

Analisis Sentimen Komentar Media Sosial Twitter Terhadap Tes CPNS dengan Algoritma Naive Bayes

Rifki Nurhidayat, Nirwana Hendrastuty*

Fakultas Teknik dan Ilmu Komputer, Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ¹rifki_nurhidayat@teknokrat.ac.id, ^{2,*}nirwanahendrastuty@teknokrat.ac.id

Email Penulis Korespondensi: nirwanahendrastuty@teknokrat.ac.id

Submitted: 29/10/2024; Accepted: 01/12/2024; Published: 03/12/2024

Abstrak—Calon Pegawai Negeri Sipil (CPNS) merupakan salah satu karier yang paling diminati di Indonesia, dengan jumlah pelamar yang terus meningkat setiap tahunnya. Proses seleksi CPNS menarik perhatian publik dan memicu beragam opini, baik positif maupun negatif, yang banyak disampaikan melalui media sosial seperti Twitter. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap proses seleksi CPNS menggunakan algoritma *Naive Bayes*. Data yang digunakan dalam penelitian ini terdiri dari 5.599 komentar di Twitter, dengan komposisi 5.269 data sentimen negatif dan 323 data sentimen positif. Pengujian dilakukan menggunakan beberapa rasio pembagian data, yaitu 80:20, 70:30, 90:10, dan 50:50. Hasil menunjukkan bahwa rasio 70:30 memberikan akurasi terbaik, yaitu sebesar 95%. Namun, ketidakseimbangan data menyebabkan model lebih fokus pada sentimen negatif. Untuk mengatasi masalah ini, teknik Synthetic Minority Over-sampling Technique (SMOTE) diterapkan, yang berhasil meningkatkan kinerja model dalam mengklasifikasikan data positif, dengan precision dan recall mencapai 85-98%. Setelah penerapan SMOTE, akurasi keseluruhan menurun sedikit menjadi 91% pada rasio 80:20, 70:30, dan 90:10, tetapi model menjadi lebih efektif dalam mendeteksi kedua sentimen. Hasil penelitian ini memberikan wawasan mengenai pandangan publik terhadap seleksi CPNS dan dapat digunakan oleh pemerintah untuk meningkatkan proses seleksi di masa mendatang. Dengan pendekatan ini, diharapkan instansi pemerintahan dapat lebih baik memahami persepsi masyarakat serta mengoptimalkan sistem rekrutmen yang lebih transparan dan adil.

Kata Kunci: CPNS, Twitter, Naive Bayes, SMOTE, Sentimen Masyarakat.

Abstract—The Calon Pegawai Negeri Sipil (CPNS) is one of the most sought-after careers in Indonesia, with the number of applicants increasing every year. The CPNS selection process attracts public attention and triggers various opinions, both positive and negative, which are widely conveyed through social media such as Twitter. This research aims to analyze public sentiment towards the CPNS selection process using the Naive Bayes algorithm. The data used in this study consists of 5,599 comments on Twitter, with a composition of 5,269 negative sentiment data and 323 positive sentiment data. Tests were conducted using several data sharing ratios, namely 80:20, 70:30, 90:10, and 50:50. The results show that the 70:30 ratio provides the best accuracy, which is 95%. However, data imbalance causes the model to focus more on negative sentiment. To address this issue, the Synthetic Minority Over-sampling Technique (SMOTE) was applied, which successfully improved the model's performance in classifying positive data, with precision and recall reaching 85-98%. After the application of SMOTE, the overall accuracy decreased slightly to 91% at 80:20, 70:30, and 90:10 ratios, but the model became more effective in detecting both sentiments. The results of this study provide insight into the public's views on CPNS selection and can be used by the government to improve the selection process in the future. With this approach, it is expected that government agencies can better understand public perceptions and optimize a more transparent and fair recruitment system.

Keywords: CPNS, Twitter, Naive Bayes, SMOTE, Community Sentiment.

1. PENDAHULUAN

Pegawai Negeri Sipil (PNS) merupakan warga negara Indonesia yang telah memenuhi syarat tertentu dan diangkat oleh pejabat berwenang untuk menduduki suatu jabatan dengan hak atas penghasilan yang diatur berdasarkan ketentuan perundang-undangan yang berlaku. Karier sebagai PNS menjadi salah satu pilihan yang sangat diminati oleh berbagai kalangan di Indonesia, khususnya lulusan baru (fresh graduate) serta alumni dari jenjang pendidikan menengah hingga pendidikan tinggi. Hal ini menyebabkan persaingan dalam seleksi Calon Pegawai Negeri Sipil (CPNS) sangat ketat, di mana satu posisi dapat diperebutkan oleh ribuan pelamar setiap tahunnya[1].

Proses seleksi Calon Pegawai Negeri Sipil (CPNS) merupakan salah satu tahapan penting dalam sistem rekrutmen pegawai di Indonesia. Setiap tahunnya, proses seleksi CPNS menarik perhatian masyarakat luas, khususnya mereka yang berharap dapat mengisi posisi di instansi pemerintahan[2]. Salah satu platform yang kerap digunakan untuk mengekspresikan pandangan dan opini masyarakat terhadap proses tes CPNS adalah media sosial Twitter. Melalui platform ini, masyarakat menyampaikan beragam komentar, baik positif maupun negatif, terkait dengan berbagai aspek tes CPNS, mulai dari transparansi proses, kriteria seleksi, hingga hasil akhir tes. Oleh karena itu, analisis sentimen terhadap komentar-komentar ini menjadi sangat relevan untuk memahami persepsi publik terhadap tes CPNS[3].

Kemajuan teknologi saat ini berkembang sangat pesat, terutama dalam bidang internet. Perkembangan internet yang cepat telah membawa perubahan signifikan pada berbagai aspek komunikasi data komputer. Kehadiran internet sebagai sarana yang sangat dibutuhkan telah memicu munculnya berbagai platform, termasuk media sosial twitter, yang kini semakin mudah diakses dan digunakan[4]. Twitter menjadi salah satu platform media sosial yang sering digunakan sebagai wadah diskusi seputar berbagai topik, termasuk pelaksanaan seleksi CPNS di Indonesia. Fitur-fitur Twitter yang memungkinkan pengguna membuat postingan publik dan menyebarkannya dengan cepat menjadikannya tempat yang ideal untuk berbagi informasi dan pendapat mengenai proses seleksi CPNS. Berdasarkan hal tersebut,

peneliti memilih Twitter sebagai platform untuk penelitian karena menyediakan kumpulan data teks yang kaya, yang dapat digunakan untuk menganalisis sentimen masyarakat terhadap seleksi CPNS[5].

Analisis sentimen merupakan proses yang bertujuan untuk memahami, mengevaluasi, dan mengklasifikasikan apakah sentimen yang terkandung dalam sebuah teks bersifat positif atau negatif. Proses ini dilakukan dengan menggunakan teknik analisis teks yang menggunakan berbagai algoritma komputasi untuk mengidentifikasi pola dan makna dalam kalimat. Analisis sentimen dapat secara sistematis mengungkap emosi tersembunyi di balik opini, komentar, dan penilaian pengguna. Teknik ini dapat membantu memahami persepsi masyarakat terhadap suatu produk, layanan, atau kebijakan tertentu, atau memberikan wawasan yang lebih mendalam untuk pengambilan keputusan berdasarkan data yang dihasilkan dari interaksi pengguna di media sosial dan platform lainnya, Analisis sentimen juga digunakan untuk menganalisis tren sosial dan opini publik agar perusahaan dan institusi dapat meningkatkan kualitas layanan dan produknya berdasarkan masukan dari masyarakat[6].

Beragam penelitian tentang analisis sentimen dengan menggunakan metode Naïve Bayes, salah satu studi yang menganalisis Mapping Opini Publik Terhadap Calon PNS Yang Mundur Setelah Lolos Dengan Teknik Berbasis Multinomial Naive Bayes, didapatkan hasil performa akurasi, presisi, recall, f-measure tertinggi yang dihasilkan metode Multinomial Naïve Bayes tanpa seleksi fitur Chi-Square menghasilkan nilai 55%. Sedangkan hasil performa akurasi tertinggi metode Multinomial Naïve Bayes menggunakan seleksi fitur Chi-Square dengan 400 fitur diperoleh nilai 82%, nilai presisi 82%, nilai recall sebesar 81%, nilai f-measure sebesar 82% [7]. Pada penelitian lain yang serupa melakukan Analisis Sentimen Penerimaan Cpnns Tahun 2021 Pada Media Sosial Twitter Menggunakan Textblob, didapatkan hasil akurasi yang dicapai adalah 89,70% data analisis sentimen positif, negatif, dan netral terkait Asumsi CPNS 2021. 32% data tweet negatif sesuai dengan ekspektasi. Data tweet negatif yang termasuk dalam prediksi positif adalah 68%, dan data tweet negatif yang diprediksi netral adalah 0% [8].

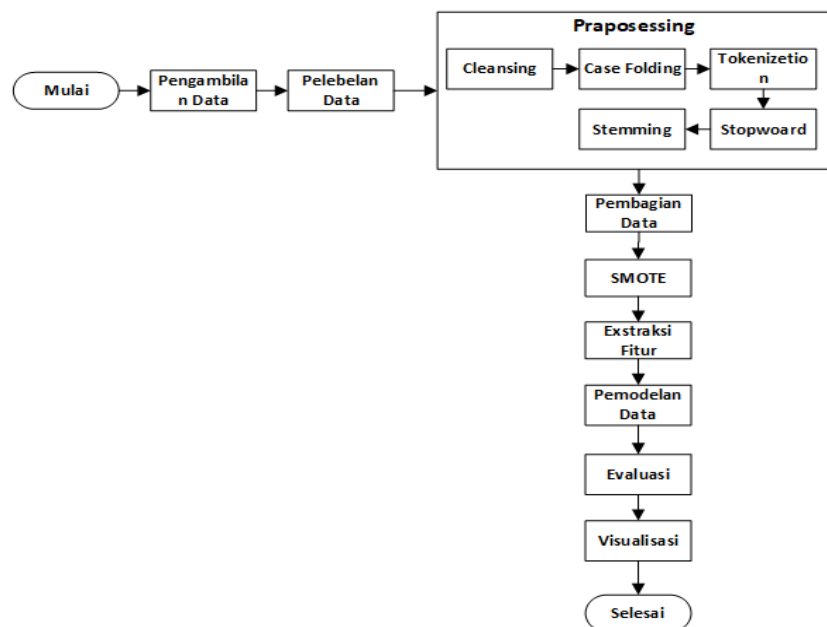
Penelitian yang serupa juga membahas Analisis Sentimen pandangan public terhadap profesi PNS (Pegawai Negeri Sipil) dari Twitter menerapkan Indonesian Roberta Base Sentiment Classifier, Didapatlah sebuah kesimpulan bahwa dari 394 data tweet yang telah di label terdapat 4.8% positif, 8.6% negatif dan netral 86.4% dan tidak hanya itu setelah melakukan uji coba pelabelan pada penelitian kali ternyata prediksi yang dilakukan memiliki akurasi yang baik dengan rata-rata keseluruhan akurasi yang didapat sebesar 90% hasil[9]. Pada penelitian Analisis Sentimen Masyarakat Terhadap Pelaksanaan P3k Guru Dengan Algoritma Naive Bayes dan decision Tree didapatkan hasil bahwa Algoritma Decision Tree, dan Nave Bayessemuanya dapat digunakan dalam penambangan data, seperti yang ditunjukkan dalam penelitian ini. Analisis sentimen pada data Twitter tentang guru PPPK menghasilkan tingkat akurasi nave bayes sebesar 100,00 persen. Dimana presisi grup untuk pred. negatif adalah 100,00% dan pred. positif adalah 100,00%, tingkat akurasi untuk pendekatan decision tree adalah 53,95 persen. Pred negatif adalah 0,00 persen dan pred positif adalah 53,95 persen[10]. Penelitian yang membahas Analisis Sentimen Calon Pegawai Negeri Sipil Indonesia Tahun 2023 Melalui Jejaring Sosial Dengan Metode Naive bayes algorithm Temuan penelitian ini menunjukkan bahwa media sosial, khususnya Twitter, memberikan pengaruh yang cukup besar terhadap sentimen publik pada tahun 2023 mengenai prosedur seleksi calon pegawai negeri sipil (CPNS)[11].

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap pelaksanaan tes CPNS dengan menggunakan algoritma Naive Bayes. Penelitian ini bertujuan untuk mengklasifikasikan sentimen positif dan negatif dari komentar-komentar yang diunggah di media sosial Twitter[12]. Dengan penerapan algoritma Naive Bayes, penelitian ini diharapkan dapat memberikan pemahaman yang lebih mendalam tentang persepsi publik terhadap proses seleksi CPNS, serta mengidentifikasi pola-pola sentimen yang dominan. Hasil penelitian ini juga diharapkan dapat memberikan kontribusi bagi pihak terkait dalam mengevaluasi dan meningkatkan kualitas pelaksanaan tes CPNS di masa mendatang[13].

2. METODOLOGI PENELITIAN

2.1 Tahapan Analisis Sentimen

Penelitian ini dilaksanakan melalui sejumlah tahapan yang dimulai dari tahap persiapan, di mana data mengenai CPNS dikumpulkan atau dicrawling dari platform media sosial Twitter (X) dengan tujuan mendapatkan opini publik yang relevan. Setelah data terkumpul, langkah berikutnya adalah proses pembersihan data untuk memastikan konsistensi serta keakuratan informasi yang ada, yang kemudian dilanjutkan dengan pemberian label pada setiap entri data agar dapat diklasifikasikan sesuai dengan kebutuhan analisis. Tahap berikutnya adalah pre-processing data, yang mencakup sejumlah teknik seperti cleansing untuk menghilangkan karakter tidak relevan, case folding untuk mengonversi seluruh teks menjadi huruf kecil, tokenisasi untuk memecah kalimat menjadi kata-kata individual, penghapusan stopword untuk menghilangkan kata-kata umum yang tidak memiliki makna signifikan, dan stemming untuk menyederhanakan kata ke bentuk dasarnya. Setelah melalui proses ini, data yang tidak seimbang diseimbangkan kembali menggunakan metode SMOTE (Synthetic Minority Over-sampling Technique) untuk memastikan distribusi kelas yang lebih merata, sehingga dapat mendukung performa model yang lebih optimal. Selanjutnya, data yang sudah siap dimasukkan ke tahap pemodelan dengan menggunakan algoritma Naïve Bayes, dilanjutkan dengan proses evaluasi untuk menilai akurasi model, dan diakhiri dengan visualisasi hasil untuk mempermudah interpretasi dan pemahaman terhadap temuan analisis yang dilakukan. Detail yang lebih lengkap mengenai metode penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian

2.2 Pengumpulan Data

Pada penelitian ini data di peroleh dengan metode crawling. Proses ini dilakukan menggunakan bahasa pemrograman Python yang dijalankan di platform Google Colaboratory[14]. Proses pengumpulan data dilakukan melalui langkah code program yang melibatkan penggunaan kata kunci untuk mencari tweet opini publik tentang tes CPNS yang sesuai dengan preferensi yang diinginkan[15].

2.3 Pelebelan Data

Pada tahap ini pemberian pelabelan pada setiap tweet. Pelabelan ini adalah proses klasifikasi data yang bertujuan untuk mengidentifikasi apakah setiap kalimat dalam dataset bersifat positif atau negatif. Label ditentukan melalui analisis teks untuk menentukan skor setimen positif atau negatif. Dengan demikian, pelabelan positif dan negatif bisa dilakukan dengan lebih presisi[16]. Dapat dilihat pada persamaan 1.

$$S_{sentiment} = \begin{cases} S_{positive} & \text{if } S_{positive} > S_{negative} \\ S_{negative} & \text{if } S_{positive} < S_{negative} \end{cases} \quad (1)$$

Dimana $S_{sentiment}$, $S_{positive}$, dan $S_{negative}$. $S_{sentiment}$ merujuk pada kalimat yang mengandung sentimen dalam dataset, baik positif maupun negatif. $S_{positive}$ adalah kalimat yang mengandung sentimen positif pada dataset, sedangkan $S_{negative}$ mengacu pada kalimat yang mengandung sentimen negatif. Ketiga definisi ini membantu dalam klasifikasi dan analisis sentimen pada data yang dikumpulkan, khususnya dalam konteks analisis sentimen komentar mengenai CPNS di media sosial Twitter.

2.4 Preprocessing Data

Tahap di mana data disaring dan diperbaiki disebut dengan preprocessing. Ketika data dikumpulkan, informasinya sering kali tidak terstruktur dan mengandung banyak karakter[17]. Dalam proses preprosesing menggunakan teknik Cleansing, Case Folding, Tokenizing, Stopword, dan Stemming[18]. Tahapan pada penelitian ini sebagai berikut:

2.4.1 Cleansing Data

Cleansing data adalah langkah penting dalam pengolahan data, di mana proses ini bertujuan untuk menghilangkan data yang tidak sesuai atau tidak relevan seperti emoji, hastag, huruf, angka dan spasi. Tahap ini membantu memastikan bahwa data yang digunakan lebih akurat dan dapat diandalkan untuk analisis lebih lanjut. Dengan menghapus inkonsistensi dan informasi yang tidak diperlukan, hasil analisis dapat menjadi lebih valid dan efektif[19].

2.4.2 Case Folding

Case folding adalah langkah untuk mengonversi semua huruf dalam teks menjadi huruf kecil, sekaligus menghapus tanda baca dan angka. Proses ini hanya memproses huruf dari "a" hingga "z", sehingga karakter lain yang tidak termasuk dalam rentang ini akan dihilangkan[20].

2.4.3 Tokenizing

Tokenizing berkaitan dengan pemisahan data teks mentah menjadi beberapa token yang terpisah, yang nantinya akan berupa kata atau karakter individual. Terdapat dua jenis tokenisasi: pertama, tokenisasi kalimat. Proses ini bertujuan untuk membagi dokumen teks menjadi sejumlah kalimat yang bermakna. Operasi ini membantu dalam memisahkan teks menjadi unit-unit kalimat dengan mendeteksi tanda-tanda pemisah, seperti titik (.) dan karakter baris baru (\n). Kedua, tokenisasi kata, yaitu proses pemecahan kalimat menjadi kata-kata individual, yang disebut token. Teknik ini bertujuan menemukan pemisah antara kata-kata, seperti titik atau spasi, sehingga setiap kata dapat dipisahkan dan diubah menjadi huruf kecil[21].

2.4.4 Stopword

Stopword adalah proses di mana kata-kata yang dianggap termasuk dalam kategori stopwords dihapus. Stopword itu sendiri merupakan kata-kata yang sering muncul, namun dipandang tidak memiliki makna penting. Contohnya : dan, ke, dari, di, yang, adalah, atau, itu[22].

2.4.5 Stemming

Stemming adalah proses penghapusan imbuhan dari kata-kata dalam kalimat laporan, atau mengubah kata kerja menjadi kata benda. Sebagai contoh, ketika imbuhan kata “diberikan”, “di-“ dan “-kan” dihapus, kata tersebut berubah menjadi “beri”[23].

2.5 Pembagian Data

Langkah berikutnya dalam penelitian ini adalah membagi data ke dalam dua kelompok, yaitu data latih (training data) dan data uji (testing data). Pada penelitian ini menggunakan empat perbandingan yaitu 50:50, 90:10, 70:30, dan 80:20. Setelah pembagian ini dilakukan, data latih akan digunakan untuk melatih model, sedangkan data uji akan dipakai untuk menguji kinerja model yang telah dibentuk. Dengan demikian, proses pemodelan dapat dimulai dan hasilnya dapat dievaluasi berdasarkan performa model terhadap data uji[24].

2.6 SMOTE

SMOTE adalah metode yang dikenal sebagai pendekatan yang efektif untuk mengatasi ketidakseimbangan data dan telah berhasil diterapkan dalam berbagai kasus. Dalam analisis sentimen, seringkali jumlah data dengan sentimen positif dan negatif tidak merata, yang dapat menyebabkan model lebih condong ke kelas yang dominan. SMOTE membantu dengan menciptakan sampel sintetis untuk kelas minoritas, sehingga distribusi data menjadi lebih seimbang[25].

2.7 Ekstraksi Fitur

Proses ekstraksi fitur diperlukan karena algoritma machine learning tidak mampu memahami karakter dan kata-kata dalam dataset. Oleh karena itu, data teks harus direpresentasikan dalam bentuk angka agar dapat dimengerti. Metode TF-IDF merupakan algoritma yang berfungsi untuk menghitung bobot dari setiap kata yang muncul dengan frekuensi tinggi. Algoritma ini umumnya digunakan untuk memberikan ciri khas pada dokumen. Dengan metode ini, kita dapat menilai seberapa sering suatu kata muncul dalam dokumen tertentu. Term Frequency (TF) dihitung dengan cara mengukur jumlah kemunculan kata i dalam data j , lalu hasilnya dibagi dengan total istilah yang ada dalam data j . Berikut ini adalah rumus yang digunakan untuk menentukan nilai TF[26].

Persamaan 1

$$IDF_t = \log \left(\frac{d}{df} \right) \quad (2)$$

IDF (Inverse Document Frequency) adalah metrik yang digunakan untuk mengukur seberapa penting suatu kata dalam kumpulan dokumen. Rumus IDF sering ditulis sebagai $IDF_t = \log \left(\frac{d}{df} \right)$, di mana t adalah kata ke- t dari kata kunci, D adalah total jumlah dokumen dalam korpus, dan df adalah jumlah dokumen yang mengandung kata yang dicari. Semakin sedikit dokumen yang mengandung kata tersebut, semakin tinggi nilai IDF, yang menunjukkan bahwa kata tersebut lebih signifikan dalam membedakan dokumen.

2.8 Pembangunan Model Naïve Bayes

Dalam tahap pembangunan, proses klasifikasi data dilakukan dengan menggunakan metode Naïve Bayes. Model ini dibangun menggunakan bahasa pemrograman Python bersama dengan library sklearn. Berikut adalah penjelasan lebih mendetail mengenai Igoritma Naïve Bayes tersebut: Naïve Bayes merupakan metode klasifikasi linear yang terkenal akan efisiensinya. Pendekatan ini didasari oleh model probabilitas dan klasifikasi yang berpijak pada teorema Bayes. Istilah "naive" sendiri merujuk pada asumsi bahwa fitur-fitur dalam dataset dianggap tidak saling bergantung satu sama lain[27]. Metode Naïve Bayes didasarkan pada teorema Bayes, berfungsi untuk melakukan klasifikasi kelas, serupa dengan yang dihasilkan oleh pohon keputusan. Metode ini telah terbukti memberikan hasil akurasi yang lebih baik serta kecepatan yang lebih tinggi saat dilakukan pengujian pada dataset yang lebih besa[28]. Berikut merupakan teorema Bayes yang dapat dilihat pada persamaan 3



$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \tag{3}$$

Dimana X adalah data dengan kelas yang belum diketahui, dan H merupakan hipotesis bahwa data tersebut termasuk dalam suatu kelas spesifik. Probabilitas hipotesis H berdasarkan data X, atau disebut posteriori probabilitas, dinyatakan sebagai P(H|X). Probabilitas hipotesis H tanpa mempertimbangkan data (prior probabilitas) ditulis sebagai P(H). Selanjutnya, P(X|H) menggambarkan probabilitas munculnya data X dengan asumsi bahwa hipotesis H benar, dan P(X) adalah probabilitas dari data X itu sendiri.

2.9 Evaluasi Model

Tahapan evaluasi dalam analisis sentimen melibatkan beberapa langkah penting. Langkah pertama adalah membagi dataset yang sudah diproses menjadi data pelatihan dan data uji. Model sentimen kemudian dilatih dengan data pelatihan tersebut. Setelah model siap, data uji digunakan untuk mengecek performanya. Metrik evaluasi seperti akurasi, presisi, recall, dan F1-score kemudian diaplikasikan untuk menilai sejauh mana model mampu mengklasifikasikan sentimen dengan baik. Hasil evaluasi ini dianalisis lebih lanjut untuk mengidentifikasi kelebihan, kekurangan, serta potensi perbaikan bagi model[29]. Akurasi diperoleh menggunakan persamaan berikut.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{4}$$

$$Presisi = \frac{TP}{TP+FP} \tag{5}$$

$$Recall = \frac{TP}{TP+FN} \tag{6}$$

Dalam evaluasi model, terdapat beberapa metrik penting, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). True Positive (TP) menunjukkan jumlah kasus sentimen positif yang benar-benar terprediksi positif oleh model, sedangkan True Negative (TN) menunjukkan jumlah kasus sentimen negatif yang benar-benar terprediksi negatif oleh model. False Positive (FP) adalah jumlah kasus yang sebenarnya negatif tetapi salah diprediksi sebagai positif, dan False Negative (FN) adalah jumlah kasus yang sebenarnya positif tetapi salah diprediksi sebagai negatif. Keempat metrik ini digunakan untuk menilai akurasi dan kinerja keseluruhan model dalam mengklasifikasikan sentimen.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pada penelitian ini, data yang berhasil dikumpulkan sebanyak 6.493 records dimulai pada 01 juli sampai 28 agustus dengan menggunakan teknik crawling. Terkait kata kunci yang dipakai untuk pengumpulan data, beberapa di antaranya adalah "Tes CPNS 2024", "Lulus CPNS", "Gagal CPNS", serta "CPNS". Setelah data dikumpulkan, hasilnya disimpan terlebih dahulu dalam format CSV untuk nantinya dipakai dalam analisis sentimen. Proses pengumpulan data diperoleh menggunakan kode twitter API yang dilakukan lewat Google Colab dengan memanfaatkan Python sebagai bahasa pemrograman. Hasil data yang di peroleh saat crawling ditunjukkan pada tabel 1.

Tabel 1. Hasil Sempel Pengumpulan Data Tweet

Username	Tweet
sanaguan	3 kali tes CPNS gak lulus. 2 kali tes lulus passing grade tp jatuhnya di SKB. Tes terakhir nangis banget soalnya nilaiku tinggi dibanding teman2 aku yg lulus cuman syangnya pesaingku rata2 top score. Alhasil gugurlah lagi dan tahun ini percobaan ke-4
chrdannas	@nowyoucatchme @woodumpies @worksfess mana ada tes cuma formalitas kalo tesnya berbasis CAT. celah dalam seleksi cpns terdapat di seleksi non-CAT kalo ga lulus SKD bayar ratusan juta sama aja ga lolos.

3.2 Pelebelan Data

Proses pelabelan sentimen dalam dataset ini dimulai dengan pengumpulan dan analisis teks seperti komentar atau ulasan. Data kemudian dibagi menjadi dua kategori, yaitu Positif dan Negatif, untuk mengidentifikasi sentimen yang terkandung. Pelabelan dilakukan menggunakan program Python di Google Colaboratory untuk efisiensi pengujian. Hasil analisis menunjukkan 5.269 data berlabel "Negatif" dan 323 data berlabel "Positif," yang menjadi dasar untuk analisis lebih lanjut terhadap sentimen dalam dataset. Hasil pelebelan dapat dilihat pada tabel 2.

Tabel 2. Hasil Pelabelan

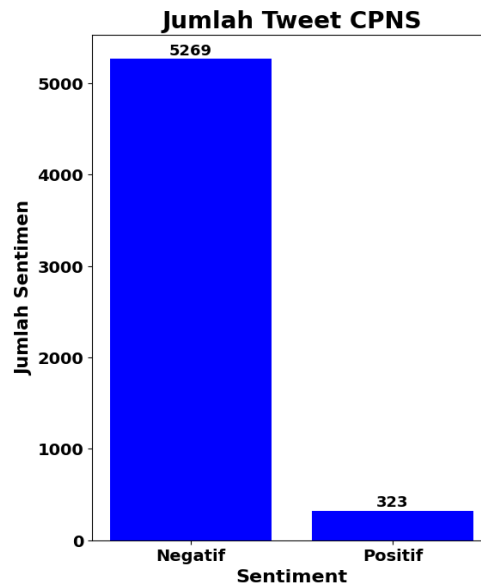
Tweet	Label
-------	-------

3 kali tes CPNS gak lulus. 2 kali tes lulus passing grade tp jatuhnya di SKB. Tes terakhir nangis banget soalnya nilaiku tinggi dibanding teman2 aku yg lulus cuman syangnya pesaingku rata2 top score. Alhasil gugurlah lagi dan tahun ini percobaan ke-4 @nowyoucatchme @woodumpies @worksfess mana ada tes cuma formalitas kalo tesnya berbasis CAT. celah dalam seleksi cpns terdapat di seleksi non-CAT kalo ga lulus SKD bayar ratusan juta sama aja ga lolos.

Positif

Negatif

Untuk memberikan gambaran yang lebih jelas tentang distribusi hasil pelabelan ini, tersedia diagram yang menunjukkan perbandingan jumlah data dalam masing-masing kategori, seperti yang ditampilkan pada Gambar 2.



Gambar 2. Grafik Hasil Labelling

3.3 Preprocessing

Pra-pemrosesan data pada analisis sentimen adalah langkah penting untuk membersihkan dan menyiapkan teks sebelum dianalisis. Tahapan ini meliputi case folding, yaitu mengubah semua teks menjadi huruf kecil, serta tokenisasi yang memecah teks menjadi kata-kata. Tanda baca, karakter spesial, URL, mention, dan hashtag dihapus karena tidak relevan. Stopwords, seperti kata "yang" atau "dan", juga dihilangkan untuk mengurangi kebisingan data. Proses stemming dan lemmatisasi digunakan untuk mengubah kata ke bentuk dasarnya, seperti "berlari" menjadi "lari". Selain itu, normalisasi dilakukan untuk menangani variasi penulisan atau kesalahan ketik, sehingga teks siap untuk dianalisis sentimennya secara lebih efektif. Tahapan dalam preprocessing data untuk analisis sentimen sebagai berikut:

3.3.1 Cleansing

Cleansing dalam analisis sentimen adalah langkah penting untuk merapikan teks dengan menghilangkan elemen-elemen yang tak diperlukan atau bisa mengganggu proses analisis. Proses ini mencakup beberapa tahap, mulai dari menghapus tanda baca, karakter spesial, angka, URL, mention seperti (@username), dan hashtag (#) yang sering terlihat di platform media sosial. Selain itu, kata-kata umum yang tidak membantu analisis, seperti "di", "ke", "dari", juga dihapus. Langkah lainnya melibatkan menghilangkan spasi berlebih, menormalkan teks untuk memperbaiki kesalahan ketik, serta menyamakan variasi penulisan kata, sehingga data jadi lebih rapi dan konsisten, memungkinkan algoritma bekerja lebih baik dalam mendeteksi sentimen. Hasil cleansing dapat dilihat pada tabel 3.

Tabel 3. Hasil Cleansing

Cleansing
kali tes CPNS gak lulus kali tes lulus passing grade tp jatuhnya di SKB Tes terakhir nangis banget soalnya nilaiku tinggi dibanding teman aku yg lulus cuman syangnya pesaingku rata top score Alhasil gugurlah lagi dan tahun ini percobaan ke worksfess mana ada tes cuma kalo tesnya berbasis CAT celah dalam seleksi cpns terdapat di seleksi nonCAT kalo ga lulus SKD bayar ratusan juta sama aja ga lolos

3.3.2 Case Folding

Case folding pada analisis sentimen adalah langkah pra-pemrosesan data yang bertujuan untuk mengubah semua huruf dalam teks menjadi huruf kecil (lowercase). Langkah ini dilakukan untuk menyamakan kata-kata yang ditulis dengan huruf kapital dan huruf kecil agar diperlakukan sama oleh algoritma. Misalnya, kata "Tes CPNS" dan "tes cpns" akan



dianggap sebagai satu entitas yang sama. Dengan melakukan case folding, variasi huruf kapital yang tidak mempengaruhi makna dapat dihilangkan, sehingga analisis sentimen menjadi lebih konsisten dan akurat. Hasil case folding dapat dilihat pada tabel 4.

Tabel 4. Hasil Case Folding

Case Folding
kali tes cpns gak lulus kali tes lulus passing grade tp jatuhnya di skb tes terakhir nangis banget soalnya nilaiku tinggi dibanding teman aku yg lulus cuman syangnya pesaingku rata top score alhasil gugurlah lagi dan tahun ini percobaan ke worksfess mana ada tes cuma kalo tesnya berbasis cat celah dalam seleksi cpns terdapat di seleksi noncat kalo ga lulus skd bayar ratusan juta sama aja ga lolos

3.3.3 Tokenizing

Tokenizing pada analisis sentimen adalah proses memecah teks menjadi unit-unit kecil yang disebut token, biasanya berupa kata, frasa, atau karakter. Tujuan dari tokenizing adalah untuk memudahkan pemrosesan teks dengan mengubah kalimat atau paragraf menjadi potongan-potongan yang lebih kecil, sehingga setiap token dapat dianalisis secara terpisah. Dalam analisis sentimen, tokenizing membantu algoritma memahami konteks setiap kata dan mengidentifikasi pola-pola emosi atau sentimen di dalam teks. Misalnya, kalimat " cara cek jumlah pelamar formasi yang lulus seleksi cpns" akan dipecah menjadi token: ['cara', 'cek', 'jumlah', 'pelamar', 'formasi', 'yang', 'lulus', 'seleksi', 'cpns']. Proses ini penting karena setiap token dapat membawa informasi yang signifikan untuk menentukan apakah teks mengandung sentimen positif dan negatif. Hasil tokenizing dapat dilihat pada tabel 5.

Tabel 5. Hasil Tokenizing

Tokenizing
['kali', 'tes', 'cpns', 'gak', 'lulus', 'kali', 'tes', 'lulus', 'passing', 'grade', 'tp', 'jatuhnya', 'di', 'skb', 'tes', 'terakhir', 'nangis', 'bangat', 'soalnya', 'nilaiku', 'tinggi', 'dibanding', 'teman', 'aku', 'yg', 'lulus', 'cuman', 'syangnya', 'pesaingku', 'rata', 'top', 'score', 'alhasil', 'gugurlah', 'lagi', 'dan', 'tahun', 'ini', 'percobaan', 'ke'] ['worksfess', 'mana', 'ada', 'tes', 'cuma', 'kalo', 'tesnya', 'berbasis', 'cat', 'celah', 'dalam', 'seleksi', 'cpns', 'terdapat', 'di', 'seleksi', 'noncat', 'kalo', 'ga', 'lulus', 'skd', 'bayar', 'ratusan', 'juta', 'sama', 'aja', 'ga', 'lolos']

3.3.4 Stopword

Stopwords pada analisis sentimen adalah kata-kata umum yang tidak memiliki makna signifikan dalam menentukan sentimen suatu teks. Kata-kata ini sering muncul dalam kalimat tetapi tidak memberikan kontribusi penting terhadap makna atau emosi teks. Contoh stopwords dalam bahasa Indonesia meliputi kata-kata seperti "yang", "dan", "di", "ke", "untuk", dan lain-lain. Dalam analisis sentimen, stopwords biasanya dihapus agar algoritma fokus pada kata-kata yang lebih relevan dalam menentukan apakah sentimen suatu teks bersifat positif dan negatif. Penghapusan stopwords membantu meningkatkan akurasi dan efisiensi analisis karena mengurangi gangguan dari kata-kata yang tidak berkontribusi langsung terhadap sentimen. Hasil Stopword dapat dilihat pada tabel 6.

Tabel 6. Hasil Stopword

Stopword
['kali', 'tes', 'cpns', 'gak', 'lulus', 'kali', 'tes', 'lulus', 'passing', 'grade', 'tp', 'jatuhnya', 'skb', 'tes', 'nangis', 'bangat', 'nilaiku', 'dibanding', 'teman', 'yg', 'lulus', 'cuman', 'syangnya', 'pesaingku', 'top', 'score', 'alhasil', 'gugurlah', 'percobaan'] ['worksfess', 'tes', 'kalo', 'tesnya', 'berbasis', 'cat', 'celah', 'seleksi', 'cpns', 'seleksi', 'noncat', 'kalo', 'ga', 'lulus', 'skd', 'bayar', 'ratusan', 'juta', 'aja', 'ga', 'lolos']

3.3.5 Stemming

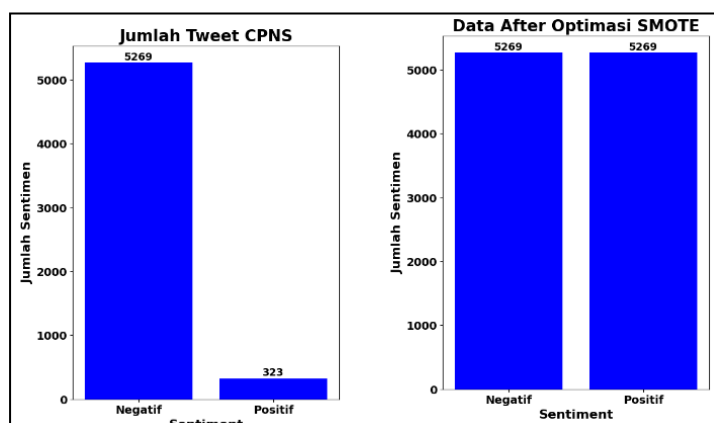
Stemming pada analisis sentimen adalah proses mengubah kata-kata menjadi bentuk dasar atau akarnya untuk mengurangi variasi kata yang memiliki arti serupa. Tujuan utamanya adalah menyederhanakan teks sehingga kata-kata dengan berbagai bentuk (misalnya, "diterima", "keterima") dikonversi menjadi satu bentuk dasar, yaitu "terima". Dengan cara ini, algoritma dapat lebih mudah mengidentifikasi pola sentimen dari kata-kata tersebut tanpa harus memperhitungkan berbagai variasi morfologisnya. Stemming membantu dalam mengurangi kompleksitas data, mempercepat proses analisis, dan meningkatkan akurasi dalam menentukan sentimen positif dan negatif. Hasil Stemming dapat dilihat pada tabel 7.

Tabel 7. Hasil Stemming

Stemming
kali tes cpns gak lulus kali tes lulus passing grade tp jatuh skb tes nang banget nilai banding teman yg lulus cuman syangnya saing top score alhasil gugur coba worksfess tes kalo tes bas cat celah seleksi cpns seleksi noncat kalo ga lulus skd bayar ratus juta aja ga lolos

3.4 Akurasi SMOTE

Setelah melalui proses preprocessing, data sentimen yang diperoleh terdiri dari 323 data positif dan 5.276 data negatif, menciptakan ketidakseimbangan yang berdampak signifikan pada kinerja algoritma Naïve Bayes. Ketidakseimbangan ini menyebabkan model cenderung belajar lebih banyak dari data negatif, sehingga akurasi dan performa klasifikasi untuk kategori negatif menjadi lebih tinggi, sementara kategori positif sering diabaikan. Oleh karena itu, penilaian terhadap model Naïve Bayes dalam penelitian ini tidak hanya berfokus pada akurasi, tetapi juga mencakup metrik lain seperti precision, recall, dan F1-Score untuk mengevaluasi efektivitas model secara keseluruhan. Untuk mengatasi masalah ketidakseimbangan data, teknik optimasi SMOTE (Synthetic Minority Over-sampling Technique) digunakan untuk meningkatkan jumlah data positif secara sintesis, menciptakan distribusi yang lebih seimbang antara kategori positif dan negatif. Hasil penerapan SMOTE dapat dilihat pada Gambar 3, yang menunjukkan peningkatan distribusi data dan performa model berdasarkan berbagai metrik evaluasi, sehingga diharapkan model Naïve Bayes dapat menghasilkan klasifikasi yang lebih akurat dan robust untuk kedua kategori sentimen.



Gambar 3. Hasil Optimasi SMOTE

Berdasarkan gambar 3, terlihat perbandingan antara data sebelum dan setelah optimasi SMOTE diterapkan. Dengan adanya optimasi ini, jumlah data untuk setiap kategori sentimen menjadi setara, sehingga dalam penelitian ini terdapat 5.269 data untuk sentimen positif dan negatif. Optimasi SMOTE memastikan model dapat mempelajari sentimen secara seimbang, tanpa ada ketidakseimbangan antara data minoritas dan mayoritas. Setelah data berhasil diseimbangkan, langkah selanjutnya adalah melatih dan menguji model algoritma agar dapat belajar lebih baik dari data baru yang telah dioptimalkan.

Dalam analisis sentimen terhadap tes CPNS menggunakan algoritma Naïve Bayes, data latih dan data uji dibagi dengan rasio yang bervariasi untuk melihat bagaimana model beradaptasi dengan proporsi data yang berbeda. Pada rasio 80:20, sebanyak 4.474 data dipakai untuk pelatihan, sementara 1.118 data digunakan untuk pengujian, memberi model lebih banyak peluang belajar. Dengan rasio 70:30, 3.914 data menjadi data latih dan 1.678 data untuk pengujian, yang memberikan lebih banyak variasi pada data uji. Di rasio 90:10, ada 5.033 data untuk pelatihan, dengan 559 data pengujian, memastikan bahwa model memiliki lebih banyak data untuk dilatih, meskipun data uji terbatas. Terakhir, rasio 50:50 membagi data secara merata, dengan 2.796 data untuk pelatihan dan 2.796 data untuk pengujian, menciptakan keseimbangan antara proses pelatihan dan pengujian. Namun, jumlah data latih yang lebih sedikit dapat memengaruhi kemampuan model dalam memahami pola sentimen. Perbedaan hasil sebelum dan sesudah SMOTE dapat dilihat pada tabel, yang menunjukkan dampak dari penyeimbangan data pada performa model.

Tabel 8. Hasil Sebelum SMOTE

Perbandingan	Modelling	Sebelum SMOTE			
		Accuracy	Precision	Recall	F1-Score
0.1	Positif	94%	94%	100%	97%
	Negatif		00%	00%	00%
0.2	Positif	94%	94%	100%	97%
	Negatif		00%	00%	00%
0.3	Positif	95%	95%	100%	97%
	Negatif		00%	00%	00%
0.5	Positif	94%	94%	100%	97%
	Negatif		00%	00%	00%

Sesudah penerapan SMOTE menunjukkan bahwa model mengalami peningkatan performa signifikan dalam mengklasifikasikan sentimen negatif setelah SMOTE diterapkan. Sebelum SMOTE, model memiliki akurasi yang tinggi pada data positif, dengan precision dan recall mencapai 94%-100%, serta F1-Score konsisten di 97%. Namun,

untuk data negatif, model gagal mengidentifikasi sentimen, dengan precision, recall, dan F1-Score sebesar 0%, menunjukkan adanya ketidak seimbangan data. Setelah penerapan SMOTE, meskipun akurasi keseluruhan sedikit menurun menjadi 89%-91%, model berhasil meningkatkan kemampuan dalam mengklasifikasikan data negatif, dengan precision, recall, dan F1-Score mencapai 85%-98%. Hal ini menandakan bahwa SMOTE mampu mengatasi ketidakseimbangan data, sehingga model dapat mengenali dan mengklasifikasikan kedua jenis sentimen, baik positif maupun negatif, dengan lebih seimbang dan akurat. Penurunan akurasi keseluruhan diimbangi oleh peningkatan signifikan pada metrik precision, recall, dan F1-Score, khususnya untuk data negatif, yang membuat model lebih efektif secara keseluruhan. Hasil sesudah penerapan SMOTE dapat di lihat pada table 9.

Tabel 9. Hasil Sesudah SMOTE

Perbandingan	Modelling	Sesudah SMOTE			
		Accuracy	Precision	Recall	F1-Score
0.1	Positif	91%	98%	83%	90%
	Negatif		85%	98%	91%
0.2	Positif	91%	98%	85%	91%
	Negatif		85%	98%	92%
0.3	Positif	91%	98%	83%	90%
	Negatif		85%	98%	91%
0.5	Positif	89%	97%	80%	88%
	Negatif		83%	90%	90%

Sebelum diterapkannya teknik penyeimbangan data seperti SMOTE (Synthetic Minority Over-sampling Technique), model yang digunakan untuk klasifikasi sering kali mengalami ketidakseimbangan dalam performa prediksinya, terutama ketika berhadapan dengan dataset yang memiliki distribusi kelas tidak merata. Pada kondisi ini, model cenderung lebih fokus pada kelas mayoritas, sehingga menyebabkan akurasi prediksi untuk kelas minoritas menjadi sangat rendah. Tabel di bawah ini menampilkan hasil confusion matrix pada beberapa perbandingan dataset sebelum diterapkannya metode SMOTE. Terlihat bahwa prediksi untuk kelas positif tidak mampu memprediksi dengan benar, yang disebabkan oleh dominasi kelas negatif di dalam dataset.

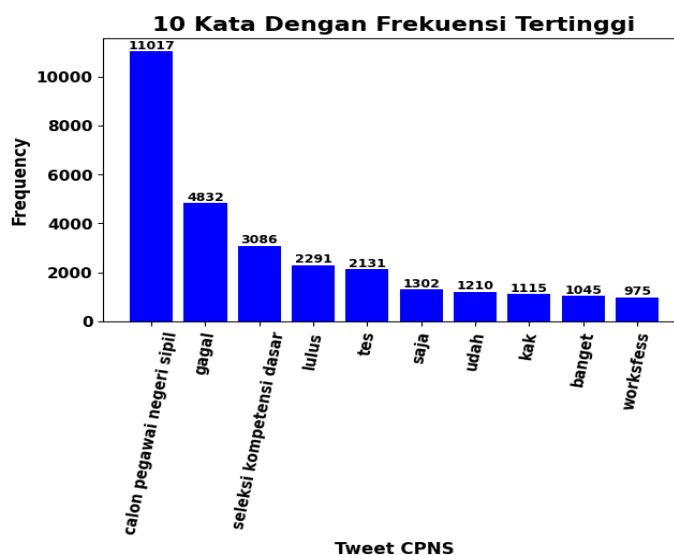
Tabel 10. Hasil Confusion Matrix Before SMOTE

Perbandingan	Prediction Class	Actual Class	
		Pred. Negative	Pred. Positive
0.1	Negatif	529	0
	Positif	31	0
0.2	Negatif	1048	0
	Positif	71	0
0.3	Negatif	1588	0
	Positif	90	0
0.5	Negatif	2639	0
	Positif	157	0

Pada tabel 10. Confusion Matrix tersebut menunjukkan hasil prediksi model Naive Bayes sebelum menggunakan SMOTE, di mana model mengalami masalah dalam memprediksi kelas "Positif". Pada semua perbandingan data 0.1, 0.2, 0.3 dan 0.5, model hanya memprediksi kelas "Negatif" dengan akurasi tinggi, tetapi gagal mengklasifikasikan data sebagai "Positif" (semua prediksi "Positif" bernilai nol). Ini menunjukkan adanya ketidakseimbangan kelas, di mana kelas mayoritas "Negatif" mendominasi dan model overfitting terhadapnya. Akibatnya, model tidak mampu mengenali pola dari kelas minoritas "Positif". Penggunaan SMOTE diperlukan untuk mengatasi masalah ini dengan menyeimbangkan data dan meningkatkan akurasi prediksi terhadap kelas minoritas. Tabel Confusion Matrix tersebut menunjukkan hasil prediksi model Naive Bayes sebelum menggunakan SMOTE, di mana model mengalami masalah dalam memprediksi kelas "Positif". Pada semua perbandingan data 0.1, 0.2, 0.3 dan 0.5, model hanya memprediksi kelas "Negatif" dengan akurasi tinggi, tetapi gagal mengklasifikasikan data sebagai "Positif" (semua prediksi "Positif" bernilai nol). Ini menunjukkan adanya ketidakseimbangan kelas, di mana kelas mayoritas "Negatif" mendominasi dan model overfitting terhadapnya. Akibatnya, model tidak mampu mengenali pola dari kelas minoritas "Positif". Penggunaan SMOTE diperlukan untuk mengatasi masalah ini dengan menyeimbangkan data dan meningkatkan akurasi prediksi terhadap kelas minoritas.

Tabel 11. Hasil Confusion Matrix After SMOTE

Perbandingan	Prediction Class	Actual Class	
		Pred. Negative	Pred. Positive
0.1	Negatif	452	502
	Positif	10	90
0.2	Negatif	908	1016



Gambar 10. Grafik Frekuensi Kata

4. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa analisis sentimen komentar media sosial Twitter terhadap tes CPNS menggunakan algoritma Naïve Bayes dengan 5.599 data bersih, yang terdiri dari 5.269 data negatif dan 323 data positif. Penelitian ini menguji metode Naïve Bayes dengan rasio pembagian data 80:20, 70:30, 90:10, dan 50:50. Hasilnya, rasio 80:20 dan 90:10 mencapai akurasi 94%, sedangkan rasio 70:30 menunjukkan hasil terbaik dengan akurasi 95%. Rasio 50:50 juga menghasilkan akurasi 94%. Dengan demikian, rasio 70:30 terbukti paling optimal dalam meningkatkan akurasi dan kemampuan model dalam mengenali pola sentimen. Hasil eksperimen menunjukkan bahwa model cenderung lebih fokus pada data dengan sentimen negatif, yang mengakibatkan ketidakseimbangan dalam dataset. Untuk mengatasi masalah ini, teknik optimasi SMOTE diterapkan agar jumlah data di kelas minoritas seimbang dengan kelas mayoritas. Penelitian ini menunjukkan bahwa penerapan SMOTE berhasil meningkatkan kinerja model Naïve Bayes dalam mengatasi ketidakseimbangan data analisis sentimen tes CPNS. Sebelum SMOTE, model memiliki akurasi tinggi untuk kelas negatif tetapi tidak mampu mengklasifikasikan kelas positif, dengan precision, recall, dan F1-Score mencapai 0% untuk data positif. Setelah penerapan SMOTE, performa model meningkat, terutama dalam mengidentifikasi kelas positif, dengan precision dan recall mencapai 85-98%. Dari beberapa rasio data, rasio 0.2 menunjukkan keseimbangan optimal antara klasifikasi positif dan negatif. Akurasi untuk rasio 80:20, 70:30, dan 90:10 mencapai 91%, sementara rasio 50:50 mencapai 89%. Meskipun akurasi keseluruhan sedikit menurun, penggunaan SMOTE membuat model lebih efektif dalam mendeteksi kedua jenis sentimen, menghasilkan klasifikasi yang lebih seimbang dan akurat.

REFERENCES

- [1] R. T. Silangen dan M. Y. Matdoan, "Klasifikasi Hasil Seleksi Kompetensi Dasar CPNS Menggunakan Metode Decision Tree," *Inferensi*, vol. 5, no. 2, hal. 69, 2022, doi: 10.12962/j27213862.v5i2.12353.
- [2] Y. H. Yudha, R. F. Marta, dan H. Panggabean, "Diskursus Faking pada Seleksi CPNS pada Video Testimoni Mengerjakan Tes Karakteristik Pribadi," *J. Komun.*, vol. 13, no. 1, hal. 118–132, 2021, [Daring]. Tersedia pada: <https://journal.untar.ac.id/index.php/komunikasi>
- [3] E. Salim dan A. Solichin, "Analisis Sentimen Pada Media Sosial Twitter Terhadap Pelayanan Dinas Kependudukan Dan Pencatatan Sipil Menggunakan Algoritma Naïve Bayes," *IDEALIS Indones. J. Inf. Syst.*, vol. 5, no. 2, hal. 79–86, 2022, doi: 10.36080/idealis.v5i2.2961.
- [4] I. A. Oktariansyah, F. R. Umbara, dan F. Kasyidi, "Klasifikasi Sentimen Untuk Mengetahui Kecenderungan Politik Pengguna X Pada Calon Presiden Indonesia 2024 Menggunakan Metode IndoBert," *Buuld. Informatics, Technol. Sci.*, vol. 6, no. 2, hal. 636–648, 2024, [Daring]. Tersedia pada: <https://ejournal.seminar-id.com/index.php/bits/article/view/5435>
- [5] D. S. Ningsih dan R. R. Suryono, "Comparison of Naïve Bayes and Information Gain Algorithms in Cyberbullying Sentiment Analysis on Twitter Perbandingan Algoritma Naïve Bayes Dan Information Gain," vol. 5, no. 4, hal. 1085–1091, 2024.
- [6] J. M. Br Sembiring dan H. H., "Naïve Bayes Algorithm Classification in Sentiment Analysis Covid-19 Wikipedia," *J. Tek. Inform.*, vol. 3, no. 4, hal. 869–875, 2022, doi: 10.20884/1.jutif.2022.3.4.311.
- [7] R. I. Lutfi, B. S. Rintyarna, dan 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, hal. 127–135, 2023.
- [8] W. Astuti, E. Firasari, F. Lia Dwi Cahyanti, dan F. Sarasati, "Analisis Sentiment on the Acceptance of Cpnas 2021 on Twitter Social Media Using Textblob," *J. Comput. Inf. Technol. As an Accredited. J. Rank*, vol. 19, no. 1, hal. 15–21, 2022, [Daring]. Tersedia pada: <https://doi.org/10.33480/techno.v19i1.2980>



- [9] A. Jaya, “Analisis Sentimen Pandangan Public Profesi PNS (Pegawai Negeri Sipil) dari Twiter menerapkan indonesian Roberta Base Sentiment Classifier,” *Indones. J. Data Sci.*, vol. 4, no. 1, hal. 38–44, 2023, doi: 10.56705/ijodas.v4i1.66.
- [10] Fitriani Fitriani, Ema Utami, dan 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, hal. 23–30, 2022, doi: 10.46764/teknimedia.v3i1.53.
- [11] Harman Akbar Tullah *et al.*, “Sentiment Analysis Of Indonesian Civil Servan Candidates 2023 Twitter Network With Naive Bayes Algorithm Method,” *Inspir. J. Teknol. Inf. dan Komun.*, vol. 13, no. 2, hal. 49–63, 2023, doi: 10.35585/inspir.v13i2.66.
- [12] F. N. Hidayat dan 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, hal. 665–672, 2023, doi: 10.55338/saintek.v5i2.1359.
- [13] R. Indrayana dan M. Solikhin, “Analisis Sentimen Pada Media Sosial Twitter Terhadap Pelayanan Dinas Kependudukan Dan Pencatatan Sipil Menggunakan Algoritma Naïve Bayes,” *Seminar Nasional Pendidikanipa Danmatematika*, vol. 12, no. 1, hal. 79–86, 2020.
- [14] D. F. Sebastian, H. Sulistiani, dan A. R. Isnain, “Sentiment Analysis of Public Opinion on the Right of Inquiry in Indonesia in 2024 Using the Support Vector Machine (Svm) Method Analisis Sentimen Opini Masyarakat Mengenai Hak Angket Di Indonesia Tahun 2024 Menggunakan Metode Support Vector Machine (Sv,” vol. 5, no. 4, hal. 1025–1034, 2024.
- [15] S. Surya Prabu Al Amin, J. Haerul Jaman, dan G. Garno, “Analisis Sentimen Masyarakat Terhadap Penanganan Kasus Penembakan Brigadir J Dengan Algoritma Naïve Bayes,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 4, hal. 2519–2526, 2024, doi: 10.36040/jati.v7i4.7126.
- [16] P. Elisa dan A. Rahman Isnain, “Comparison of Random Forest, Support Vector Machine and Naive Bayes Algorithms To Analyze Sentiment Towards Mental Health Stigma,” *J. Tek. Inform.*, vol. 5, no. 1, hal. 321–329, 2024, [Daring]. Tersedia pada: <https://doi.org/10.52436/1.jutif.2024.5.1.1817>
- [17] R. Sanjaya, E. Tohidi, E. Wahyudi, dan K. Kaslani, “Analisis Sentimen Terhadap Berhentinya Tiktokshop Pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, hal. 507–514, 2024, doi: 10.36040/jati.v8i1.8443.
- [18] F. S. Mufidah, S. Winarno, F. Alzami, E. D. Udayanti, dan R. R. Sani, “Analisis Sentimen Masyarakat Terhadap Layanan Shopeefood Melalui Media Sosial Twitter Dengan Algoritma Naïve Bayes Classifier,” *JOINS (Journal Inf. Syst.*, vol. 7, no. 1, hal. 14–25, 2022, doi: 10.33633/joins.v7i1.5883.
- [19] S. Alam, M. G. Resmi, dan N. Masripah, “Classification of Covid-19 vaccine data screening with Naive Bayes algorithm using Knowledge Discovery in database method,” *J. Comput. Networks, Archit. High Perform. Comput.*, vol. 4, no. 2, hal. 177–185, 2022, doi: 10.47709/cnahpc.v4i2.1584.
- [20] T. Ridwansyah, “Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 5, hal. 178–185, 2022, doi: 10.30865/klik.v2i5.362.
- [21] M. B. Rissan dan R. F. Hassan, “Naïve-Bayes family for sentiment analysis during COVID-19 pandemic and classification tweets,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 28, no. 1, hal. 375–383, 2022, doi: 10.11591/ijeecs.v28.i1.pp375-383.
- [22] Tommy Suhendra, B. Intan, dan A. T. Martadinata, “ANALISIS SENTIMEN PENGGUNA APLIKASI NETFLIX PADA ULASAN GOOGLE PLAYSTORE MENGGUNAKAN METODE NAÏVE BAYES Tommy,” *ESCAF 3rd*, vol. 2, no. 2, hal. 1011–1022, 2024, doi: 10.47065/bits.v6i2.5528.
- [23] F. Sidik, I. Suhada, A. H. Anwar, dan F. N. Hasan, “Analisis Sentimen Terhadap Pembelajaran Daring Dengan Algoritma Naive Bayes Classifier,” *J. Linguist. Komputasional*, vol. 5, no. 1, hal. 34, 2022, doi: 10.26418/jlk.v5i1.79.
- [24] Y. A. Singgalen, “Analisis Sentimen dan Sistem Pendukung Keputusan Menginap di Hotel Menggunakan Metode CRISP-DM dan SAW,” *J. Inf. Syst. Res.*, vol. 4, no. 4, hal. 1343–1353, 2023, doi: 10.47065/josh.v4i4.3917.
- [25] H. Utami, “Analisis Sentimen dari Aplikasi Shopee Indonesia Menggunakan Metode Recurrent Neural Network,” *Indones. J. Appl. Stat.*, vol. 5, no. 1, hal. 31, 2022, doi: 10.13057/ijas.v5i1.56825.
- [26] S. A. Pratomo, S. Al Faraby, dan M. D. Purbolaksono, “Analisis Sentimen Pengaruh Kombinasi Ekstraksi Fitur TF-IDF dan Lexicon Pada Ulasan Film Menggunakan Metode KNN,” *e-Proceeding Eng.*, vol. 8, no. 5, hal. 10116–10126, 2021, [Daring]. Tersedia pada: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15726>
- [27] A. Muzaki dan A. Witanti, “Sentimen Analisis Masyarakat Di Twitter Terhadap Pilkada 2020Ditengah Pandemic Covid-19 Dengan Metode Naïve Bayes Classifier,” *J. Tek. Inform.*, vol. 2, no. 2, hal. 101–107, 2021.
- [28] A. Septiarini, Rizqi Saputra, Andi Tejawati, dan Masna Wati, “Deteksi Sarung Samarinda Menggunakan Metode Naive Bayes Berbasis Pengolahan Citra,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 5, hal. 927–935, 2021, doi: 10.29207/resti.v5i5.3435.
- [29] R. Faiz Ananda, A. Syahri, dan F. N. Hasan, “Sentiment Analysis of Customer Satisfaction in Gojek and Grab Application Reviews Using the Naive Bayes Algorithm,” *J. Tek. Inform. (JUTIF).1*, vol. 5, no. 1, hal. 233–241, 2024, [Daring]. Tersedia pada: <https://doi.org/10.52436/1.jutif.2024.5.1.1680>