bayes menggunakan Genetic Algorithm dan Bagging,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 3, pp. 504–510, 2021, doi: 10.29207/resti.v5i3.3067.
- [23] S. A. Aaputra, Didi Rosiyadi, Windu Gata, and Syepri Maulana Husain, “Sentiment Analysis Analysis of E-Wallet Sentiments on Google Play Using the Naive Bayes Algorithm Based on Particle Swarm Optimization,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 377–382, 2019, doi: 10.29207/resti.v3i3.1118.
- [24] A. R. Isnain, H. Sulistiani, B. M. Hurohman, A. Nurkholis, and S. Styawati, “Analisis Perbandingan Algoritma LSTM dan Naive Bayes untuk Analisis Sentimen,” *J. Edukasi dan Penelit. Inform.*, vol. 8, no. 2, p. 299, 2022, doi: 10.26418/jp.v8i2.54704.
- [25] A. R. Isnain, N. S. Marga, and D. Alita, “Sentiment Analysis Of Government Policy On Corona Case Using Naive Bayes Algorithm,” *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 15, no. 1, p. 55, 2021, doi: 10.22146/ijccs.60718.
- [26] E. Laia and M. Yamin, “Penerapan Algoritma Naive Bayes dalam Menganalisis Sentimen pada Review Pengguna E-Commerce,” *Media Online*, vol. 4, no. 1, pp. 305–316, 2023, doi: 10.30865/klik.v4i1.1186.
- [27] I. Oktavia and A. R. Isnain, “Analisis Sentimen Opini Terhadap Tools Artificial Intelligence (AI) Berdasarkan Twitter Menggunakan Algoritma Naive Bayes,” *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 777, 2024, doi: 10.30865/mib.v8i2.7524.
- [28] Rayuwati, Husna Gemasih, and Irma Nizar, “IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK MEMREDIKSI TINGKAT PENYEBARAN COVID,” *Jural Ris. Rumpun Ilmu Tek.*, vol. 1, no. 1, pp. 38–46, 2022, doi: 10.55606/jurritek.v1i1.127.
- [29] C. C. Dan and L. Saufa Yardha, “Analisis Sentimen Pengguna Aplikasi Mobile Jkn Menggunakan Algoritma Naive Bayes,” *J. Sci. Soc. Res.*, vol. 4307, no. 2, pp. 555–563, 2024, [Online]. Available: <http://jurnal.goretanpena.com/index.php/JSSR>
- [30] Rita Retnosari, “Analisa kelayakan kredit usaha mikro berjalan pada perbankan dengan metode naive bayes,” *PROSISKO J. Pengemb. Ris. dan Obs. Sist. Komput.*, vol. 8, no. 1, pp. 53–59, 2021, doi: 10.30656/prosisko.v8i1.2848.
- [31] A. Aziz and I. Fitri, “Analisis Sentimen Terhadap Kebijakan Pemerintah ... Metode Naive Bayes (Abdul Aziz) |842 Teknik Informatika, Fakultas Teknologi Komunikasi dan Informatika,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, p. 7806700, 2021.
- [32] M. Zalukhu, “Analsis dan Implementasi Metode Naive Bayes dan SVM Pada Sentimen Pemilihan Calon Presiden RI Info Artikel Abstrak,” *KETIK J. Inform.*, vol. 1, no. 1, pp. 18–26, 2023, [Online]. Available: <https://jurnal.faatuatua.com/index.php/KETIK>