



Analisis Sentimen Terkait Hilirisasi Industri Pada Opini Masyarakat X dengan Menggunakan Naive Bayes

Aditya Budi Pratama, Dimas Febriawan*

Fakultas Teknik Industri dan Informatika, Teknik Informatik, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta

Jl. Limau II No.2, RT.3/RW.3, Kramat Pela, Kec. Kby. Baru, City, Daerah Khusus Ibukota Jakarta, Indonesia

Email: ¹aditbudi361@gmail.com, ²*dimas.febriawan@uhamka.ac.id

Email Penulis Korespondensi: dimas.febriawan@uhamka.ac.id

Submitted: 20/01/2025; Accepted: 31/01/2025; Published: 31/01/2025

Abstrak—Penelitian ini mengevaluasi sentimen masyarakat terhadap kebijakan hilirisasi industri di Indonesia menggunakan data dari X. Metode yang diterapkan adalah algoritma Naive Bayes untuk mengklasifikasikan opini publik ke dalam tiga kategori sentimen: positif, negatif, dan netral. Data dikumpulkan melalui proses crawling menggunakan API X dengan alat seperti Tweepy, kemudian diproses melalui tahapan preprocessing, termasuk pembersihan data, tokenisasi, konversi ke huruf kecil, penghapusan kata berhenti (stopword removal), serta stemming atau lemmatization. Setelah itu, data dilabeli secara manual menggunakan pendekatan berbasis kamus sentimen guna memastikan akurasi klasifikasi. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes mencapai tingkat akurasi sebesar 81,75% dalam mengklasifikasikan sentimen, dengan kinerja terbaik pada deteksi sentimen positif. Studi ini memberikan wawasan penting mengenai pandangan masyarakat terhadap kebijakan hilirisasi industri serta menawarkan rekomendasi bagi pembuat kebijakan untuk menyusun strategi yang lebih selaras dengan persepsi publik. Pemanfaatan X sebagai sumber data memungkinkan analisis secara real-time yang relevan dengan dinamika opini masyarakat. Pengumpulan data dimulai dari bulan September 2023 sampai bulan September 2024 dan mendapatkan data sebanyak 1400 data.

Kata Kunci: Hilirisasi Industri; Analisis Sentimen; Naive Bayes; X

Abstract—This research examines public sentiment toward Indonesia's industrial downstreaming policy using data sourced from X. The study employs the Naive Bayes algorithm to categorize public opinions into three sentiment types: positive, negative, and neutral. Data collection was conducted via a crawling process utilizing the X API and tools like Tweepy, followed by preprocessing steps such as data cleansing, tokenization, case normalization, stopword removal, and either stemming or lemmatization. Subsequently, the data was manually annotated using a lexicon-based sentiment method to ensure accurate classification. The findings reveal that the Naive Bayes algorithm achieved an accuracy rate of 81.75% in sentiment classification, with the highest performance observed in identifying positive sentiments. This research offers valuable insights into public perspectives on the industrial downstreaming policy and suggests recommendations for policymakers to develop strategies that better resonate with public sentiment. Leveraging X as a data source allows for real-time analysis that adapts to shifts in public opinion.

Keywords: Industrial Downstreaming; Sentiment Analysis; Naive Bayes; X

1. PENDAHULUAN

Indonesia merupakan negara dengan potensi sumber daya alam yang melimpah, yang seharusnya menjadi kekuatan utama dalam pembangunan ekonomi nasional. Namun, ketergantungan negara ini pada ekspor bahan mentah dan rendahnya nilai tambah produk menjadi tantangan besar dalam mencapai kemajuan ekonomi yang berkelanjutan. Untuk mengatasi masalah ini, pemerintah Indonesia telah mendorong hilirisasi industri sebagai langkah strategis untuk mengolah sumber daya alam secara lebih efisien dan meningkatkan daya saing global. Hilirisasi industri diharapkan tidak hanya meningkatkan nilai tambah produk, tetapi juga menciptakan lapangan pekerjaan, meningkatkan kesejahteraan masyarakat, dan memperkuat struktur ekonomi nasional.[1]

Namun, kebijakan hilirisasi industri tidak selalu mendapatkan respons yang sepenuhnya positif dari masyarakat. Berbagai pandangan muncul, baik yang mendukung maupun yang mengkritik pelaksanaan kebijakan ini. Dalam era digital saat ini, opini masyarakat terkait kebijakan hilirisasi industri banyak dituangkan melalui media sosial, salah satunya X, yang menjadi salah satu platform utama untuk berbagi pendapat dan berdiskusi mengenai isu-isu terkini. Oleh karena itu, menganalisis sentimen publik terhadap kebijakan ini melalui X dapat memberikan gambaran yang jelas tentang bagaimana masyarakat memandang dan merespons hilirisasi industri di Indonesia. Penerapan kebijakan hilirisasi ini telah memicu berbagai reaksi dari masyarakat. Di satu sisi, kebijakan ini dianggap mampu mendorong pertumbuhan ekonomi yang lebih berkelanjutan. Di sisi lain, ada yang berpendapat bahwa kebijakan ini menimbulkan tantangan baru, seperti kebutuhan infrastruktur yang lebih baik dan ketergantungan pada teknologi asing. [2]

Respon publik, khususnya yang terekam melalui media sosial seperti X, mencerminkan pandangan masyarakat terhadap kebijakan hilirisasi industri. Dengan jumlah pengguna aktif mencapai 24 juta pada Januari 2024 menurut data We Are Social, X menjadi platform strategis untuk menganalisis persepsi masyarakat secara cepat dan menyeluruh.[3] Analisis sentimen terhadap opini masyarakat di X dapat memberikan pemahaman mendalam tentang pandangan publik terhadap kebijakan ini. Penelitian ini menggunakan metode Naive Bayes untuk mengelompokkan opini masyarakat di X ke dalam kategori sentimen positif, negatif, atau netral, sekaligus mengidentifikasi faktor-faktor yang memengaruhi terbentuknya sentimen tersebut.[4]

Naive Bayes merupakan algoritma pembelajaran mesin yang sederhana namun sangat efektif untuk tugas klasifikasi teks. Mengacu pada Teorema Bayes, algoritma ini terbukti andal dalam melakukan analisis sentimen, terutama pada data dari media sosial. Penelitian sebelumnya menunjukkan bahwa Naive Bayes mampu mencapai tingkat akurasi lebih dari 80% dalam menganalisis sentimen data X, menjadikannya pilihan yang tepat untuk penelitian ini. Dalam studi ini, algoritma Naive Bayes akan diterapkan untuk menganalisis opini masyarakat di X mengenai kebijakan hilirisasi industri di Indonesia. Diharapkan, hasil analisis ini dapat memberikan wawasan lebih mendalam tentang persepsi publik terhadap kebijakan tersebut serta mendukung pengambil kebijakan dalam merancang strategi yang lebih efektif dan responsif terhadap opini masyarakat. [5]

Analisis sentimen terhadap opini masyarakat di X tentang hilirisasi industri memiliki pengaruh signifikan dalam hasil penelitian ini karena media sosial, khususnya X, telah menjadi sumber data yang kaya untuk memahami persepsi publik secara real-time. Informasi yang diperoleh dari analisis ini dapat memberikan wawasan mendalam tentang bagaimana kebijakan hilirisasi diterima oleh masyarakat, apakah cenderung didukung atau ditolak, serta faktor-faktor apa saja yang mempengaruhi sentimen tersebut.[6]

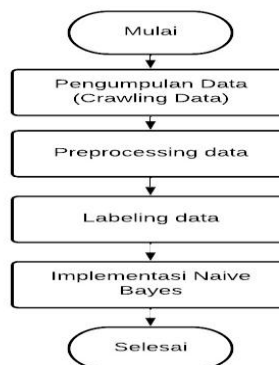
Analisis ini penting karena dapat membantu pembuat kebijakan memahami respons masyarakat dan menyesuaikan strategi mereka untuk mengurangi resistensi dan meningkatkan dukungan. Dalam konteks kebijakan publik, opini masyarakat yang tercermin dalam media sosial dapat menjadi indikator keberhasilan atau tantangan yang mungkin dihadapi oleh kebijakan tertentu. Jika dibandingkan dengan penelitian terdahulu, banyak studi telah menggunakan analisis sentimen untuk memahami persepsi masyarakat terhadap berbagai isu sosial dan ekonomi. Sebagai contoh, penelitian sebelumnya mungkin telah memanfaatkan analisis sentimen untuk mengukur reaksi publik terhadap kebijakan fiskal atau perubahan regulasi. Namun, penelitian yang secara khusus mengkaji opini masyarakat tentang hilirisasi industri di Indonesia melalui media sosial masih terbatas, sehingga penelitian ini menawarkan kontribusi baru dalam literatur akademik dan kebijakan publik. Metode Naive Bayes digunakan dalam penelitian ini karena kemampuannya yang sederhana namun efektif dalam tugas klasifikasi teks, terutama untuk analisis sentimen.[7] Algoritma ini didasarkan pada Teorema Bayes, yang dikembangkan oleh matematikawan Inggris, Thomas Bayes (1701-1761). Teorema Bayes merupakan dasar dari probabilitas kondisional, yang memungkinkan model untuk memperbarui kemungkinan hipotesis berdasarkan bukti baru yang diperoleh.[8] Meskipun Thomas Bayes adalah orang yang pertama kali merumuskan teorema ini, pengembangan algoritma Naive Bayes untuk klasifikasi teks dilakukan jauh setelahnya, terutama pada era modern dengan munculnya pembelajaran mesin

Naive Bayes disebut "naive" karena asumsi independensinya antar fitur, yang dalam analisis teks berarti bahwa kemunculan satu kata dalam sebuah dokumen dianggap tidak memengaruhi kemunculan kata lainnya. Meskipun asumsi ini sering kali tidak sepenuhnya sesuai dengan kenyataan, algoritma ini tetap menunjukkan kinerja yang sangat baik dalam berbagai aplikasi, termasuk analisis sentimen. Dalam penelitian ini, penerapan Naive Bayes diharapkan mampu menghasilkan klasifikasi sentimen masyarakat terhadap hilirisasi industri dengan tingkat akurasi yang memadai. Selain itu, hasil analisis ini diharapkan dapat memberikan wawasan yang berharga bagi pembuat kebijakan untuk lebih memahami pandangan publik dan menyesuaikan kebijakan mereka agar selaras dengan kebutuhan serta harapan masyarakat.[9]

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Dalam penelitian ini peneliti fokus pada penerapan metode Naive Bayes pada proses klasifikasi Analisis Sentimen Terkait Hilirisasi Industri Pada Opini Masyarakat X Dengan Menggunakan Naive Bayes dari data ulasan yang telah dikumpulkan. Dataset ulasan dikumpulkan menggunakan layanan X untuk pengolahan data. Lalu proses klasifikasi dilakukan dengan bantuan dari RapidMiner.[10] Gambar 1 menunjukkan proses tahapan metode penelitian.



Gambar 1. Tahapan Penelitian



Dalam penelitian ini, penulis melaksanakan serangkaian tahapan analisis sentimen terhadap komentar pengguna X. Proses dimulai dengan pengumpulan data melalui metode crawling menggunakan Tweepy pada platform X. Setelah data terkumpul, dilakukan tahap preprocessing untuk membersihkan dan mempersiapkan data sehingga siap digunakan. Data yang telah diproses kemudian diberi label sebagai sentimen positif atau negatif. Selanjutnya, algoritma Naive Bayes diterapkan untuk mengklasifikasikan data berdasarkan kategori sentimen tersebut. Evaluasi model dilakukan untuk mengukur kinerjanya dalam aspek akurasi, presisi, dan recall. Hasil penelitian menunjukkan bahwa algoritma ini mampu mengklasifikasikan sentimen pada dataset secara efektif. Temuan dari penelitian ini diharapkan dapat memberikan wawasan berharga bagi bisnis e-commerce dalam memanfaatkan analisis sentimen untuk merancang strategi yang lebih tepat dan efisien.[11]

2.2 Pengumpulan Data (Crawling Data)

Pengumpulan data, atau data crawling, adalah proses otomatis yang dilakukan untuk mengumpulkan informasi dari internet dengan menggunakan program komputer yang dikenal sebagai "web crawler" atau "spider." Proses ini bertujuan untuk mengindeks informasi dari berbagai halaman web untuk berbagai keperluan, seperti mesin pencari, analisis pasar, atau pengolahan data. Selama prosesnya, web crawler menelusuri halaman web, mengikuti tautan yang tersedia, dan mengumpulkan berbagai jenis konten, termasuk teks, gambar, serta metadata, yang kemudian diproses untuk analisis lebih lanjut.[12]

Tahap ini bertujuan untuk mengumpulkan data opini masyarakat dari platform media sosial X yang berkaitan dengan hilirisasi industri. Proses ini dilakukan dengan memanfaatkan API X atau alat crawling data seperti Tweepy. Dalam proses pengumpulan data, kata kunci yang relevan dengan hilirisasi industri digunakan untuk mengekstraksi tweet yang berhubungan dengan topik ini. Data yang dikumpulkan kemudian disimpan dalam format seperti CSV atau JSON agar mudah diolah pada tahap berikutnya.[13]

2.3 Preprocessing Data.

Pengolahan data (data processing) adalah proses mengubah data mentah menjadi informasi yang lebih berguna dan bermakna. Tahapan ini mencakup berbagai langkah yang dirancang untuk mengolah data sehingga menghasilkan output yang dapat dimanfaatkan untuk pengambilan keputusan atau analisis lanjutan[14]. Tahap ini sangat penting karena data yang diambil dari X biasanya masih mentah dan memerlukan pembersihan agar dapat digunakan untuk analisis lebih lanjut. Proses preprocessing mencakup beberapa langkah:

1. **Cleaning** adalah tahap krusial dalam pengolahan data yang bertujuan untuk meningkatkan kualitas data dengan mengidentifikasi, memperbaiki, atau menghapus data yang tidak lengkap, tidak akurat, atau tidak relevan. Proses ini bertujuan memastikan bahwa data yang digunakan untuk analisis atau pemodelan bebas dari kesalahan yang dapat memengaruhi hasil akhir. Cleaning mencakup penghapusan elemen seperti simbol, angka, tanda baca, tautan, dan karakter khusus lainnya yang tidak relevan untuk analisis.[15]
2. **Lowercasing** adalah proses mengubah semua huruf dalam teks menjadi huruf kecil. Langkah ini dilakukan untuk memastikan konsistensi dalam analisis teks, sehingga kata-kata yang sama dengan perbedaan kapitalisasi, seperti "Data" dan "data," dianggap identik oleh model analitik. Dengan menerapkan lowercasing, variasi akibat penggunaan huruf besar dan kecil dapat diminimalkan, sehingga membantu meningkatkan akurasi dalam pemrosesan bahasa alami dan tugas-tugas seperti klasifikasi teks. Menurut penelitian yang diterbitkan dalam *International Research Journal of Engineering and Technology (IRJET)*, lowercasing merupakan teknik preprocessing sederhana yang melibatkan konversi setiap token dalam teks menjadi huruf kecil, sehingga memastikan keseragaman dalam proses analisis lebih lanjut. [16]
3. **Stopword removal** adalah proses menghilangkan kata-kata umum yang sering muncul dalam teks tetapi memiliki nilai informasi rendah, seperti "dan," "atau," "yang," dan sebagainya. Proses ini bertujuan meningkatkan efisiensi dan efektivitas analisis teks dengan mengurangi fitur-fitur yang kurang relevan, sehingga model dapat lebih fokus pada kata-kata yang lebih penting dan bermakna. Menurut penelitian yang diterbitkan di arXiv dengan judul "Accelerating Text Mining Using Domain-Specific Stop Word Lists", penghapusan stopwords dapat mengurangi dimensionalitas data teks sekaligus meningkatkan kinerja dalam tugas penambangan teks. Studi tersebut juga menekankan pentingnya penggunaan daftar stopwords yang disesuaikan dengan domain tertentu untuk mencapai hasil analisis yang lebih optimal.[17]
4. **Labeling Data** adalah tahap di mana data yang telah melalui proses preprocessing diberi label sesuai dengan sentimennya. Pada tahap ini, setiap tweet dikategorikan ke dalam salah satu sentimen: positif, negatif, atau netral. Pemberian label ini dapat dilakukan secara manual oleh peneliti pada sampel data, yang kemudian akan digunakan sebagai data pelatihan. Sebagai alternatif, metode semi-supervised dapat digunakan, yang menggabungkan teknik manual dan otomatis untuk meningkatkan efisiensi proses pemberian label. Penting untuk menetapkan kriteria yang jelas untuk setiap kategori sentimen guna memastikan konsistensi dalam proses labeling.[19]

2.4 Implementasi Naive Bayes

Algoritma Naive Bayes digunakan untuk mengklasifikasikan sentimen pada data yang telah melalui tahap preprocessing. Proses dimulai dengan melatih model Naive Bayes menggunakan data yang telah diberi label.

Setelah proses pelatihan selesai, model diuji menggunakan data uji untuk mengevaluasi kinerjanya. Pengukuran performa dilakukan menggunakan metrik seperti akurasi, presisi, recall, dan F1-score, yang bertujuan menilai kemampuan model dalam mengklasifikasikan sentimen dari tweet dengan tepat.[20]

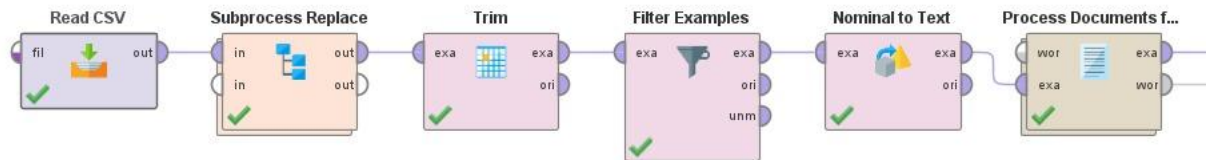
2.5 Selesai

Di mana semua hasil yang diperoleh dari analisis sentimen disusun dan dianalisis lebih lanjut. Hasil ini kemudian diinterpretasikan untuk memberikan wawasan yang lebih dalam mengenai opini masyarakat tentang kebijakan hilirisasi industri di Indonesia. Berdasarkan hasil analisis sentimen ini, rekomendasi atau saran yang bermanfaat dapat diberikan kepada pembuat kebijakan untuk memperbaiki atau menyesuaikan strategi mereka agar lebih sesuai dengan persepsi dan kebutuhan masyarakat.

3. HASIL DAN PEMBAHASAN

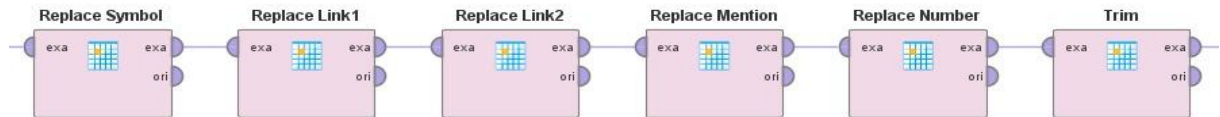
3.1 Preprocessing

Tahapan preprocessing meliputi cleansing, tokenizing, transform cases, filter stopwords. Alat yang digunakan untuk proses preprocessing adalah Rapid Miner. Gambar 3 menggambarkan proses cleansing, di mana tahap ini dimulai dengan membaca seluruh data ulasan menggunakan operator read CSV. Setelah itu, operator replace digunakan untuk menghilangkan karakter-karakter khusus dan tanda baca dari setiap ulasan yang muncul



Gambar 2. Proses Cleansing

Tahap pertama dalam processing data adalah adalah proses Cleansing, yang bertujuan untuk menghapus tanda baca serta karakter-karakter khusus dalam data ulasan, Seperti pada Tabel 1.

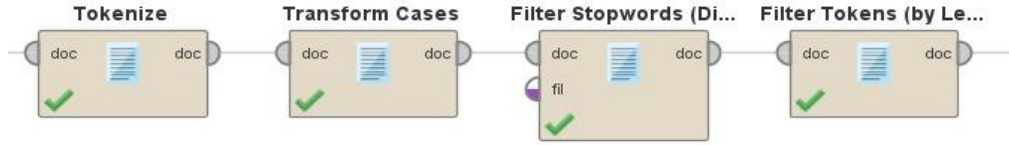


Gambar 3. Process document from data

Tabel 1. Proses Cleansing data

Sebelum Cleansing	Sesudah Cleansing
@kentvrama Asli ini baru hilirisasi industri kreatif. Investasi jangka panjang yang dilakukan Sampoerna di Indonesia sejalan dengan prioritas negara untuk mendorong hilirisasi industri inovasi dan penyerapan tenaga kerja.	hilirisasi industri kreatif investasi jangka panjang Sampoerna Indonesia prioritas negara hilirisasi industri inovasi penyerapan tenaga kerja
@Bisniscom Pemerintah memberikan apresiasi atas komitmen Freeport dan PT Amman Mineral Industri dalam membangun smelter di dalam negeri dengan memperpanjang relaksasi izin ekspor konsentrat tembaga. Semoga langkah ini mendorong terwujudnya hilirisasi industri.	apresiasi komitmen Freeport PT Amman Mineral Industri smelter dalam negeri relaksasi izin ekspor konsentrat tembaga hilirisasi industri
@i_kulitayam @madokafc1 @mwisbc @kompascom Lu masih gak paham berarti. Lu mau nuntut enak (pendidikan & kesehatan gratis) tapi gak siap dengan poin gak enak (pajak tinggi). Finlandia maju karena semua dikerjakan berbarengan, antara perbaikan SDM dan hilirisasi industri.	Finlandia maju pendidikan kesehatan gratis pajak tinggi SDM hilirisasi industri
@fariss_rachman Makanya jangan kemakan ajaran Greenpeace dkk di Eropa yang menentang ekspansi dan hilirisasi industri perkebunan sawit dkk karena itu salah satu sumber kekayaan yang sebenarnya sustainable buat Indonesia.	Greenpeace Eropa ekspansi hilirisasi industri perkebunan sawit sumber kekayaan sustainable Indonesia
Nah, Jokowi baru saja konfirmasi bakal perpanjang izin ekspor konsentrat tembaga PTFI. Ini jadi bentuk apresiasi atas upaya Freeport dan PT Amman Mineral Industri dalam hilirisasi dengan membangun smelter di dalam negeri.	Jokowi perpanjang izin ekspor konsentrat tembaga PTFI apresiasi upaya Freeport PT Amman Mineral Industri hilirisasi smelter dalam negeri

Menunjukkan bagian operator preprocessing. Setelah menyelesaikan tahap awal pembersihan, proses dilanjutkan dengan beberapa langkah berikutnya, yaitu tokenisasi, transformasi huruf, dan penyaringan stopwords.



Gambar 4. Proses Tokenzing

Langkah kedua adalah tokenzing. Tokenzing diperlukan untuk memecah teks menjadi kata-kata yang memiliki makna dalam sebuah kalimat. Kata-kata dalam ulasan diubah menjadi potongan-potongan menggunakan operator tokenizing, seperti yang ditunjukkan pada Tabel 2.[21]

Tabel 2. Hasil Tokenzing

Sebelum Tokenzing	Sesudah Tokenzing
@kentvrama Asli ini baru hilirisasi industri kreatif. Investasi jangka panjang yang dilakukan Sampoerna di Indonesia sejalan dengan prioritas negara untuk mendorong hilirisasi industri inovasi dan penyerapan tenaga kerja.	[hilirisasi, industri, kreatif, investasi, jangka, panjang, Sampoerna, Indonesia, prioritas, negara, hilirisasi, industri, inovasi, penyerapan, tenaga, kerja]
@Bisniscom Pemerintah memberikan apresiasi atas komitmen Freeport dan PT Amman Mineral Industri dalam membangun smelter di dalam negeri dengan memperpanjang relaksasi izin ekspor konsentrat tembaga. Semoga langkah ini mendorong terwujudnya hilirisasi industri.	[apresiasi, komitmen, Freeport, PT, Amman, Mineral, Industri, smelter, dalam, negeri, relaksasi, izin, ekspor, konsentrat, tembaga, hilirisasi, industri]
@i_kulitayam @madokafc1 @mwisbc @kompascom Lu masih gak paham berarti. Lu mau nuntut enak (pendidikan & kesehatan gratis) tapi gak siap dengan poin gak enak (pajak tinggi). Finlandia maju karena semua dikerjakan berbarengan, antara perbaikan SDM dan hilirisasi industri.	[Finlandia, maju, pendidikan, kesehatan, gratis, pajak, tinggi, SDM, hilirisasi, industri]
@fariss_rachman Makanya jangan kemakan ajaran Greenpeace dkk di Eropa yang menentang ekspansi dan hilirisasi industri perkebunan sawit dkk karena itu salah satu sumber kekayaan yang sebenarnya sustainable buat Indonesia.	[Greenpeace, Eropa, ekspansi, hilirisasi, industri, perkebunan, sawit, sumber, kekayaan, sustainable, Indonesia]
Nah, Jokowi baru saja konfirmasi bakal perpanjang izin ekspor konsentrat tembaga PTFI. Ini jadi bentuk apresiasi atas upaya Freeport dan PT Amman Mineral Industri dalam hilirisasi dengan membangun smelter di dalam negeri.	[Jokowi, perpanjang, izin, ekspor, konsentrat, tembaga, PTFI, apresiasi, upaya, Freeport, PT, Amman, Mineral, Industri, hilirisasi, smelter, dalam, negeri]

Tabel 3. Hasil Transform Case

Sebelum Transform Case	Sesudah Transform Case
@kentvrama Asli ini baru hilirisasi industri kreatif. Investasi jangka panjang yang dilakukan Sampoerna di Indonesia sejalan dengan prioritas negara untuk mendorong hilirisasi industri inovasi dan penyerapan tenaga kerja.	@kentvrama asli ini baru hilirisasi industri kreatif. investasi jangka panjang yang dilakukan sampoerna di indonesia sejalan dengan prioritas negara untuk mendorong hilirisasi industri inovasi dan penyerapan tenaga kerja.
@Bisniscom Pemerintah memberikan apresiasi atas komitmen Freeport dan PT Amman Mineral Industri dalam membangun smelter di dalam negeri dengan memperpanjang relaksasi izin ekspor konsentrat tembaga. Semoga langkah ini mendorong terwujudnya hilirisasi industri.	@bisniscom pemerintah memberikan apresiasi atas komitmen freeport dan pt amman mineral industri dalam membangun smelter di dalam negeri dengan memperpanjang relaksasi izin ekspor konsentrat tembaga. semoga langkah ini mendorong terwujudnya hilirisasi industri.
@i_kulitayam @madokafc1 @mwisbc @kompascom Lu masih gak paham berarti. Lu mau nuntut enak (pendidikan & kesehatan gratis) tapi gak siap dengan poin gak enak (pajak tinggi). Finlandia maju karena semua dikerjakan berbarengan, antara perbaikan SDM dan hilirisasi industri.	@i_kulitayam @madokafc1 @mwisbc @kompascom lu masih gak paham berarti. lu mau nuntut enak (pendidikan & kesehatan gratis) tapi gak siap dengan poin gak enak (pajak tinggi). finlandia maju karena semua dikerjakan berbarengan, antara perbaikan sdm dan hilirisasi industri.



Sebelum Transform Case	Sesudah Transform Case
<p>@fariss_rachman Makanya jangan kemakan ajaran Greenpeace dkk di Eropa yang menentang ekspansi dan hilirisasi industri perkebunan sawit dkk karena itu salah satu sumber kekayaan yang sebenarnya sustainable buat Indonesia.</p> <p>Nah, Jokowi baru saja konfirmasi bakal perpanjang izin ekspor konsentrat tembaga PTFI. Ini jadi bentuk apresiasi atas upaya Freeport dan PT Amman Mineral Industri dalam hilirisasi dengan membangun smelter di dalam negeri.</p>	<p>@fariss_rachman makanya jangan kemakan ajaran greenpeace dkk di eropa yang menentang ekspansi dan hilirisasi industri perkebunan sawit dkk karena itu salah satu sumber kekayaan yang sebenarnya sustainable buat indonesia.</p> <p>nah, jokowi baru saja konfirmasi bakal perpanjang izin ekspor konsentrat tembaga ptfi. ini jadi bentuk apresiasi atas upaya freeport dan pt amman mineral industri dalam hilirisasi dengan membangun smelter di dalam negeri.</p>

Tabel 4. Hasil Filter Stopword

Sebelum Filter Stopword	Sesudah Filter Stopword
<p>@kentvrama Asli ini baru hilirisasi industri kreatif. Investasi jangka panjang yang dilakukan Sampoerna di Indonesia sejalan dengan prioritas negara untuk mendorong hilirisasi industri inovasi dan penyerapan tenaga kerja.</p> <p>@Bisniscom Pemerintah memberikan apresiasi atas komitmen Freeport dan PT Amman Mineral Industri dalam membangun smelter di dalam negeri dengan memperpanjang relaksasi izin ekspor konsentrat tembaga. Semoga langkah ini mendorong terwujudnya hilirisasi industri.</p> <p>@i_kulitayam @madokafcl @mwisbc @kompascom Lu masih gak paham berarti. Lu mau nuntut enak (pendidikan & kesehatan gratis) tapi gak siap dengan poin gak enak (pajak tinggi). Finlandia maju karena semua dikerjakan berbarengan, antara perbaikan SDM dan hilirisasi industri.</p> <p>@fariss_rachman Makanya jangan kemakan ajaran Greenpeace dkk di Eropa yang menentang ekspansi dan hilirisasi industri perkebunan sawit dkk karena itu salah satu sumber kekayaan yang sebenarnya sustainable buat Indonesia. Nah, Jokowi baru saja konfirmasi bakal perpanjang izin ekspor konsentrat tembaga PTFI. Ini jadi bentuk apresiasi atas upaya Freeport dan PT Amman Mineral Industri dalam hilirisasi dengan membangun smelter di dalam negeri.</p>	<p>hilirisasi industri kreatif investasi jangka panjang Sampoerna Indonesia prioritas negara hilirisasi industri inovasi penyerapan tenaga kerja</p> <p>apresiasi komitmen Freeport PT Amman Mineral Industri smelter dalam negeri relaksasi izin ekspor konsentrat tembaga hilirisasi industri</p> <p>Finlandia maju pendidikan kesehatan gratis pajak tinggi SDM hilirisasi industri</p> <p>Greenpeace Eropa ekspansi hilirisasi industri perkebunan sawit sumber kekayaan sustainable Indonesia</p> <p>Jokowi perpanjang izin ekspor konsentrat tembaga PTFI apresiasi upaya Freeport PT Amman Mineral Industri hilirisasi smelter dalam negeri</p>

3.2 Labelling

Setelah tahap pembersihan data dalam proses preprocessing, jumlah data berkurang dari 1400 menjadi 1266, lalu kemudian diberikan sentiment data secara otomatis menjadi 582. Dataset ini kemudian dibagi menjadi dua bagian dengan rasio 40:60 untuk pelatihan dan pengujian, menghasilkan 580 data untuk pelatihan dari 1250 data.

Pelabelan dilakukan secara manual dengan pendekatan berbasis kamus (Lexicon-Based), menggunakan kamus sentimen untuk menentukan dominasi kata positif atau negatif dalam sebuah ulasan. Data yang telah dilabeli kemudian digunakan untuk melatih model Naive Bayes, sementara data uji dipakai untuk mengevaluasi kinerja model. Hasil pengujian dengan menggunakan 411 data uji menunjukkan performa model sebagai berikut: Akurasi: 81.75%, Precision (Positif): 86.63%, Recall (Positif): 92.01%, Precision (Negatif): 34.25%, Recall (Negatif): 34.25%. Model menunjukkan performa yang sangat baik untuk prediksi sentimen positif, tetapi performa untuk sentimen negatif perlu ditingkatkan, mungkin dengan menyeimbangkan dataset atau menggunakan teknik pengolahan data lainnya.

Tabel 5. Labelling Data

Ulasan	Sentimen
Hilirisasi industri kreatif, investasi jangka panjang Sampoerna di Indonesia, sejalan dengan prioritas negara untuk mendorong hilirisasi industri inovasi dan penyerapan tenaga kerja.	Positif
Apresiasi atas komitmen Freeport dan PT Amman Mineral Industri dalam membangun smelter di dalam negeri dengan memperpanjang relaksasi izin ekspor konsentrat tembaga.	Positif
Finlandia maju karena perbaikan SDM dan hilirisasi industri. Pendidikan dan kesehatan gratis namun dengan pajak tinggi.	Positif
Menentang ekspansi dan hilirisasi industri perkebunan sawit karena Greenpeace dan kelompok anti-industri, yang dapat menghambat sumber kekayaan yang sustainable.	Negatif

Ulasan

Sentimen

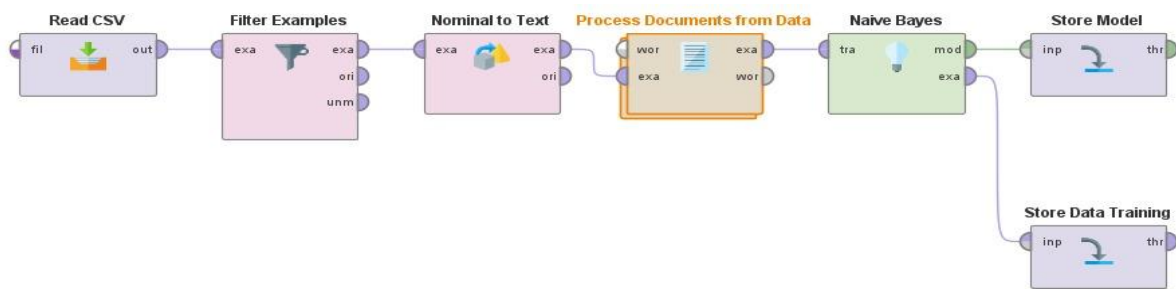
Jokowi memperpanjang izin ekspor konsentrat tembaga PTFI sebagai bentuk apresiasi atas upaya Freeport dan PT Amman Mineral Industri dalam hilirisasi dengan membangun smelter.

Positif

Data yang sudah dilabeli berdasarkan acuan dari kamus sentiment ((Lexicon-Based Approach, ada 4 4 ulasan yang termasuk kedalam ulasan positif dan 1 ulasan yang termasuk kedalam ulasan negative, seperti yang terlihat pada Tabel 5.

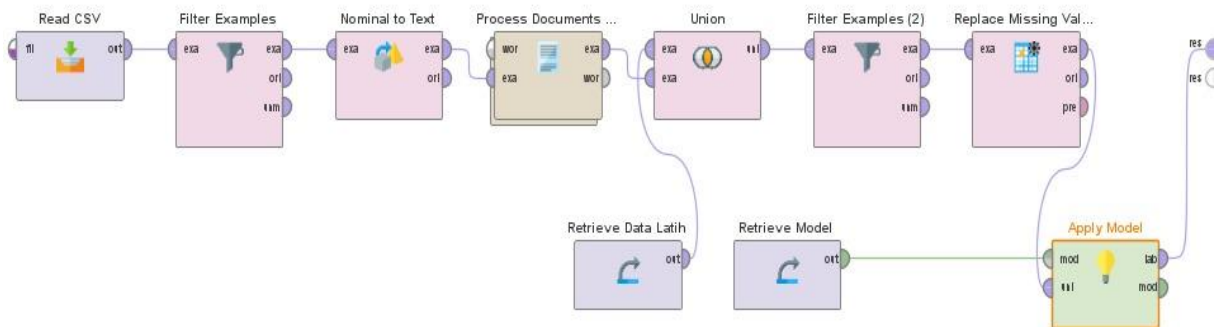
3.3 Implementasi Naïve Bayes

Tahapan implementasi Naive Bayes selanjutnya dijelaskan pada Gambar 5, dimulai dengan penggunaan operator read csv untuk membaca file yang telah diberi label manual maupun yang belum. Untuk menetapkan kolom sentimen sebagai label, operator set role dihubungkan dengan operator read csv. Kemudian, operator filter example digunakan untuk menyaring data yang tidak memiliki nilai kosong atau yang telah dilabeli manual. Selanjutnya, operator nominal to text diterapkan untuk mengubah data nominal menjadi format teks. Proses berikutnya mencakup process documents, yang melibatkan tokenisasi, transformasi huruf, dan pemfilteran stopword. Setelah itu, data dilatih menggunakan Naive Bayes dengan menghubungkan data latih dan model, kemudian hasil pelatihan disimpan.[22]



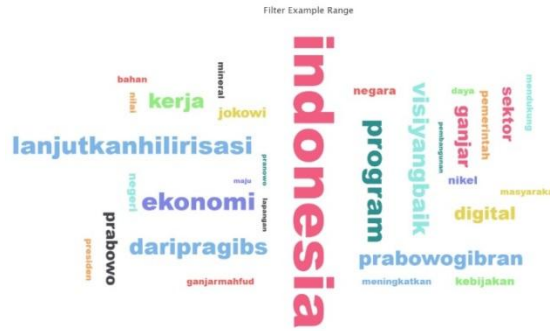
Gambar 5. Proses Pembuatan Data latih

Setelah dataset pelatihan disiapkan, langkah selanjutnya adalah penerapan Naive Bayes, yang digambarkan pada Gambar 5. Proses dimulai dengan pembacaan data menggunakan operator readCSV, kemudian dilakukan penyaringan data yang hilang atau belum dilabeli dengan menggunakan filter examples. Penjelasan mengenai operator nominal to text dan process document sama seperti pada tahap sebelumnya. Data hasil analisis dari store data latih digabungkan dengan hasil dari process document dan dihubungkan dengan filter examples, kemudian dilanjutkan dengan penggantian nilai yang hilang menggunakan replace missing values. Sementara itu, output dari storemodel dihubungkan dengan apply model untuk menghasilkan model prediksi.[13]



Gambar 6. Proses Pengimplementasian Naïve Bayes

Setelah melalui tahap preprocessing, pelabelan, dan penerapan algoritma Naïve Bayes, prediksi dilakukan terhadap 582 data ulasan. Dari jumlah tersebut, 349.2 ulasan dikategorikan sebagai sentimen positif, sedangkan 232.8 ulasan masuk ke dalam sentimen negatif. Hasil ini menunjukkan bahwa ulasan positif lebih dominan dibandingkan ulasan negatif, dengan selisih 116.4 ulasan. Analisis data mengungkap bahwa sentimen negatif terutama disebabkan oleh ketidakpuasan pengguna terhadap kebijakan industri, seperti ekspansi dan hilirisasi, yang dianggap merugikan sebagian pihak. Banyak ulasan negatif yang mengkritik kebijakan ekspansi industri perkebunan sawit beserta dampaknya terhadap ekonomi dan lingkungan. Hal ini tercermin dari kata-kata yang sering muncul, seperti "pajak tinggi," "Greenpeace," dan "ekspansi industri," yang menunjukkan ketidakpuasan terhadap kebijakan yang diterapkan oleh pemerintah dan perusahaan terkait. Untuk mengidentifikasi kata-kata yang paling sering digunakan oleh pengguna, hasil analisis Naïve Bayes divisualisasikan dalam bentuk wordcloud. Gambar 6 menampilkan kata-kata tersebut berdasarkan frekuensi kemunculan dari yang tertinggi hingga terendah.[23]



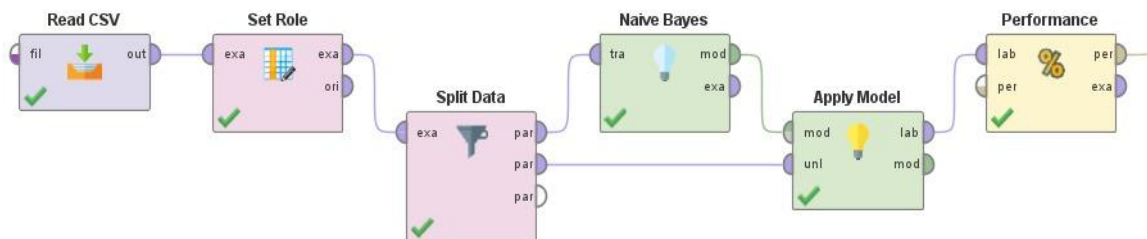
Gambar 7. Bentuk Wordcloud

3.4 Evaluasi

Dengan merujuk pada gambar confusion matrix yang disediakan, kita dapat mengevaluasi kinerja model menggunakan berbagai metrik evaluasi. Algoritma Naive Bayes diterapkan untuk mengklasifikasikan sentimen pada data yang telah diproses. Proses ini dimulai dengan melatih model Naive Bayes menggunakan data yang telah diberi label. Setelah pelatihan selesai, pengujian dilakukan menggunakan data uji untuk menilai kinerja model. Metrik seperti akurasi, presisi, recall, dan F1-score digunakan untuk mengukur sejauh mana model berhasil mengklasifikasikan sentimen dari tweet dengan akurat.[20] Akurasi model dihitung dengan membagi jumlah prediksi yang benar (baik positif maupun negatif) dengan total jumlah data yang diuji. Perhitungan ini dilakukan menggunakan rumus yang tepat untuk menghitung akurasi:

Tabel 6. Confusion Matrix

Evaluasi	Prediction Class		
	Positif	Negatif	
Actual Class Positif	True Positif (TP)	False Positis (FP)	
Actual Class Negatif	False Negatif (FN)	True Negatif (FN)	



Gambar 8. Proses Confussion Matrix

Hasil dari analisis confusion matrix menunjukkan bahwa model mencapai akurasi sebesar 81,75%, dengan 311 True Positive (TP), 27 False Negative (FN), 48 False Positive (FP), dan 25 True Negative (TN). Tabel 7 menyajikan hasil akurasi dari proses klasifikasi ini.

Tabel 7. Hasil Dari Klasifikasi

	True Negatif	True Positif	Class Precision
Pred. Negatif	25	27	40.08%
Pred. Positif	48	311	86.63%
Class Recall	34.25%	92.01%	

Hasil dari perhitungan akurasi, presisi, dan recall yang dilakukan menggunakan persamaan confusion matrix berikut ini

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{311+25}{311+27+48+25} = \frac{336}{411} = 81.75\%$$

$$Precision = \frac{TP}{TP+FP} = \frac{311}{311+48} = \frac{311}{359} = 86.63\%$$

$$Recall = \frac{TP}{TP+FN} = \frac{311}{311+27} = \frac{311}{338} = 92.01\%$$

$$F1-Score = \frac{2 \times Recall \times Precision}{Recall+Precision} = \frac{2 \times 0,9201 \times 0,8662}{0,9201+0,8662} = \frac{1,5939}{1,7863} = 0,8922 = 89 \%$$

Model memiliki akurasi tinggi sebesar 81.75% dan performa yang sangat baik dalam mendeteksi instance positif, dengan presisi 86.63% dan recall 92.01%. Namun, recall untuk kelas negatif hanya 34.25%, menunjukkan model kurang efektif dalam mendeteksi instance negatif. Secara keseluruhan, model lebih baik dalam



mengklasifikasikan instance positif dan memerlukan perbaikan untuk meningkatkan deteksi pada instance negatif agar hasil prediksi lebih seimbang.

4. KESIMPULAN

Penelitian ini berhasil mengklasifikasikan opini masyarakat mengenai kebijakan hilirisasi industri di Indonesia dengan menggunakan algoritma Naive Bayes. Hasil analisis menunjukkan bahwa mayoritas opini masyarakat di X bersifat positif terhadap kebijakan tersebut. Algoritma Naive Bayes mencapai akurasi **81,75%**, dengan precision untuk sentimen positif sebesar **86,63%** dan recall sebesar **92,01%**. Namun, model ini menunjukkan kinerja yang kurang memadai dalam mendeteksi sentimen negatif, dengan recall hanya sebesar **34,25%**. Kesimpulan ini mengindikasikan bahwa meskipun sebagian besar masyarakat mendukung kebijakan hilirisasi industri, ada kelompok yang menyatakan ketidakpuasan, terutama terkait dengan tantangan seperti infrastruktur dan ketergantungan pada teknologi asing. Penelitian ini memberikan kontribusi penting bagi pembuat kebijakan dalam memahami pandangan masyarakat terhadap kebijakan hilirisasi industri, yang dapat menjadi landasan untuk merancang strategi yang lebih efektif dan responsif terhadap opini publik. Selain itu, penelitian ini juga menyoroti pentingnya perbaikan model dalam mendeteksi sentimen negatif, sehingga analisis yang dihasilkan dapat lebih seimbang. Dalam konteks kebijakan publik, menganalisis opini masyarakat melalui media sosial seperti X menjadi alat yang sangat berharga untuk menilai keberhasilan kebijakan dan mengidentifikasi area yang perlu disesuaikan lebih lanjut.

REFERENCES

- [1] S. A. Damayanty. "Pengelolaan Industri Ekstraktif di Indonesia: Kebijakan Fiskal dan Tantangan ke Depan." Menggali Potensi Penerimaan Negara di Tengah Lesunya Ekonomi Global, pp. 259–283, 2018.
- [2] M. Hidayat and U. Budiyo. "Sentimen Analisis Tentang Hilirisasi Industri Berdasarkan Opini Masyarakat di X Menggunakan Metode K-Nearest Neighbor." Prosiding Seminar Nasional, vol. 2, no. September, pp. 826–835, 2023. Available: <https://senafiti.budiluhur.ac.id/index.php/senafiti/article/view/921>.
- [3] P. Irenniza Aulia, F. Tomi Saputra, and R. Hardiyanti. "Pengaruh Penggunaan Media Sosial X terhadap Pemenuhan Kebutuhan Informasi (Survei terhadap Pengikut Akun @Habisnontonfilm)." Jurnal Ilmiah Wahana Pendidik, vol. 10, no. 8, pp. 410–418, 2024. Available: <https://jurnal.peneliti.net/index.php/JIWP>.
- [4] Mubarak, R. "Analisis Sentimen Pengguna X terhadap Kebijakan Pemberlakuan Pembatasan Sosial Berskala Besar (PSBB) dengan Metode Naive Bayes." Jurnal Siliwangi Seri Sains dan Teknologi, vol. 7, no. 1, pp. 19–24, 2021. Available: <http://jurnal.unsil.ac.id/index.php/jssainstek/article/view/3726>.
- [5] Rahman, R. F., and Irwiensyah, F. "Analisis Sentimen Ulasan Pengguna Aplikasi Alibaba.com pada Marketplace Indonesia." Journal of System Computing (JOSYC), vol. 6, no. 1, pp. 64–72, 2024. doi: 10.47065/josyc.v6i1.6132..
- [6] Febriyani, E., and Februariyanti, H. "Analisis Sentimen terhadap Program Kampus Merdeka Menggunakan Algoritma Naive Bayes Classifier di X." Jurnal Tekno Kompak, vol. 17, no. 1, p. 25, 2023. doi: 10.33365/jtk.v17i1.2061.
- [7] Krisdiyanto, T. "Analisis Sentimen Opini Masyarakat Indonesia terhadap Kebijakan PPKM pada Media Sosial X Menggunakan Naive Bayes Classifiers." Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi, vol. 7, no. 1, p. 32, 2021. doi: 10.24014/coreit.v7i1.12945.
- [8] Saputra, R., and Hasan, F. N. "Analisis Sentimen terhadap Program Makan Siang & Susu Gratis Menggunakan Algoritma Naive Bayes." Jurnal Teknologi dan Sistem Informasi Bisnis, vol. 6, no. 3, pp. 411–419, 2024. doi: 10.47233/jteksis.v6i3.1378.
- [9] Martiti, and Juliane, C. "Implementation of Naive Bayes Algorithm on Sentiment Analysis Application." Proceedings of the 2nd International Seminar on Science and Applied Technology (ISSAT 2021), vol. 207, pp. 193–200, 2021. doi: 10.2991/aer.k.211106.030.
- [10] Diana, S. R., and Farida, F. "Applying Bag of Words Approach to Determine Remote Sensing Technology Acceptance among Smallholder Plantations." Arab Gulf Journal of Scientific Research, vol. 42, pp. 904–919, July 2023. doi: 10.1108/AGJSR-02-2023-0056.
- [11] Aditya, D. R., Supriyati, E., and Listyorini, T. "Analisis Sentimen Pengguna X terhadap Rokok Elektrik (Vape) di Indonesia Menggunakan Metode Naive Bayes." Jurnal Ilmiah Penelitian dan Pembelajaran Informatika (JIPI), vol. 7, no. 1, pp. 43–50, 2022. doi: 10.29100/jipi.v7i1.2145.
- [12] Danang, R., et al. "Implementasi Web Crawling pada Website BPS Jawa Timur pada Data Inflasi di Jawa Timur." UMAT: Jurnal Pengabdian Al-Shobar untuk Masyarakat, vol. 1, no. 2, pp. 6–13, 2024. Available: <https://jurnal.alshobar.or.id/index.php/umat>.
- [13] Khofifah, W., Rahayu, D. N., and Yusuf, A. M. "Analisis Sentimen Menggunakan Naive Bayes untuk Melihat Review Masyarakat terhadap Tempat Wisata Pantai di Kabupaten Karawang pada Ulasan Google Maps." Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi, vol. 16, no. 4, pp. 28–38, 2022. doi: 10.35969/interkom.v16i4.192.
- [14] Fathiarahma, A., Voutama, A., Ridwan, T., & Heryana, N. (2023). "Analisis Text Mining Klasifikasi Kegiatan Keluarga menggunakan Orange dengan Metode Naive Bayes." Jurnal Teknologi Terpadu, 9(1), 35–41. doi: 10.54914/jtt.v9i1.606..
- [15] A. Fathiarahma, A. Voutama, T. Ridwan, and N. Heryana, "Analisis Text Mining Klasifikasi Kegiatan Keluarga menggunakan Orange dengan Metode Naive Bayes," J. Teknol. Terpadu, vol. 9, no. 1, pp. 35–41, 2023, doi: 10.54914/jtt.v9i1.606.
- [16] Jakhotiya, A., Jain, H., Jain, B., & Chaniyara, C. (2022). "Text Pre-Processing Techniques in Natural Language Processing: A Review." International Research Journal of Engineering and Technology, 9(2), 878–880.



- [17] Alshaniq, F., Apon, A., Herzog, A., Safro, I., & Sybrandt, J. (2020). "Accelerating Text Mining Using Domain-Specific Stop Word Lists." In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data) (pp. 2639–2648). doi: 10.1109/BigData50022.2020.9378226.
- [18] Mandala, R., Koryanti, E., Munir, R., & Harlili. (2004). "Sistem Stemming Otomatis untuk Kata dalam Bahasa Indonesia." In Seminar Nasional Aplikasi Teknologi Informasi (pp. 29–36).
- [19] Noviana, R., & Rasal, I. (2023). "Penerapan Algoritma Naive Bayes dan SVM untuk Analisis Sentimen Boy Band BTS pada Media Sosial X." *Jurnal Teknik dan Sains*, 2(2), 51–60. doi: 10.56127/jts.v2i2.791.
- [20] Ratnawati, F. (2018). "Implementasi Algoritma Naive Bayes terhadap Analisis Sentimen Opini Film pada Twitter." *Jurnal Inovtek Polbeng Seri Informatika*, 3(1), 50–59. doi: 10.35314/isi.v3i1.335.
- [21] Agustina, N., Citra, D. H., Purnama, W., Nisa, C., & Kurnia, A. R. (2022). "Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store." *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 2(1), 47–54. doi: 10.57152/malcom.v2i1.195..
- [22] Faesal, A., Muslim, A., Ruger, A. H., & Kusriani, K. (2020). "Sentimen Analisis terhadap Komentar Konsumen terhadap Produk Penjualan Toko Online menggunakan Metode K-Means." *MATRIK: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, 19(2), 207–213. doi: 10.30812/matrik.v19i2.640.
- [23] Casandy, M. H., & Mahdiana, D. (2022). "Penerapan Algoritma Naive Bayes untuk Melakukan Analisis Sentimen pada PT Pos Indonesia (Persero)." *KRESNA: Jurnal Riset dan Pengabdian Masyarakat*, 2(2), 221–229. doi: 10.36080/jk.v2i2.51.