

PERSEPSI PUBLIK TERHADAP KEPEMIMPINAN FIRLI BAHURI DI KPK: PENDEKATAN SENTIMEN TWITTER DENGAN NAÏVE BAYES DAN SVM

Jimmy Juliantono*¹⁾, Parjito²⁾

1. Sistem Informasi, Fakultas Teknik Ilmu Komputer, Universitas Teknokrat Indonesia, Indonesia
2. Informatika, Fakultas Teknik Ilmu Komputer, Universitas Teknokrat Indonesia, Indonesia

Article Info

Kata Kunci: Analisis Sentimen; Naïve Bayes; SVM; Firli Bahuri; KPK

Keywords: *Sentiment Analysis; Naïve Bayes; SVM; Firli Bahuri; KPK*

Article history:

Received 11 November 2024

Revised 15 Desember 2024

Accepted 14 Januari 2025

Available online 1 Maret 2025

DOI :

<https://doi.org/10.29100/jupi.v10i2.6181>

* Corresponding author.

Corresponding Author

E-mail address:

jimmy_j@teknokrat.ac.id

ABSTRAK

Komisi Pemberantasan Korupsi (KPK) merupakan lembaga independen Indonesia yang bertujuan melakukan pemberantasan korupsi. Di bawah kepemimpinan Firli Bahuri (2019-2023), KPK menangani kasus besar di sektor infrastruktur dan keuangan, serta meluncurkan program pencegahan korupsi. Namun, kepemimpinannya diwarnai kontroversi, termasuk dugaan pelanggaran etik dan tuduhan korupsi, seperti kasus pemerasan terhadap Menteri Pertanian Syahrul Yasin Limpo. Untuk menganalisis respon masyarakat dengan menggunakan data Twitter, peneliti melakukan perbandingan dengan menggunakan metode pengklasifikasian Naïve Bayes dan Support Vector Machine (SVM). Sebelum dilakukan perbandingan antara kedua model, penelitian ini terlebih dahulu melakukan proses optimasi SMOTE untuk mengatasi ketidakseimbangan data agar data minority dan data majority dapat memiliki jumlah data yang sama. Hasil dari perbandingan kedua model mendapatkan nilai algoritma Naïve bayes dengan akurasi 97% dan algoritma support vector machine dengan akurasi 100%, presisi 100%, *recall* 100%, dan FI-score 100%. Jika dilihat pada perbandingan pada kedua model algoritma, dapat disimpulkan bahwa algoritma SVM mampu melakukan proses analisis sentimen dengan sangat akurat pada setiap tahapannya dibandingkan dengan naïve bayes. Oleh karena itu, dalam penelitian ini terkait opini masyarakat di media sosial twitter terhadap kepemimpinan firli bahuri di kpk, dapat disimpulkan bahwa algoritma SVM menjadi pilihan yang lebih baik dibandingkan algoritma Naïve bayes.

ABSTRACT

The Corruption Eradication Commission (KPK) is an independent Indonesian institution whose aim is to eradicate corruption. Under the leadership of Firli Bahuri (2019-2023), the KPK handled major cases in the infrastructure and financial sectors, and launched a corruption prevention program. However, his leadership was marred by controversy, including allegations of ethical violations and accusations of corruption, such as the blackmail case against the Minister of Agriculture Syahrul Yasin Limpo. To analyze public responses using Twitter data, researchers carried out comparisons using the Naïve Bayes and Support Vector Machine (SVM) classification methods. Before comparing the two models, this research first carried out a SMOTE optimization process to overcome data imbalance so that minority data and majority data can have the same amount of data. The results of the comparison of the two models show that the Naïve Bayes algorithm has an *accuracy* of 97% and the support vector machine algorithm has an *accuracy* of 100%, *precision* of 100%, *recall* of 100%, and FI-score of 100%. If we look at the comparison of the two algorithm models, it can be concluded that the SVM algorithm is able to carry out the sentiment analysis process very accurately at each stage compared to naïve Bayes. Therefore, in this research regarding public opinion on Twitter social media regarding Firli Bahuri's leadership at KPK, it can be concluded that the SVM algorithm is a better choice than the Naïve Bayes algorithm.

I. PENDAHULUAN

KOMISI Pemberantasan Korupsi Merupakan Institus Independen pada Pemerintahan Indonesia yang dibentuk pada tahun 2002. Lembaga ini didirikan atas dasar hukum Undang-undang Nomor 30 Tahun 2002 tentang Komisi Pemberantasan Korupsi. Dalam gambarannya, KPK melibatkan masyarakat umum dan menyampaikan laporan secara terbuka serta kepada Presiden, DPR, dan BPK secara berkala[1]. Kasus-kasus penurunan citra KPK telah menjadi isu umum di Indonesia, yang konon disebabkan oleh nepotisme, kolusi untuk mendapatkan imbalan, dan bentuk-bentuk korupsi lainnya termasuk penyalahgunaan wewenang. Tindak pidana korupsi dapat terjadi di mana saja di masyarakat, tidak hanya di lembaga pemerintah saja, tetapi juga di industri swasta[2]. Salah satu kasus korupsi yang paling mengejutkan adalah kasus tindak pidana korupsi oleh ketua KPK Firli Bahuri. Polda Metro Jaya resmi menetapkan Firli Bahuri sebagai tersangka kasus korupsi pada 22 November 2023. Firli dituduh melakukan suap, pemerasan, dan gratifikasi terkait pengurusan urusan hukum di Kementerian Pertanian dibawah kepemimpinan SYL[3]. Direktur Reserse Kriminal Khusus Polda Metro Jaya Kombes Ade Safri Simanjuntak mengklaim penetapan tersangka Firli Bahuri berdasarkan fakta penyidikan yang mengungkap cukup bukti terkait dugaan tindak pidana korupsi seperti menerima suap dan pemerasan. Ade Safri memberikan penjelasan mengenai prosedur perkara yang berujung pada penetapan tersangka. Dalam kasus ini, Firli Bahuri didakwa melanggar Pasal 65 KUHP dan Pasal 12 e alias Pasal 12B atau Pasal 11 UU Tipikor[4].

Dalam perkembangannya, dugaan korupsi yang dilakukan oleh Firli dikuatkan oleh sejumlah barang bukti yang telah dimiliki oleh penyidik Polda Metro Jaya, yang diantaranya merupakan dokumen penukaran valuta asing dalam pecahan dolar Singapura dan Amerika Serikat dengan total sebesar Rp7,4 miliar. Selain itu, untuk menguatkan bukti, penyidik juga telah meminta keterangan dan konfirmasi terhadap 91 saksi dan ahli. Kasus ini akan diingat sebagai sebuah skandal yang memalukan bagi Indonesia di tataran global, karena pimpinan lembaga anti korupsi yang seharusnya menjadi garda terdepan pemberantasan korupsi, justru terlibat dalam praktik rasuah. Untuk itu, ICW mendorong agar Polda Metro Jaya harus menangani gugatan dugaan pungli secara imparial dan akuntabel. Penyidik bisa memecat Firli setelah menaikkan status tersangka. Hal ini dilakukan mengingat adanya kekhawatiran yang bersangkutan akan mencopot atau menghilangkan dan menghilangkan barang bukti, Serta Presiden selaku atasan langsung harus memastikan kinerja lembaga antirasuah ini akuntabel dan bebas dari kepentingan politik setelah dipastikan terbitnya Keputusan Presiden (Keppres) tentang pemberhentian sementara Firli Bahuri sebagai Pimpinan KPK. dan menunjuk Nawawi Pomolango yang sebelumnya menjabat Wakil menjadi Pj Ketua KPK. Kasus ini harus dijadikan alat pemulihan dan penilaian oleh otoritas publik mengenai cara yang paling umum dalam memilih pesaing otoritas KPK di masa depan. Masukan masyarakat mengenai rekam jejak calon pimpinan lembaga antirasuah ini sudah sepatutnya diperhitungkan dalam proses seleksi[5]

Twitter adalah salah satu media sosial yang sangat populer di kalangan masyarakat untuk berbagi informasi, baik positif maupun negatif, tentang topik-topik yang sedang dibicarakan. Analisis sentimen adalah metode yang digunakan untuk mengumpulkan dan menilai pendapat masyarakat mengenai suatu isu di media sosial berbasis web. Hal ini menjadi acuan penting bagi peneliti untuk menganalisis opini publik mengenai kepemimpinan firli bahuri selama menjabat sebagai ketua KPK[6].

Metode yang digunakan untuk mengklasifikasikan pengujian ini adalah Naïve Bayes dan Support Vector Machine (SVM). Naïve Bayes sangat populer karena kemudahannya dalam penggunaan. Metode ini mampu memenuhi persyaratan dalam berbagai domain yang berbeda. Di sisi lain, keunggulan dari Support Vector Machine adalah kemampuannya dalam mengidentifikasi hyperplane yang memisahkan dua kelas dengan margin yang lebih besar. Namun, Support Vector Machine memiliki kelemahan dalam hal pemilihan fitur yang tepat[7]. Analisis sentimen dapat diterapkan untuk mengelompokkan komentar dan *tweet* mengenai stigma kepemimpinan Firli Bahuri di Twitter ke dalam tiga kategori yaitu negatif, positif, dan netral[8]. Dengan analisis sentimen menggunakan metode Naïve Bayes dan Support Vector Machine (SVM), diharapkan hasil analisis ini dapat memberikan gambaran mengenai persepsi publik terhadap Firli Bahuri berdasarkan data dari Twitter.

Penelitian Sebelumnya oleh Dedi Darwis tentang analisis sentimen pada data twitter Komisi Pemberantasan Korupsi Republik Indonesia dengan metode SVM mengklasifikasikan sentimen masyarakat pengguna Twitter terhadap kinerja KPK ke dalam tiga kelas yaitu positif (8%), negatif (77%), dan netral (15%). Evaluasi hasil pengujian menunjukkan nilai akurasi sebesar 82%, presisi 90%, *recall* 88%, dan *F1-score* 89%. Temuan ini mengindikasikan bahwa mayoritas sentimen masyarakat terhadap kinerja KPK bersifat negatif, dengan persentase sentimen negatif mencapai 77%. Dengan demikian, dapat disimpulkan bahwa persepsi masyarakat terhadap kinerja KPK sangat kurang baik, seperti yang ditunjukkan oleh dominasi sentimen negatif[9].

Penelitian yang dilakukan oleh Bergas Pamungkas dengan topik analisis sentimen pada kasus benih lobster 2020 dengan menggunakan metode SVM. Hasil penelitian menunjukkan bahwa pengkategorian emosi *tweet*

menggunakan SVM menghasilkan 28% emosi negatif dan 72% emosi positif. Mayoritas tanggapan di Twitter mendukung program KPK untuk mencegah korupsi, serta mendukung kebijakan baru dalam perikanan laut, termasuk larangan ekspor benih lobster. Model SVM yang digunakan menunjukkan performa yang baik dengan akurasi 84,21%, sensitivitas 73,38%, dan spesifisitas 82,10%. Analisis ini menegaskan bahwa media sosial seperti Twitter dapat menjadi sumber data alternatif yang efektif untuk mengidentifikasi dan menyelesaikan masalah terkini menggunakan teknik SVM. Studi ini juga menyoroti potensi penggunaan teknik pembelajaran terawasi lainnya, seperti jaringan saraf, dalam analisis data dari internet untuk menghasilkan wawasan baru[10].

Penelitian lainnya yang dilakukan oleh Ayu Sri Rahayu dengan topik Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Spotify. Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes dengan nilai akurasi sebesar 86.4% sedangkan pada algoritma Support Vector Machine (SVM) memiliki nilai akurasi sebesar 84%. Dapat disimpulkan bahwa Naïve Bayes memberikan unjuk kerja lebih baik daripada Support Vector Machine (SVM) untuk klasifikasi sentimen review Spotify dengan menghasilkan nilai akurasi yang tinggi[11]

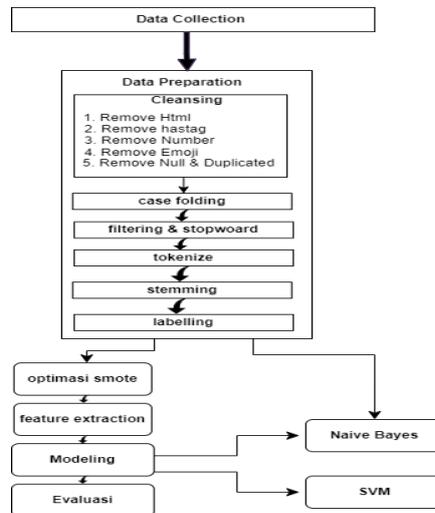
Selanjutnya penelitian yang dilakukan oleh Ni Wayan Ayu Sekar Aprilia dengan topik Analisis Sentimen Terhadap Media Sosial Twitter dengan Kasus Kampanye Anti-Korupsi di Indonesia Menggunakan Naïve Bayes. Hasil penelitian menunjukkan bahwa sebagian besar *tweet* terkait Kampanye Anti-Korupsi memiliki sentimen positif. Analisis sentimen Twitter ini memberikan wawasan tentang persepsi masyarakat terhadap Kampanye Anti-Korupsi di Indonesia, yang dapat membantu dalam merumuskan kebijakan dan strategi anti-korupsi di masa depan. Dari total *tweet* yang dianalisis, 58,64% memiliki sentimen positif, sementara 41,36% memiliki sentimen negatif. Evaluasi data uji menunjukkan tingkat akurasi 74%, presisi 79%, dan *recall* 76%[2].

Berdasarkan pada penelitian sebelumnya terdapat beberapa keterbatasan penelitian yang ditemukan, diantaranya kurangnya fokus pada analisis sentimen terhadap kepemimpinan individu di institusi anti-korupsi, seperti KPK. Penelitian sebelumnya lebih banyak mengkaji sentimen terhadap kebijakan tertentu atau kinerja institusi secara umum, tanpa menyoroti dampak kepemimpinan individu. Penelitian ini mengisi kekurangan tersebut dengan menawarkan perspektif baru yang menilai persepsi publik secara spesifik terhadap kepemimpinan Firli Bahuri di KPK. sehingga dapat membantu dalam memahami dinamika opini publik dan memberikan panduan untuk strategi komunikasi dan perbaikan kinerja di lembaga anti-korupsi.

Penelitian ini bertujuan untuk membandingkan kinerja algoritma Naïve Bayes dan SVM dalam menganalisis sentimen publik terhadap kepemimpinan Firli Bahuri. Melalui analisis data *tweet* di Twitter, diharapkan penelitian ini dapat mengungkap persepsi masyarakat terkait kepemimpinan Firli Bahuri di KPK, memberikan wawasan yang lebih mendalam dan akurat tentang opini publik. Dengan memahami opini publik, institusi dapat merumuskan strategi komunikasi yang lebih efektif, meningkatkan transparansi, serta memperbaiki praktik dan kebijakan yang ada. Penelitian ini juga dapat menjadi panduan dalam proses seleksi calon pemimpin di masa depan, memastikan bahwa mereka memiliki rekam jejak dan kinerja yang dapat diterima oleh masyarakat luas, sehingga meningkatkan kepercayaan publik terhadap lembaga anti-korupsi. Dengan menggunakan tahapan preprocessing data dan evaluasi menggunakan akurasi, presisi, *recall*, dan F1-Score, penelitian ini diharapkan dapat menghasilkan pemahaman yang lebih baik tentang cara kedua algoritma menginterpretasikan serta mengelompokkan opini publik.

II. METODOLOGI PENELITIAN

Penelitian ini dilaksanakan melalui beberapa tahapan, sehingga memiliki alur kerja yang terstruktur dari awal hingga akhir. Tahapan-tahapan penelitian ini dapat dilihat pada Gambar 1 berikut:



Gambar 1. Tahapan Penelitian

A. Data Collection

Penelitian ini melakukan tahapan awal dengan mengumpulkan data *tweet* dari twitter dengan memanfaatkan bantuan Application Programming Interface (API). dalam proses crawling data di Twitter, API (Application Programming Interface) memungkinkan pengguna mengakses langsung dan real-time ke beberapa *tweet*, profil pengguna, serta informasi lainnya. API memfasilitasi pencarian dan filtrasi data berdasarkan kriteria tertentu, otomatisasi pengambilan data dalam skala besar, dan integrasi dengan alat analitik lainnya. Selain itu, API memberikan kontrol dan keamanan melalui mekanisme otentikasi dan otorisasi ke pengguna. Pada Proses Crawling Data di twitter Pengguna membutuhkan autho token dari media sosial twitter yang nantinya akan dilakukan pengolahan library Harvest pada bahasa pemrograman Python. Untuk mendapatkan data *tweet* dengan skala besar kata kunci yang digunakan untuk menemukan *tweet* relevan dalam penelitian ini adalah "Firli Bahuri KPK". Peneliti mendapatkan data *tweet* sebanyak 7000 data dengan proses Crawling pada periode waktu yang berbeda. Data *tweet* yang didapat memiliki rentang waktu sejak tanggal 8 Desember 2023 – 4 Mei 2024

B. Data Understanding

Data understanding adalah tahap dalam analisis sentimen di mana peneliti harus memahami terlebih dahulu data yang ada. Melalui pemahaman data tersebut, peneliti dapat mengatasi masalah sesuai dengan tujuan yang telah ditetapkan. Pada tahap ini, peneliti melakukan analisis mendalam terhadap data yang telah dikumpulkan. Dengan pemahaman yang lebih baik mengenai data tersebut, peneliti dapat merencanakan langkah-langkah yang diperlukan untuk memecahkan masalah dalam penelitian[12].

C. Preprocessing Data

Tahap preprocessing adalah langkah awal dalam text mining pada. Langkah ini mencakup semua rutinitas dan proses untuk mempersiapkan data yang akan digunakan dalam operasi knowledge discovery dalam sistem text mining. Proses yang dilakukan pada tahap ini yaitu *case folding*, *Tokenizing*, *stemming*, *labeling*, dan *filtering*[13]. Berikut penjelasan dari beberapa tahapan yang dilakukan di dalam preprocessing:

- 1) Pada tahap *case folding*, teks diubah dari huruf besar menjadi huruf kecil dan semua tanda baca dihapus.
- 2) Pada tahap *Tokenizing*, setiap kata dipisahkan berdasarkan spasi yang ada.
- 3) Pada tahap *stemming*, kata-kata berimbuhan diubah menjadi bentuk dasarnya
- 4) Pada tahap *filtering*, kata-kata yang tidak penting dihapus dari hasil tokenisasi.

D. Cleaning dan Case folding

Untuk memastikan bahwa *Wordcloud* dan analisis sentimen dapat mencerminkan hasil yang akurat, perlu dilakukan pembersihan data terlebih dahulu. Proses ini meliputi penghapusan URL, mention, hashtag, *stopwords*, emoticon, tanda baca, dan angka, serta mengubah semua huruf menjadi kecil. Pada penelitian ini, pembersihan *stopwords* dilakukan menggunakan library Python *pySastrawi*[14].

TABEL 1.
 CLEANING DATA DAN CASE FOLDING

<i>Tweet_text</i>	<i>Cleansing</i>	<i>Case folding</i>
KPK periode 2019-2024 paling ribet masalah intern. Ketua KPK Firli Bahuri jadi tersangka. Antara Dewas dan Komisioner KPK tidak berjalan efektif. Apalagi saat Pilpres yg lalu...?? ICW Sebut Nurul Ghufron Pengecut karena Mangkir Sidang Etik DewasKPK https://t.co/h9huQJCWCu	KPK periode paling ribet masalah intern KetuaKPK Firli Bahuri jadi tersangka Antara Dewas dan KPK tidak berjalan efektif Apalagi saat Pilpresyg lalu ICW Sebut Nurul Ghufron Pengecut karena Mangkir Sidang Etik Dewas KPK	kpk periode paling ribet masalah intern ketua kpk firli bahuri jadi tersangka antara dewas dan kpk tidak berjalan efektif apalagi saat pilpres yg lalu icw sebut nurul ghufron pengecut karena mangkir sidang etik dewas kpk

E. Tokenizing dan Removal Stopword atau Filtering

Setelah proses *Cleaning* dan *Case folding*, langkah selanjutnya adalah *Tokenizing* dan *Filtering*. *Tokenizing* sendiri merupakan langkah dimana teks atau data teks diubah menjadi sejumlah token. Token adalah unit terkecil dalam pemrosesan teks yang memiliki arti[15]. Sedangkan *Filtering* digunakan untuk menyaring kata-kata yang tidak diperlukan sehingga data menjadi lebih berkualitas[16]. Pada tahapan ini, digunakan perintah *split()* dalam Python untuk memisahkan setiap kata menggunakan tanda koma. *Stopword* atau *filtering* sendiri dalam proses data mining memiliki tujuan untuk menghapus kata-kata yang tidak penting yang terdapat dalam daftar *stopword*, sehingga kata-kata tersebut tidak akan muncul dalam analisis teks. Biasanya, kata-kata yang dihapus adalah kata-kata penghubung atau konjungsi seperti "dan," "atau," "tetapi," "sehingga," dan sebagainya[12]. Hasil dari tahapan ini dapat dilihat pada tabel 2 berikut ini :

TABEL 2.
 TOKENIZING DAN FILTERING

<i>Tokenizing</i>	<i>Filtering</i>
['kpk', 'periode', 'paling', 'ribet', 'masalah', 'intern', 'ketua', 'kpk', 'firli', 'bahuri', 'jadi', 'tersangka', 'antara', 'dewas', 'dan', 'kpk', 'tidak', 'berjalan', 'efektif', 'apalagi', 'saat', 'pilpres', 'yg', 'lalu', 'icw', 'sebut', 'nurul', 'ghufron', 'pengecut', 'karena', 'mangkir', 'sidang', 'etik', 'dewas', 'kpk']	['kpk', 'periode', 'paling', 'ribet', 'masalah', 'intern', 'ketua', 'kpk', 'firli', 'bahuri', 'tersangka', 'dewas', 'kpk', 'berjalan', 'efektif', 'pilpres', 'yg', 'lalu', 'icw', 'sebut', 'nurul', 'ghufron', 'pengecut', 'mangkir', 'sidang', 'etik', 'dewas', 'kpk']

F. Stemming dan Labeling

Stemming merupakan proses mengubah kata-kata dalam dokumen menjadi kata dasar atau akar (*root word*). Proses *stemming* pada dokumen Bahasa Indonesia cukup kompleks karena harus menghilangkan seluruh imbuhan pada kata-kata dalam *tweets*. Untuk ini, digunakan library *Sastrawi Stemming* yang berbasis algoritme dalam Bahasa Indonesia[17]. sementara pada proses *labeling* pelabelan dilakukan dengan Mengkategorikan komentar ke dua kategori, yaitu positif dan negatif. Metode pelabelan ini bekerja dengan memahami makna keseluruhan kalimat, bukan per kata. Hasil dari tahapan ini dapat dilihat pada tabel 3 berikut ini:

TABEL 3.
 STEMMING DAN LABELING

<i>Stemming</i>	<i>Labeling</i>
kpk periode paling ribet masalah intern ketua kpk firli bahuri sangka dewasa kpk jalan efektif pilpres yg lalu icw sebut nurul ghufron kecut mangkir sidang etik dewasa kpk	Positif
kasus mantan ketua kpk firli bahuri bagaimana nya	Positif

Berdasarkan tabel 3, peneliti menggunakan library Sastrawi pada tahap *stemming* untuk mengolah kata-kata yang memiliki imbuhan, sehingga dapat diubah menjadi kata dasar. Hasil *stemming* ini adalah kalimat yang terdiri dari kata-kata tanpa imbuhan. Library Sastrawi menangani kata-kata berimbuhan dengan memecah dan menghilangkan prefiks (awalan), infiks (sisipan), sufiks (akhiran), dan konfiks (kombinasi awalan dan akhiran) sesuai aturan morfologis Bahasa Indonesia. Namun, metode ini memiliki sedikit kendala, seperti kurang akurat dalam menangani kata-kata yang memiliki lebih dari satu bentuk dasar atau yang berasal dari kata serapan, serta mungkin kesulitan mengenali kata-kata baru atau bahasa informal dan non-standar yang digunakan oleh kelompok-kelompok tertentu untuk berkomunikasi dalam konteks sosial yang akrab, seperti "nongkrong" (berkumpul), "nyokap" (ibu), atau "bokap" (ayah). Dimana kata-kata tersebut tidak terdapat dalam kamus dasar.

Pada tahap pelabelan, proses ini dilakukan oleh mesin menggunakan package NLTK. Peneliti menggunakan library Valence Aware Dictionary and sentiment Reasoner (VADER) lexicon untuk menentukan skor polaritas setiap kata, menghasilkan 5234 data dengan sentimen positif dan 175 data dengan sentimen negatif[12]. VADER lexicon digunakan dalam pelabelan sentimen dengan menilai intensitas emosional setiap kata dalam teks berdasarkan kamus kata-kata yang telah diberi skor sentimen. Skor ini dapat berupa nilai positif, negatif, atau netral. Setelah setiap kata diberi skor, skor-skor ini diakumulasikan untuk menentukan skor polaritas keseluruhan teks. VADER juga memperhitungkan faktor-faktor yang mempengaruhi intensitas sentimen, seperti tanda seru, huruf kapital, dan adverbial yang memperkuat atau melemahkan sentimen. Hasilnya adalah skor polaritas yang mencerminkan seberapa positif, negatif, atau netral teks tersebut, yang kemudian digunakan untuk mengklasifikasikan teks sebagai positif, negatif, atau netral. Pada penelitian ini, proses *stemming* dan *labeling* adalah tahapan terakhir pada preprocessing data.

G. Feature Extraction (TF-IDF)

TF-IDF adalah metode yang digunakan untuk menghitung bobot kata dalam konteks information retrieval. Metode ini terkenal pada efisiensinya, kemudahannya, dan hasil yang akurat[18]. Teknik ini membantu mengidentifikasi kata-kata yang paling informatif untuk membedakan antara sentimen positif dan negatif dalam teks. Dalam konteks penelitian ini, TF-IDF dipilih karena kemampuannya untuk menyaring kata-kata yang relevan, mengurangi kebisingan, meningkatkan akurasi model analisis sentimen, serta efisiensinya dalam menangani volume data besar dari Twitter. Dengan menggunakan TF-IDF, penelitian ini dapat memberikan wawasan yang lebih akurat tentang sentimen masyarakat terhadap kepemimpinan firli bahuri di KPK, yang penting untuk pengambilan keputusan yang lebih tepat dalam meningkatkan layanan.

H. Term Frequency (TF)

Term Frequency berfungsi membantu mengidentifikasi kata-kata yang paling sering muncul dalam teks yang dianalisis. Fungsi utama TF adalah untuk menghitung frekuensi kemunculan setiap kata dalam dokumen sehingga

kita dapat mengetahui kata-kata mana yang paling relevan atau menonjol dalam konteks tertentu. TF (Term Frequency) memiliki perhitungan yang bisa dilihat pada persamaan 1 seperti berikut:

$$tf_{t,d} = \frac{n_{t,d}}{\text{(Total number of term in document)}} \quad (1)$$

Fungsi $tf_{t,d}$ dalam konteks *Term Frequency (TF)* adalah untuk mengukur frekuensi kemunculan sebuah kata t dalam dokumen d . Nilai $tf_{t,d}$ adalah rasio jumlah kemunculan kata t terhadap total jumlah kata dalam dokumen d . Perhitungan ini memberikan informasi tentang seberapa sering sebuah kata muncul dalam sebuah dokumen, yang kemudian digunakan dalam proses analisis teks, seperti analisis sentimen, untuk menentukan relevansi dan pentingnya kata tersebut dalam dokumen.

I. Inverse Document Frequency (IDF)

Inverse Document Frequency (IDF) digunakan untuk mengukur seberapa penting sebuah kata dalam sebuah dokumen relatif terhadap seluruh kumpulan dokumen. Secara khusus, IDF menekan kata-kata umum yang muncul di banyak dokumen dan memberikan bobot lebih pada kata-kata yang jarang muncul, sehingga membantu dalam pengidentifikasian kata-kata yang lebih relevan dan signifikan dalam analisis teks. Berikut adalah rumus IDF yang dapat ditemukan pada persamaan 2 sebagai berikut:

$$idf_d = \log \frac{\text{Number of document}}{\text{(Total Number of tern in document)}} \quad (2)$$

Pada perhitungan IDF, logaritma digunakan untuk menghitung frekuensi kemunculan kata dalam seluruh dokumen atau korpus yang tersedia. Total dokumen tersebut dibagi dengan jumlah term atau kata dalam suatu dokumen, kemudian hasilnya dikalikan dengan nilai logaritma.

$$tfidf_d = tf_{t,d} \times idf_d \quad (3)$$

Dengan mengalikan nilai TF dengan IDF, TF-IDF memberikan bobot tinggi pada kata-kata yang sering muncul dalam dokumen tertentu tetapi jarang muncul di seluruh kumpulan dokumen. Pada persamaan tersebut, pembobotan TF-IDF dihasilkan dari perkalian $tf_{t,d}$ dengan idf_d , yang merepresentasikan bobot setiap kata. Pembobotan ini menghitung bobot kata secara proporsional, sehingga hasil TF-IDF dapat digunakan untuk tahap visualisasi dan pemodelan.

J. Modelling Menggunakan Klasifikasi Naïve Bayes dan Support Vector Machine

Setelah data diberi label dan pembobotan kata menggunakan TF-IDF selesai, selanjutnya masuk dalam proses metode Algoritma Naïve Bayes dan Support Vector Machine (SVM). Pada kedua metode tersebut akan dilakukan pengujian untuk menentukan performa terbaik dari setiap model. Pada penelitian ini Algoritma Naïve Bayes dipilih karena kesederhanaan, kecepatan komputasi, dan kemampuannya dalam menangani dataset besar dan data teks, sementara algoritma Support Vector Machine (SVM) dipilih karena kemampuannya dalam menangani data berdimensi tinggi dan kemampuannya menemukan hyperplane optimal yang memisahkan kelas. Kriteria yang digunakan untuk menentukan performa terbaik antara kedua algoritma ini meliputi akurasi, presisi, *recall*, F1-score, serta waktu komputasi dan kemampuan generalisasi model terhadap data baru.

K. Naïve Bayes Classifier

Naïve Bayes adalah salah satu algoritma yang digunakan dalam text mining. Algoritma ini memprediksi kejadian di masa depan berdasarkan data dari pengalaman sebelumnya, sehingga dikenal dengan nama Teorema Bayes [19]. Berikut adalah rumus IDF yang dapat ditemukan pada persamaan 4 sebagai berikut:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (4)$$

Pada persamaan 4 dijelaskan bahwa Rumus Naive Bayes diatas digunakan untuk menghitung probabilitas posterior yaitu probabilitas suatu kelas c terjadi diberikan input x . Untuk ini, pertama-tama dihitung probabilitas likelihood $P(x|c)$, yang merupakan probabilitas munculnya input x jika diketahui kelasnya c . Kemudian, dikalikan dengan probabilitas prior $P(c)$, yang adalah probabilitas kelas c dalam keseluruhan data. Hasil perkalian ini dibagi dengan probabilitas prior input $P(x)$, yaitu probabilitas bahwa input x muncul dalam data secara keseluruhan, untuk menormalkan hasil dan menghasilkan probabilitas posterior yang akurat.

L. Support Vector Machine

Support Vector Machine adalah sistem pembelajaran yang memanfaatkan fungsi-fungsi linear dalam ruang fitur berdimensi tinggi dan dilatih menggunakan algoritma pembelajaran yang berdasarkan teori optimasi. Diperkenalkan pertama kali oleh Vapnik pada tahun 1992, Support Vector Machine menggabungkan berbagai konsep unggulan dalam bidang pengenalan pola secara harmonis[20]. Metode SVM memungkinkan penyelesaian masalah linear dengan menerapkan transformasi matematis pada ruang belajar menggunakan fungsi kernel. Konsep sentral dalam metode SVM adalah mengklasifikasikan data dengan menentukan hyperplane terbaik yang memisahkan dua kelas yang telah ditentukan. SVM bekerja dengan menemukan hyperplane optimal yang memiliki margin maksimum, yaitu jarak antara titik data terdekat dengan hyperplane. Titik-titik data terdekat dengan hyperplane ini disebut sebagai vektor pendukung[21]. Berikut adalah rumus SVM yang dapat dilihat pada persamaan 5 sebagai berikut:

$$f(x) = \sum a_i y_i K(x, x') + b \quad \forall i = 1 \quad (5)$$

Persamaan $f(x) = \sum a_i y_i K(x, x') + b \quad \forall i = 1$ adalah fungsi keputusan SVM untuk klasifikasi biner. Dalam persamaan ini, Dimana a_i merupakan bobot untuk setiap vektor support x_i , y_i Pada rumus diatas, $f(x)$ adalah hasil prediksi, a_i adalah koefisien Lagrange yang diperoleh dari pelatihan, y_i adalah label kelas dari $K(x, x')$ adalah fungsi kernel yang memetakan data ke dimensi yang lebih tinggi agar bisa dipisahkan secara linear, dan b adalah bias. Penjumlahan \sum mencakup semua vektor support. Fungsi ini digunakan untuk memprediksi kelas dari data baru x dengan mengevaluasi $f(x)$; jika $f(x) > 0$, maka x diklasifikasikan ke dalam kelas positif, jika tidak maka ke kelas negatif. Hanya vektor support yang mempengaruhi penjumlahan berbobot ini, dan pemilihan fungsi kernel yang tepat sangat penting untuk memisahkan data yang tidak linear menjadi linear di ruang fitur yang lebih tinggi[22].

M. Optimasi Smote

Menurut Siringoringo, SMOTE merupakan metode yang digunakan untuk mengatasi masalah data yang tidak seimbang, di mana satu kelompok data memiliki jumlah yang jauh lebih besar daripada kelompok lainnya. Ketidakeimbangan ini bisa menyebabkan model yang dikembangkan menjadi terlalu fokus pada kelompok mayoritas, sehingga performa prediksinya buruk untuk kelompok minoritas. SMOTE mengatasi masalah ini dengan menghasilkan data tambahan untuk kelompok minoritas, sehingga jumlah data di kedua kelompok menjadi seimbang. Dengan demikian, model yang dikembangkan dapat belajar dengan lebih adil dan menghindari overfitting[23].

N. Evaluasi

Penelitian ini menggunakan *Confusion matrix* sebagai metode evaluasi untuk menilai hasil dari algoritma yang digunakan. Metode ini sangat membantu dalam menganalisis kualitas classifier. Evaluasi akan mencakup perhitungan nilai akurasi, *recall*, *precision*, dan *f1-score*, hasil dari tahapan ini nantinya akan disajikan dalam bentuk persentase[24]. Berikut tabel *confusion matrix* yang dapat dilihat pada tabel 4 sebagai berikut:

TABEL IV.
CONFUSION MATRIX

Nilai Aktual	Nilai Prediksi Positif	Nilai Prediksi Negatif
Positif	True Positive (TP)	False Negative (FN)
Negatif	False Positive (FP)	True Negative (TN)

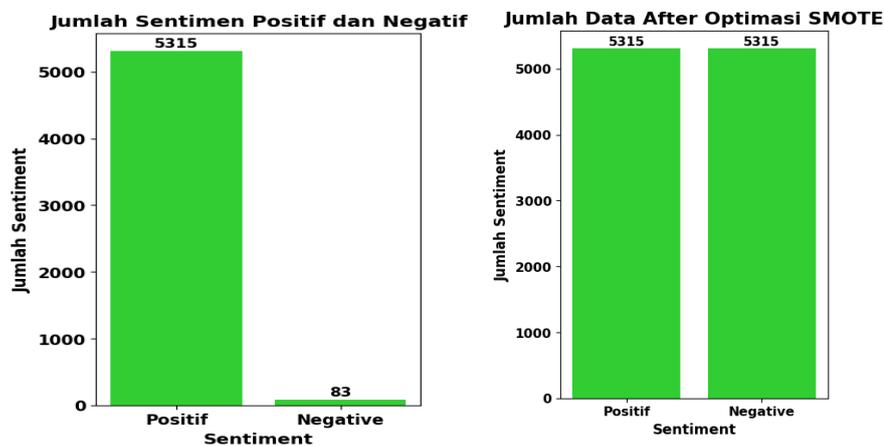
Pada Tabel 4, *True Positive (TP)* adalah jumlah data positif yang berhasil diprediksi dengan benar sebagai positif. *False Negative (FN)* adalah jumlah data positif yang salah diprediksi sebagai negatif. *False Positive (FP)* adalah

jumlah data negatif yang salah diprediksi sebagai positif, dan *True Negative (TN)* adalah jumlah data negatif yang berhasil diprediksi dengan benar sebagai negatif. Dengan menganalisis *confusion matrix*, kita dapat menghitung berbagai metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score* untuk menilai kinerja model klasifikasi secara lebih mendalam[22].

III. HASIL DAN PEMBAHASAN

A. Perbandingan Akurasi

Dalam penelitian ini, setelah melalui tahapan preprocessing, data yang dihasilkan terdiri dari 5315 data dengan sentimen positif dan 83 data dengan sentimen negatif. Ketidakseimbangan antara data sentimen positif dan negatif ini menyebabkan model algoritma Naïve Bayes dan Support Vector Machine lebih cenderung berlatih pada kata-kata positif, sehingga akurasi dan performa klasifikasi untuk kelas positif menjadi tinggi. Oleh karena itu, hasil eksperimen tidak hanya berfokus pada nilai akurasi untuk mengevaluasi suatu model algoritma. Untuk menentukan model algoritma Naïve Bayes atau dan Support Vector Machine yang memiliki performa terbaik, diperlukan evaluasi melalui classification report yang mencakup nilai *precision*, *recall*, dan *F1-Score*. Untuk mengatasi ketidakseimbangan data ini, penelitian ini menerapkan optimasi SMOTE (Synthetic Minority Over-sampling Technique). Dengan menggunakan SMOTE, data minoritas akan memiliki jumlah yang sama dengan data mayoritas, sehingga model algoritma tidak hanya berlatih pada satu jenis sentimen saja.



Gambar 2. Optimasi Smote

Pada Gambar 2 ditunjukkan jumlah sentimen positif dan negatif sebelum dan sesudah proses optimasi SMOTE. Setelah dilakukan optimasi SMOTE, jumlah data sentimen positif dan negatif menjadi seimbang. Dengan optimasi menggunakan SMOTE, model algoritma dapat mempelajari setiap sentimen dengan lebih seimbang karena jumlah data antara sentimen positif dan negatif menjadi setara. SMOTE bekerja dengan meningkatkan performa model dan memperbaiki distribusi data yang tidak seimbang antara kelas mayoritas dan minoritas. Ketika dataset tidak seimbang, model cenderung lebih memfokuskan pelatihan pada kelas mayoritas, mengabaikan fitur penting dari kelas minoritas yang mengakibatkan bias dan akurasi yang rendah pada prediksi kelas minoritas. SMOTE mengatasi masalah ini dengan menghasilkan contoh sintesis baru dari kelas minoritas melalui interpolasi antara contoh-contoh yang ada, sehingga jumlah data antara kelas mayoritas dan minoritas menjadi seimbang. Dengan distribusi data yang lebih seimbang, model dapat mempelajari karakteristik dari kedua kelas dengan lebih efektif, mengurangi bias terhadap kelas mayoritas, dan meningkatkan kemampuan generalisasi model. Hasilnya, model yang dioptimasi dengan SMOTE menunjukkan peningkatan akurasi secara keseluruhan karena mampu membuat prediksi yang lebih akurat dan adil terhadap semua kelas dalam dataset. Setelah proses optimasi, langkah selanjutnya adalah melakukan proses pelatihan (*training*) dan pengujian (*testing*) pada model algoritma. Hal ini penting untuk memastikan bahwa model dapat belajar dan menggeneralisasi dengan baik terhadap data baru yang telah dioptimasi dengan SMOTE, sehingga meningkatkan performa dan akurasi dalam klasifikasi sentimen secara keseluruhan.

B. Tahap Pengujian

Dalam penelitian ini, untuk menentukan performa terbaik dari setiap model algoritma peneliti menggunakan 70% data untuk pelatihan dan 30% untuk pengujian pada model Naïve Bayes dan Support Vector Machine. Data

yang dipakai untuk menerapkan model algoritma ini telah mengalami proses optimasi melalui teknik SMOTE sehingga data tersebut lebih seimbang dan dapat meningkatkan akurasi model.

TABEL IV.
 MODELLING NAÏVE BAYES DAN SUPPORT VECTOR MACHINE

Modeling		accuracy	precision	recall	F1-score
Naïve Bayes	Negatif	97%	96%	99%	97%
	Positif		99%	96%	97%
Support Vector Machine	Negatif	100%	100%	100%	100%
	Positif		100%	100%	100%

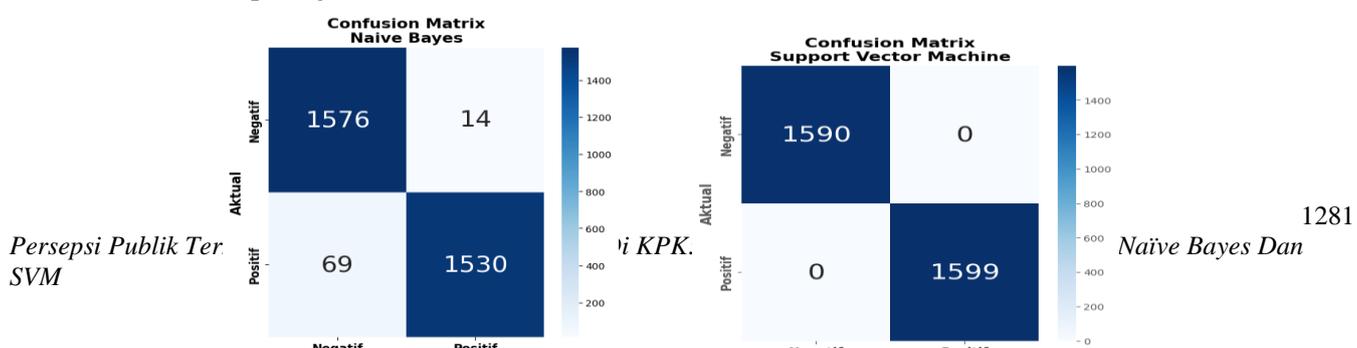
Pada tabel 4 Berdasarkan hasil eksperimen menggunakan 70% data untuk pelatihan dan 30% data untuk pengujian setelah dilakukan optimasi SMOTE, dapat disimpulkan bahwa model Support Vector Machine (SVM) menunjukkan performa yang sempurna dengan akurasi, presisi, *recall*, dan *F1-score* semuanya mencapai 100% untuk kedua kelas (negatif dan positif). Di sisi lain, model Naive Bayes juga menunjukkan performa yang sangat baik dengan akurasi 97% dan keseimbangan yang sangat baik antara presisi dan *recall* untuk kedua kelas. SVM menunjukkan keunggulan dalam konsistensi dan keseimbangan performa, menjadikannya pilihan yang lebih baik untuk mengklasifikasikan sentimen negatif dan positif dalam penelitian ini. Tetapi sebelum kita benar-benar menentukan model mana yg terbaik, kita harus melihat untuk aspek lain seperti presisi, *recall*, dan *F1-score*. Pada algoritma Naïve Bayes, akurasi 97% dan sedangkan untuk Support Vector Machine (SVM) 100%. Untuk kasus sentimen negatif, presisi Naïve Bayes berada di angka 96% dan SVM di 100%, *recall* Naïve Bayes berada di 99% dan SVM di 100%, serta *F1-score* dari Naïve Bayes 97% dan SVM 100%. Untuk sentimen positif, presisi Naïve Bayes adalah 99% dan SVM 100%, *recall* Naïve Bayes adalah 96% dan SVM 100%, serta *F1-score* dari Naïve Bayes 97% dan SVM 100%.

Pada hasil eksperimen, SVM menunjukkan kinerja yang lebih baik dalam analisis sentimen dibandingkan Naïve Bayes pada dataset dengan interaksi fitur yang kompleks atau non-linear. SVM dapat memisahkan fitur sentimen positif dari fitur netral atau negatif dengan efektif menggunakan kernel non-linear, seperti pada kalimat "Saya sangat menghargai kepemimpinan Firli Bahuri di KPK, meskipun ada beberapa tantangan yang harus dihadapi." Sebaliknya, Naïve Bayes, yang mengasumsikan independensi antara fitur, mungkin gagal menangkap interaksi kompleks, seperti dalam kalimat "saya rasa kinerja Firli Bahuri tidak buruk sama sekali," yang bisa menyebabkan kesalahan dalam klasifikasi sentimen.

Jika dibandingkan dengan metode lain seperti Random Forest dan Gradient Boosting, terdapat perbedaan penting. Random Forest, sebagai metode berbasis ensemble, efektif dalam menangani data tidak seimbang dan mengurangi risiko overfitting dengan menggabungkan prediksi dari banyak pohon keputusan, namun mungkin tidak mencapai performa yang sama dalam konsistensi dan presisi pada data dengan fitur non-linear yang sangat kompleks. Sementara itu, Gradient Boosting membangun model secara iteratif dengan memperbaiki kesalahan model sebelumnya, memberikan performa tinggi pada dataset besar dan kompleks, tetapi memerlukan tuning parameter yang lebih intensif dibandingkan SVM.

Secara keseluruhan Random Forest dan Gradient Boosting dapat memberikan hasil yang sangat baik dalam analisis sentimen, performa sempurna SVM dalam studi ini menunjukkan keunggulan dalam hal konsistensi dan akurasi pada dataset spesifik, menunjukkan bahwa SVM mungkin lebih sesuai untuk aplikasi ini dibandingkan dengan metode klasifikasi lain.

Untuk menentukan setiap model algoritma terbaik, peneliti melakukan tahapan eksperimen perbandingan pada kedua algoritma yang digunakan dengan *confusion matrix*. Hasil perbandingan untuk kedua model algoritma dapat dilihat secara visual pada gambar 3 dibawah.



Gambar 3. *Confusion matrix*

Setelah melalui proses optimasi, setiap algoritma dilakukan proses *Confusion matrix* untuk mengevaluasi kinerja model klasifikasi dengan membandingkan prediksi model terhadap nilai sebenarnya. Berdasarkan gambar diatas, algoritma Support Vector Machine mengklasifikasikan 1590 data sebagai True Negatif, dan 1599 True Positif, Sedangkan algoritma Naïve Bayes mengklasifikasikan data 1576 True Negatif dan 1530 True Positif. Jika dilihat dalam perbandingan kedua model, algoritma SVM menunjukkan akurasi yang lebih tinggi dibandingkan Naïve Bayes, dengan akurasi mencapai 100%. Selanjutnya kedua algoritma akan dilakukan perbandingan berdasarkan akurasi (*precision*) dan kelengkapan (*recall*) pada klasifikasi sentimen, baik untuk kelas negatif maupun positif. Untuk mengevaluasi algoritma terbaik, pada tahapan ini *F1-score* digunakan sebagai indikator kinerja. *F1-score* dihitung berdasarkan *precision* dan *recall*.

Pada Perbandingan perbandingan berdasarkan akurasi (*precision*) dan kelengkapan (*recall*), SVM cenderung memiliki kinerja yang lebih baik dalam analisis sentimen dibandingkan Naïve Bayes pada dataset yang memiliki interaksi fitur yang kompleks atau tidak linear. Hal ini karena SVM mampu menangani margin pemisahan yang optimal dan menggunakan kernel non-linear untuk menangkap pola-pola sentimen yang sulit dipisahkan secara linear, seperti frasa dengan interaksi kompleks antara kata-kata atau negasi. Misalnya, dalam kalimat "Saya sangat menghargai kepemimpinan Firli Bahuri di KPK, meskipun ada beberapa masalah yang harus dihadapi," SVM dapat memisahkan fitur sentimen positif dari fitur netral atau negatif dengan lebih efektif menggunakan kernel non-linear.

Sebaliknya, Naïve Bayes, yang mengasumsikan independensi antara fitur, mungkin gagal menangkap interaksi kompleks ini, seperti pada kalimat "saya rasa kinerja Firli Bahuri tidak buruk sama sekali", di mana asumsi independensi dapat menyebabkan kesalahan dalam klasifikasi sentimen keseluruhan. Ini menunjukkan bahwa SVM memiliki keunggulan dalam menangani data dengan hubungan fitur yang kompleks, sementara Naïve Bayes mungkin tidak selalu efektif dalam kasus-kasus dengan pola sentimen yang lebih rumit. Ketika dibandingkan, algoritma SVM memiliki nilai *F1-score* sebesar 100% menjadikan algoritma SVM sebagai pilihan yang terbaik dalam analisis ini. Performanya yang optimal menunjukkan kemampuan algoritma SVM dalam menangani data dengan presisi dan *recall* yang tinggi, sehingga dapat memberikan hasil yang lebih akurat dibandingkan algoritma Naïve Bayes.

C. Visualisasi Data

Pada penelitian ini untuk memudahkan mengidentifikasi kata-kata yang paling sering muncul dalam dataset, dalam memahami topik utama atau sentimen dominan, peneliti menggunakan eksperimen dengan memanfaatkan visualisasi *Wordcloud*. setiap kata ditampilkan dengan ukuran font yang bervariasi, di mana kata yang lebih besar menunjukkan frekuensi kemunculannya lebih tinggi dibandingkan dengan kata yang lebih kecil.

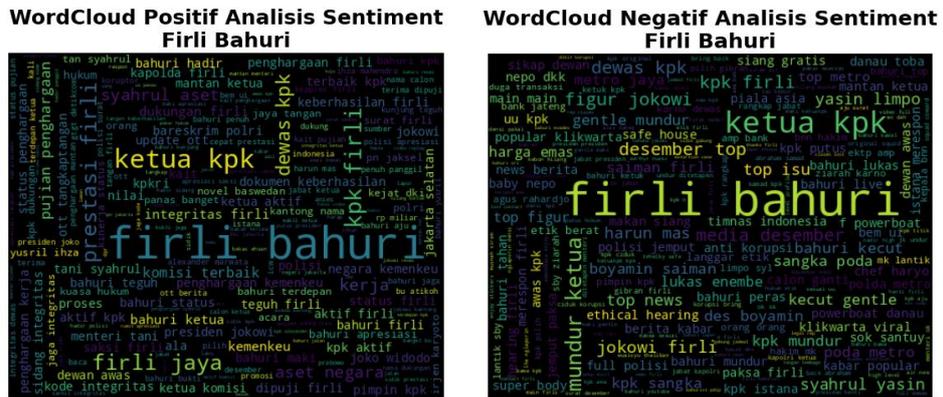
WordCloud Analisis Sentiment Firli Bahuri



Gambar 4. *Wordcloud* Firli Bahuri

Pada hasil *wordcloud* diatas menunjukkan bahwa kata Firli Bahuri adalah salah satu kata yg paling banyak

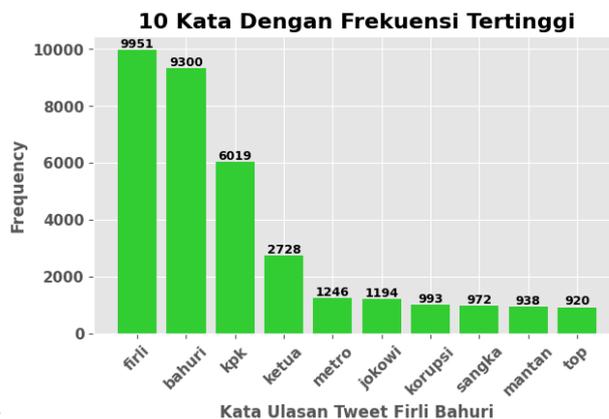
disebut dalam komentar, selain itu ada kata “ketua kpk”, “polda metro”, “yasin limpo”, “syahrul yasin”, “perkara firli”, “awas kpk” dan sebagainya. Dengan beberapa kata kunci yg disebut kan menunjukkan bahwa kebanyakan masyarakat di twitter membahasa tentang firli bahuri. sehingga diperlukan analisis lebih lanjut untuk menentukan apakah kata kunci tersebut termasuk dalam komentar positif atau negatif. Berikut ini adalah *Wordcloud* dari data komentar positif dan negatif tentang firli bahuri, yang dapat dilihat pada Gambar 5.



Gambar 5. Sentimen positif dan negatif

Berdasarkan pada sentimen negatif dari hasil *wordcloud* terdapat frekuensi kata yang sering muncul yaitu “Firli Bahuri”, “ketua kpk”, “mundur ketua”, “desember top”, “top isu”, “top news”, “bahuri peras” dan sebagainya. Pada hasil analisis dengan menggunakan *wordcloud* dapat dilihat bahwa kata “firli bahuri” menjadi fokus utama pada penelitian ini. Kata “firli bahuri” dengan frekuensi tertinggi menunjukkan bahwa *wordcloud* ini berfokus pada pribadi dan kinerja beliau selama di kpk. Kata “ketua kpk” dan “mundur ketua” menandai bahwa banyak dari pengguna twitter atau masyarakat yg ingin firli bahuri mundur dari ketua kpk. Sementara kata “desember top”, “top isu”, “yasin limpo”, “bahuri peras” mengindikasikan bahwa kasus suap dan pemerasan oleh firli bahuri terhadap yasin limpo menjadi isu dan berita teratas yang banyak dibicarakan dibulan desember.

Berdasarkan pada sentimen positif dari hasil *wordcloud* terdapat frekuensi kata yang sering muncul yaitu “firli bahuri”, “pujian penghargaan”, “kemenkue” “aset negara”, “integritas firli”, dan “dokumen keberhasilan”. Pada kata-kata tersebut, dapat dilihat bahwa *wordcloud* tersebut membahas tentang prestasi atau penghargaan firli bahuri selama menjabat sebagai ketua kpk. Kata “firli bahuri”, “pujian penghargaan”, “aset negara”, dan “kemenkue” menunjukkan bahwa firli bahuri mendapatkan penghargaan anugerah rekso bandha dari kementerian keuangan karena telah berhasil mengelola aset negara. Kata “integritas firli” mengindikasikan firli sebagai ketua kpk yg telah mencerminkan komitmen yang teguh terhadap prinsip-prinsip moral dan etika dalam pemberantasan korupsi. Ini tercermin dalam kejujuran dalam tindakan dan keputusan, transparansi dalam proses kerja, akuntabilitas yang ketat terhadap publik dan lembaga terkait, serta keteguhan dalam mempertahankan prinsip-prinsip anti-korupsi tanpa terpengaruh oleh tekanan eksternal. Sementara kata “update ott” mencakup perbaikan strategi dan teknologi untuk meningkatkan efektivitas dan transparansi operasi tangkap tangan dalam pemberantasan korupsi. Langkah ini bertujuan untuk memperkuat integritas KPK dan meningkatkan kepercayaan publik.



Gambar 6. Frekuensi kata tertinggi

Berdasarkan pada hasil 10 frekuensi kata tertinggi, kata "Firli" dan "Bahuri" menjadi kata yg memiliki frekuensi tertinggi dibandingkan kata lainnya, kata tersebut merujuk kepada nama dan identitas pribadi Ketua KPK yang sedang di analisis. Kata "KPK" merujuk sebagai lembaga yang dipimpinnya yang terfokus pada pemberantasan korupsi. Kata "korupsi" dan "sangka" sendiri merujuk kepada firli bahuri yang telah ditetapkan sebagai tersangka dalam kasus dugaan pemerasan terhadap SYL sejak November 2023 dan dijerat tindak pidana pemberantasan korupsi berupa pemerasan atau gratifikasi atau suap terkait dengan penanganan permasalahan hukum di Kementerian Pertanian Republik Indonesia pada kurun 2020-2023,. "Ketua" menunjukkan peran jabatan firli bahuri selamat dia menjabat, Kata "metro" merujuk pada Polda Metro Jaya yang menangani kasus firli bahuri selama masa penyidikan. Kata "Mantan" merujuk pada firli bahuri sendiri yang telah dicopot jabatannya sebagai ketua kpk oleh presiden Jokowi pada tanggal 28 Desember 2023, Kata "top" mengindikasikan perhatian dan sorotan publik terhadap kasus yang telah menjerat firli bahuri yang kebetulan menjadi trending topik di berbagai sosial media dan berita televisi.

D. Pembahasan Hasil

Dalam penelitian ini, analisis sentimen terhadap opini masyarakat Indonesia mengenai kepemimpinan Firli Bahuri di KPK menggunakan data sebanyak 7032 *tweet*. Penelitian ini membandingkan performa dua algoritma, yaitu Naïve Bayes dan Support Vector Machine (SVM), setelah dilakukan optimasi SMOTE untuk mengatasi ketidakseimbangan data. Hasil penelitian menunjukkan bahwa SVM memiliki performa yang sangat baik dengan mencapai akurasi, *precision*, *recall*, dan *F1-score* masing-masing sebesar 100%. Sebaliknya, algoritma Naïve Bayes hanya mencapai akurasi 97%. Temuan ini menunjukkan bahwa SVM tidak hanya mampu mengklasifikasikan sentimen dengan lebih akurat tetapi juga lebih konsisten dalam menangani data yang tidak seimbang.

Penelitian lain terkait opini masyarakat mengenai Luhut Binsar Pandjaitan di media sosial Twitter, yang menggunakan 4008 data *tweet*, juga membandingkan performa algoritma Naïve Bayes dan SVM dengan optimasi SMOTE untuk keseimbangan data. Hasilnya menunjukkan bahwa SVM kembali unggul dengan akurasi 99%, *precision* 99%, *recall* 100%, dan *F1-score* 99%, sedangkan Naïve Bayes hanya mencapai akurasi 95%. Hasil ini konsisten dengan penelitian terhadap Firli Bahuri, di mana SVM menunjukkan keunggulan yang jelas dalam mengklasifikasikan sentimen dengan lebih akurat dan konsisten dibandingkan Naïve Bayes. [22].

Perbandingan hasil dari kedua penelitian ini menunjukkan bahwa SVM mampu menangani analisis sentimen dengan sangat baik di berbagai topik, baik itu mengenai kepemimpinan Firli Bahuri maupun Luhut Binsar Pandjaitan. Dalam kedua kasus, SVM menunjukkan keunggulan yang jelas dalam hal akurasi dan konsistensi hasil dibandingkan dengan Naïve Bayes. Temuan ini memperkuat validasi bahwa SVM adalah pilihan yang lebih baik untuk analisis sentimen di media sosial Twitter. Selain itu, penelitian ini juga memberikan kontribusi signifikan dalam metodologi analisis sentimen publik di Indonesia, memberikan bukti empiris yang kuat untuk penggunaan SVM dalam penelitian dan aplikasi praktis di masa depan.

IV. KESIMPULAN

Berdasarkan hasil dari penelitian ini telah dilakukan analisis sentimen terhadap opini masyarakat Indonesia tentang kepemimpinan firli bahuri di kpk. data dalam penelitian ini menggunakan 7032 data mengenai firli bahuri. Dalam penelitian ini dilakukan perbandingan model algoritma naïve bayes dan support vector machine. Sebelum dilakukan perbandingan antara kedua model, penelitian ini terlebih dahulu melakukan proses optimasi SMOTE untuk mengatasi ketidakseimbangan data agar data minority dan data majority dapat memiliki jumlah data yang sama. Hasil dari perbandingan kedua model mendapatkan nilai algoritma Naïve bayes dengan akurasi 97% dan algoritma support vector machine dengan akurasi 100%. Jika dilihat pada perbandingan terkait tingkat akurasi pada kedua model dapat disimpulkan bahwa algoritma dengan dengan performa terbaik yaitu support vector machine dengan nilai akurasi 100%, *precision* 100%, *recall* 100%, dan *FI-score* 100%. Berdasarkan hasil tersebut, dapat disimpulkan bahwa dalam penelitian ini, algoritma Support Vector Machine (SVM) menunjukkan performa yang sangat baik dibandingkan dengan algoritma Naïve bayes dalam melakukan analisis sentimen terhadap opini masyarakat Indonesia mengenai kepemimpinan firli bahuri di kpk pada media sosial twitter. Walaupun kedua algoritma memiliki akurasi yang tinggi pada setiap tahapannya, SVM unggul dengan nilai nilai akurasi 100%, *precision* 100%, *recall* 100%, dan *FI-score* 100%. Hasil ini menunjukkan bahwa SVM mampu melakukan proses analisis sentimen dengan sangat akurat pada setiap tahapannya dibandingkan dengan naïve bayes. Oleh karena itu,

dalam penelitian ini terkait opini masyarakat di media sosial twitter terhadap kepemimpinan firli bahuri di kpk, dapat disimpulkan bahwa algoritma SVM menjadi pilihan yang lebih baik dibandingkan algoritma Naïve bayes.

DAFTAR PUSTAKA

- [1] R. Nooraeni, A. Fikri Fadhilah I, H. Dwi, S. Fatimatul, S. Pertiwi, and Y. Ronaldias, "Analisis Sentimen Data Twitter Mengenai Isu RUU KPK Dengan Metode Support Vector Machine (SVM)," *Jurnal Paradigma, Jurnal Komputer dan Informatika*, vol. 22, no. 1, pp. 2579–3500, 2020, doi: 10.31294/p.v21i2.
- [2] N. Wayan Ayu Sekar Aprilia and A. Rahman Isnain, "JURNAL MEDIA INFORMATIKA BUDIDARMA Analisis Sentimen Terhadap Media Sosial Twitter dengan Kasus Kampanye Anti-Korupsi di Indonesia Menggunakan Naive Bayes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 2, pp. 695–703, 2024, doi: 10.30865/mib.v8i2.7582.
- [3] Haris Fadhil, "Akhir Drama Firli Bahuri di KPK," detikNews. Accessed: Jun. 14, 2024. [Online]. Available: <https://news.detik.com/berita/d-7114211/akhir-drama-firli-bahuri-di-kpk>
- [4] Admin DPMPTSP, "Skandal Korupsi Ketua KPK: Firli Bahuri Tersangka Kasus Pemerasan Terhadap Mantan Menteri Pertanian," <https://dpmptsp.acehprov.go.id/berita/kategori/blog/delegitimasi-putusan-mk-perspektif-politik>. Accessed: Jun. 14, 2024. [Online]. Available: <https://dpmptsp.acehprov.go.id/berita/kategori/blog/delegitimasi-putusan-mk-perspektif-politik>
- [5] Siaran Pers, "Menanggapi Penetapan Tersangka Firli Bahuri dan Bukti Konkrit Pelemahan Agenda Pemberantasan Korupsi di Indonesia," antikorupsi.org. Accessed: Jun. 14, 2024. [Online]. Available: <https://www.antikorupsi.org/id/menanggapi-penetapan-tersangka-firli-bahuri-dan-bukti-konkrit-pelemahan-agenda-pemberantasan>
- [6] M. I. Petiwi, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 1, pp. 542–550, Jan. 2022, doi: 10.30865/mib.v6i1.3530.
- [7] R. R. Ratino, N. H. H. Hafidz, S. A. A. Anggraeni, and W. G. G. Gata, "Sentimen Analisis Informasi Covid-19 menggunakan Support Vector Machine dan Naïve Bayes," *JUPITER: Jurnal Penelitian Ilmu dan Teknologi Komputer*, vol. 12, no. 2, pp. 1–11, 2020, Accessed: Jun. 23, 2024. [Online]. Available: <https://jurnal.polsri.ac.id/index.php/jupiter/article/view/2388>
- [8] P. Elisa and A. R. Isnain, "COMPARISON OF RANDOM FOREST, SUPPORT VECTOR MACHINE AND NAIVE BAYES ALGORITHMS TO ANALYZE SENTIMENT TOWARDS MENTAL HEALTH STIGMA," *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 1, pp. 321–329, 2024, Accessed: Jun. 23, 2024. [Online]. Available: <http://jutif.if.unsoed.ac.id/index.php/jurnal/article/view/1817>
- [9] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *Jurnal Ilmiah EduTic: Pendidikan dan Informatika*, vol. 7, no. 1, pp. 1–11, 2020, Accessed: Jun. 24, 2024. [Online]. Available: <https://journal.trunojoyo.ac.id/edutic/article/view/8779>
- [10] B. Pamungkas, M. E. Purbaya, and D. J. AK, "Analisis Sentimen Twitter Menggunakan Metode Support Vector Machine (SVM) pada Kasus Benih Lobster 2020," *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, vol. 3, no. 2, pp. 10–20, 2021, Accessed: Jun. 25, 2024. [Online]. Available: <https://journal.itelkom-pwt.ac.id/index.php/inista/article/view/243>
- [11] A. S. Rahayu, A. Fauzi, and R. Rahmat, "Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Spotify," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 2, pp. 349–354, 2022, Accessed: Jul. 18, 2024. [Online]. Available: <https://ejurnal.stmik-budidarma.ac.id/index.php/JSON/article/view/5398/3175>
- [12] B. Ramadhani and R. R. Suryono, "Komparasi Algoritma Naïve Bayes dan Logistic Regression Untuk Analisis Sentimen Metaverse," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 2, pp. 714–725, 2024, Accessed: Jun. 27, 2024. [Online]. Available: <http://www.stmik-budidarma.ac.id/ejurnal/index.php/mib/article/view/7458>
- [13] A. Z. Amrullah, A. S. Anas, and M. A. J. Hidayat, "Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square," *Jurnal Bumigora Information Technology (BITe)*, vol. 2, no. 1, pp. 40–44, 2020, Accessed: Jun. 28, 2024. [Online]. Available: <https://journal.universitasmumigora.ac.id/index.php/bite/article/view/804>
- [14] A. Perdana, A. Hermawan, and D. Avianto, "Analisis Sentimen Terhadap Isu Penundaan Pemilu di Twitter Menggunakan Naive Bayes Classifier," *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, vol. 11, no. 2, pp. 195–200, 2022.
- [15] R. R. S. Elvika Alya Junita, "ANALISIS SENTIMEN HATE SPEECH MENGENAI CALON WAKIL PRESIDEN INDONESIA MENGGUNAKAN ALGORITMA BERT," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 3, pp. 1–13, Jun. 2019, doi: 10.29100/jipi.v4i1.781.
- [16] F. Rozi and K. Khomsatun, "ANALISIS SENTIMEN APLIKASI BCA MOBILE MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 4, no. 1, pp. 1–11, Jun. 2019, doi: 10.29100/jipi.v4i1.781.
- [17] A. M. Pravina, I. Cholisoddin, and P. P. Adikara, "Analisis sentimen tentang opini maskapai penerbangan pada dokumen twitter menggunakan algoritme support vector machine (svm)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 3, pp. 2789–2797, 2019, Accessed: Jun. 30, 2024. [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4793>
- [18] N. S. Marga, "Sentimen Analisis Tentang Kebijakan Pemerintah Terhadap Kasus Corona Menggunakan Metode Naive Bayes," *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, vol. 2, no. 4, pp. 453–463, 2021, Accessed: Jul. 02, 2024. [Online]. Available: <https://jim.teknokrat.ac.id/index.php/informatika/article/view/1602>
- [19] D. D. Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 1, 2022, Accessed: Jul. 03, 2024. [Online]. Available: <https://journal.eng.unila.ac.id/index.php/jitet/article/view/2262>
- [20] A. R. Isnain, A. I. Sakti, D. Alita, and N. S. Marga, "Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm," *Jurnal Data Mining Dan Sistem Informasi*, vol. 2, no. 1, pp. 31–37, 2021, Accessed: Jul. 04, 2024. [Online]. Available: <https://ejurnal.teknokrat.ac.id/index.php/JDMSI/article/view/1021>
- [21] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis sentimen terhadap layanan indihome berdasarkan twitter dengan metode klasifikasi support vector machine (SVM)," *Jurnal Media Informatika Budidarma*, vol. 4, no. 3, pp. 650–658, 2020, Accessed: Jul. 04, 2024. [Online]. Available: <https://ejurnal.stmik-budidarma.ac.id/index.php/mib/article/view/2181>
- [22] Ardiyansah and Parjito, "KLIK: Kajian Ilmiah Informatika dan Komputer Perbandingan Metode Naive Bayes dan Support Vector Machine Dalam Analisis Sentimen Terhadap Tokoh Publik," *Media Online*, vol. 4, no. 6, pp. 2813–2821, 2024, doi: 10.30865/klik.v4i6.1928.
- [23] H. Hidayatullah, P. Purwanto, and Y. Umaidah, "Penerapan Naive Bayes Dengan Optimasi Information Gain Dan Smote Untuk Analisis Sentimen Pengguna Aplikasi Chatgpt," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 3, pp. 1546–1553, 2023, Accessed: Jul. 04, 2024. [Online]. Available: <https://ejournal.itn.ac.id/index.php/jati/article/view/6887>
- [24] F. A. Indriyani, A. Fauzi, and S. Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naive bayes dan support vector machine," *TEKNOSAINS: Jurnal Sains, Teknologi dan Informatika*, vol. 10, no. 2, pp. 176–184, 2023, Accessed: Jul. 05, 2024. [Online]. Available: <https://jurnal.stmcileungsi.ac.id/index.php/tekn/article/view/419>