

# Analisis Sentimen Masyarakat Terhadap Kebijakan Ekspor Pasir Laut Berdasarkan Ulasan Twitter Menggunakan Algoritma Naive Bayes dan Support Vector Machine

Zarqani Zarqani<sup>\*</sup>, M Afdal, Rice Novita, Megawati Megawati

Fakultas Sains dan Teknologi, Sistem Informasi, Universitas Islam Negeri Sultan Syarif Kasim, Pekanbaru, Indonesia

Email: <sup>1,\*</sup>12150314283@students.uin-suska.ac.id, <sup>2</sup>m.afdal.uin-suska.ac.id, <sup>3</sup>rice.novita@uin-suska.ac.id,

<sup>4</sup>megawati@uin-suska.ac.id

Email Penulis Korespondensi: 12150314283@students.uin-suska.ac.id

Submitted: 24/05/2025; Accepted: 19/06/2025; Published: 05/09/2025

**Abstrak**—Ekspor pasir laut telah dilarang sejak tahun 2003 melalui Surat Keputusan Menteri Perindustrian dan Perdagangan. Namun, pada 15 Mei 2023, Presiden Joko Widodo kembali mengizinkan ekspor pasir laut melalui Peraturan Pemerintah No. 26 Tahun 2023. Kebijakan ini memicu kontroversi dan menjadi viral di media sosial, salah satunya di platform Twitter. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap kebijakan tersebut berdasarkan ulasan di Twitter dengan menggunakan algoritma Naive Bayes dan Support Vector Machine (SVM). Data dikumpulkan melalui teknik crawling, kemudian diproses dengan metode text preprocessing, pembobotan kata menggunakan TF-IDF, serta random oversampling untuk menyeimbangkan data. Selanjutnya, data dikategorikan ke dalam empat variabel tematik, yaitu ekonomi, lingkungan, sosial, dan kebijakan geologis, untuk melihat distribusi sentimen yang lebih terfokus. Hasil analisis terhadap 2.765 data menunjukkan bahwa mayoritas sentimen bersifat negatif (55%), yang mengindikasikan kecenderungan penolakan publik terhadap kebijakan ekspor pasir laut, diikuti sentimen netral (30%) dan positif (15%). Evaluasi performa menunjukkan bahwa SVM unggul pada kategori Ekonomi dengan akurasi hampir 95%, sementara pada kategori lain perbedaannya dengan Naive Bayes relatif tipis. Penelitian ini diharapkan dapat memberikan wawasan mengenai persepsi masyarakat Indonesia terhadap kebijakan ekspor pasir laut dan implikasinya di berbagai sector.

**Kata kunci:** Analisis Sentimen; Ekspor Pasir Laut; Naive Bayes; Support Vector Machine; Twitter

**Abstract**—The export of sea sand has been banned since 2003 through a Decree of the Minister of Industry and Trade. However, on May 15, 2023, President Joko Widodo once again allowed the export of sea sand through Government Regulation No. 26 of 2023. This policy sparked controversy and went viral on social media, including on Twitter. This study aims to analyze public sentiment toward the policy based on reviews on Twitter using the Naive Bayes and Support Vector Machine (SVM) algorithms. Data was collected through crawling techniques, then processed using text preprocessing methods, word weighting using TF-IDF, and random oversampling to balance the data. The data was then categorized into four thematic variables—economy, environment, social, and geological policy—to examine a more focused distribution of sentiment. Analysis of 2,765 data points revealed that the majority of sentiment was negative (55%), indicating public opposition to the sea sand export policy, followed by neutral sentiment (30%) and positive sentiment (15%). Performance evaluation shows that SVM excels in the Economy category with nearly 95% accuracy, while in other categories the difference with Naive Bayes is relatively small. This study is expected to provide insights into the Indonesian public's perception of the sea sand export policy and its implications across various sectors.

**Keywords:** Sentiment Analysis; Sea Sand Export; Naive Bayes; Support Vector Machine; Twitter

## 1. PENDAHULUAN

Kebijakan ekspor pasir laut merupakan salah satu topik yang menjadi perhatian penting dalam konteks ekonomi dan lingkungan di Indonesia karena banyak diminati untuk berbagai keperluan, seperti konstruksi, manufaktur, dan industri lainnya. Tingginya permintaan pasir laut telah menjadikan kegiatan penambangan dan eksportnya sebagai topik perdebatan yang penting terkait dampak lingkungan, sosial, dan ekonomi [1].

Pasir laut adalah jenis pasir yang ditemukan di pantai atau dasar laut. Material ini terbentuk dari hasil erosi batuan dan mineral di daratan, yang kemudian dibawa oleh air dan diendapkan di wilayah pantai atau dasar laut [2]. Penambangan pasir adalah aktivitas ekstraksi atau pengambilan pasir dari alam. Pasir merupakan material yang sering dimanfaatkan dalam berbagai industri, seperti konstruksi, pembuatan beton, kaca, dan produksi logam [2]. Dengan metode yang digunakan bervariasi tergantung pada lokasi, skala operasi, dan jenis pasir yang ditambang.

Pada tahun 2003 diberlakukanlah larangan ekspor pasir laut melalui Surat Keputusan Menperindag mengenai penghentian sementara ekspor pasir laut. Namun, pada 15 Mei 2023, Presiden Joko Widodo mengizinkan kembali ekspor pasir laut melalui Peraturan Pemerintah No. 26 Tahun 2023, dalam Pasal 9 ayat (1) disebutkan bahwa hasil sedimentasi di laut berupa dua jenis, yakni diatur dalam huruf (a) pasir laut dan huruf (b) material sedimen lain misalnya yang berbentuk lumpur. Kemudian dalam Pasal 9 ayat (2), berbunyi "Pemanfaatan Hasil Sedimentasi di Laut berupa pasir laut sebagaimana dimaksud pada ayat (1) huruf a digunakan untuk ekspor sepanjang kebutuhan dalam negeri terpenuhi dan sesuai dengan ketentuan peraturan perundang-undangan. Tentu, dengan diizinkan kembali ekspor pasir laut banyak memicu kritik dan pertentangan dari berbagai pihak berkepentingan [3], semenjak dibukanya kembali kebijakan ekspor pasir laut ini, sudah ada 66 perusahaan yang mengajukan permohonan rekomendasi ekspor pasir laut kepada Kementerian Kelautan dan Perikanan, dua diantaranya adalah PT Gajamina Sakti Nusantara dan PT Bumi Lautan Samudera.

Penambangan pasir laut membawa dampak serius terhadap ekosistem laut dan kehidupan masyarakat pesisir. Aktivitas pengerukan pasir laut seringkali mengakibatkan erosi pantai, membahayakan ekosistem pesisir, dan merusak

habitat laut [4]. Dampak terhadap kehidupan masyarakat pesisir adalah hilangnya area berburu ikan karena sudah diambil oleh perusahaan tambang, sehingga berdampak terhadap perubahan ekonomi para nelayan, seorang nelayan bernama daeng senja yang berusia 45 tahun dari Dusun Sampulungan, biasanya mencari nafkah dengan memancing menggunakan perahu jolloro. Ia melaut sekitar 20 hingga 24 jam, dengan jarak tangkap 1 hingga 8 mil dari pesisir pantai Galesong, fokus pada ikan-ikan kecil seperti ikan merah kecil dan ikan katombo. Namun, setelah maraknya penambangan pasir di wilayah pesisir Galesong, Daeng Senja dan nelayan kecil lainnya kehilangan area tangkap mereka. Untuk bertahan hidup, ia terpaksa bergabung dengan nelayan laut dalam sebagai sawi (anak buah kapal) di perahu-perahu besar, yang biasanya melaut hingga 15 hari. Ini menjadi pilihan satu-satunya agar keluarganya tetap bisa memenuhi kebutuhan sehari-hari [5].

Industri penambangan pasir laut juga memiliki potensi besar untuk memberikan dampak positif di berbagai bidang, termasuk ekonomi, sosial, lingkungan, dan ilmu pengetahuan serta teknologi. Secara ekonomi, pengelolaan penambangan pasir laut yang baik dapat memberikan kontribusi signifikan melalui penjualan komoditas, pemasukan pajak (seperti Pendapatan Asli Daerah atau PAD), dan penyerapan tenaga kerja, terutama bagi penduduk lokal [6].

Penelitian sebelumnya menggunakan algoritma Multinomial Naïve Bayes dengan metode 10-fold cross validation menghasilkan rata-rata akurasi sebesar 88,20%, dengan akurasi tertinggi tercatat pada fold ke-8, yaitu mencapai 94%. Hasil ini menunjukkan bahwa Multinomial Naïve Bayes memiliki performa yang konsisten dan andal dalam mengklasifikasikan data teks, serta mampu memberikan hasil prediksi yang cukup akurat pada sebagian besar pembagian data pelatihan dan pengujian[7]. Kemudian hasil akhir dari penelitian Hidayat dkk menunjukkan Hasil akhir dengan pemrosesan Naïve Bayesmendapatkan hasil dengan accuracysesbesar 80,06%, percissionsebesar 83,11 dan recallsebesar 75,21 [8].

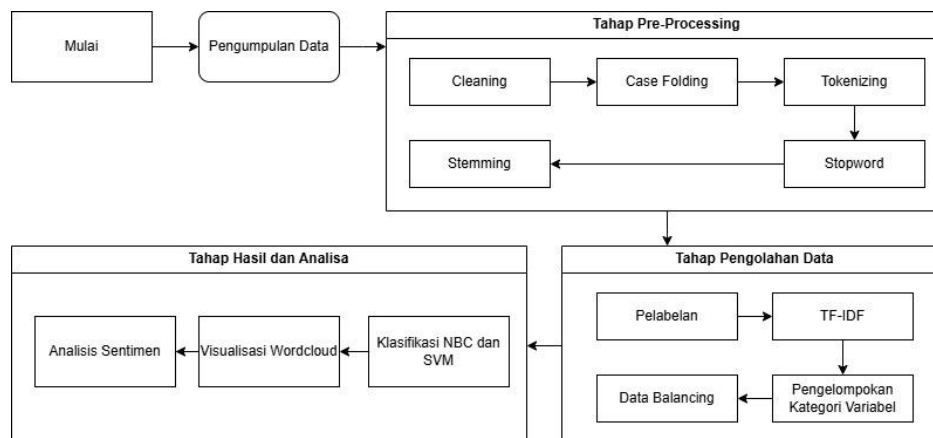
Penerapan algoritma yang dilakukan oleh Catur dkk, SVM dengan kernel LBF untuk deteksi katarak menunjukkan akurasi sebesar 95,04% [9]. Penelitian terdahulu oleh Hasibuan dkk dengan topik analisis sentimen kebijakan ekspor pasir laut pada sosial media twitter menggunakan algoritma support vector machine, pengujian terhadap kernel SVM menunjukkan kinerja yang optimal, khususnya pada kernel polynomial yang mencapai akurasi 80,94%, precision 80,91%, recall 80,94%, dan f-measure 80,75% setelah dilakukan penanganan terhadap ketidakseimbangan kelas[10]. Namun, belum ada penelitian sebelumnya yang secara langsung membandingkan dua algoritma, yaitu Naïve Bayes dan Support Vector Machine (SVM), dalam menganalisis isu kebijakan ekspor pasir laut, serta menilai performa keduanya dengan metode balancing data SMOTE. Penelitian ini mencakup jumlah data yang lebih besar dan rentang waktu yang lebih luas, yaitu dari Mei 2023 hingga Oktober 2024, sehingga mampu memberikan gambaran yang lebih representatif terhadap opini publik. Selain itu, penelitian ini juga memperkaya analisis dengan mengelompokkan data ke dalam empat kategori variabel, yaitu ekonomi, lingkungan, sosial, dan kebijakan geologis, guna mengetahui fokus utama respons publik terhadap kebijakan tersebut. Fokus utama penelitian ini adalah menganalisis persepsi masyarakat terhadap kebijakan ekspor pasir laut tahun 2023 berdasarkan ulasan yang diunggah di media sosial Twitter.

Berdasarkan uraian tersebut, peneliti ingin menggunakan algoritma *Naive Bayes* dan *Support Vector Machine* mengenai Sentimen Masyarakat Terhadap Kebijakan Ekspor Pasir Laut Berdasarkan Ulasan Twitter, agar hasil analisis ini dapat memberikan gambaran performa dari kedua algoritma dan persepsi masyarakat dan menjadi masukan penting bagi pembuat kebijakan dalam menyempurnakan kebijakan ekspor pasir laut.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Untuk mendukung pencapaian tujuan penelitian, prosedur metodologi yang digunakan digambarkan secara sistematis, mulai dari tahap pengumpulan data sampai analisis sentiment yang ditampilkan pada Gambar 1.



Gambar 1. Metode Penelitian

Gambar alur penelitian menunjukkan proses analisis sentimen dimulai dari pengumpulan data Twitter, dilanjutkan dengan pre-processing seperti cleaning, case folding, tokenizing, stopword removal, dan stemming. Setelah itu, data diproses melalui pelabelan, pembobotan kata dengan TF-IDF, pengelompokan kategorikal variabel dan data balancing menggunakan SMOTE. Data yang telah siap kemudian diklasifikasikan menggunakan algoritma Naïve Bayes dan SVM, hasilnya divisualisasikan dalam bentuk wordcloud, dan dianalisis untuk mengetahui sentimen masyarakat terhadap kebijakan ekspor pasir laut.

## 2.1 Tahap Pengumpulan Data

Pengumpulan data ulasan dilakukan melalui teknik scraping menggunakan bahasa pemrograman Python, dengan data yang diambil dari ulasan pengguna Twitter terhadap kebijakan ekspor pasir laut.

## 2.2 Tahap Pre-Processing

Setelah data berhasil dikumpulkan, tahap selanjutnya adalah proses pra-pemrosesan (pre-processing), yang diawali dengan pembersihan (cleaning) untuk menghapus emoji, simbol, serta tanda baca yang tidak relevan dalam teks. Selanjutnya dilakukan case folding, yaitu mengubah seluruh huruf dalam teks menjadi huruf kecil (lowercase) guna memastikan konsistensi. Proses tokenisasi (tokenizing) kemudian diterapkan untuk memecah kalimat menjadi kata-kata terpisah. Setelah itu, tahap stopword removal dilakukan dengan menghilangkan kata-kata yang dianggap tidak memiliki makna penting dalam analisis. Terakhir, dilakukan stemming untuk menghapus imbuhan sehingga, untuk memastikan bahwa data yang akan digunakan adalah representatif dan bebas dari gangguan yang dapat menurunkan performa model analitik[11].

## 2.3 Tahap Pengolahan Data

### a. Pelabelan

Proses pelabelan data ke dalam kategori positif, negatif, dan netral dilakukan secara manual oleh pakar untuk mendapatkan hasil yang lebih akurat, karena kualitas dan kuantitas label sangat memengaruhi performa model[12]

### b. Pembobotan TF-IDF

Tahapan ini memberikan bobot pada setiap istilah yang terdapat dalam dokumen dengan mencari banyaknya kemunculan term pada dokumen [13]. Proses pembobotan dilakukan dengan memanfaatkan modul scikit-learn (sklearn) menggunakan bahasa pemrograman Python.

### c. Pengelompokan Kategori Variabel

Evaluasi didasarkan pada variabel kategori, yaitu ekonomi, lingkungan, sosial, dan kebijakan geologis. Setelah proses pelabelan sentimen selesai, perbandingan dilakukan di antara keempat kategori tersebut guna mengidentifikasi perbedaan kecenderungan opini publik.

### d. Data Balancing

masalah ketidakseimbangan kelas adalah tantangan umum dalam klasifikasi data nyata[14]Proses menyeimbangkan jumlah data antar kelas dalam dataset, terutama saat ada ketimpangan (misalnya kelas positif jauh lebih banyak dari negatif. Tujuannya agar model tidak bias dan bisa mengenali semua kelas secara adil. Teknik yang digunakan pada penelitian ini adalah SMOTE yang menghasilkan sampel baru dengan interpolasi antara data minoritas yang ada. Keunggulan utama dari SMOTE terletak pada kemampuannya mempertahankan informasi yang ada [15]. penggunaan SMOTE dapat meningkatkan akurasi klasifikasi[16].

## 2.4 Tahap Hasil dan Analisa

### a. Klasifikasi NBC dan SVM

Pada tahap klasifikasi, dilakukan evaluasi kinerja model dari dua algoritma, yakni Naïve Bayes Classifier dan Support Vector Machine. Naive Bayes merupakan metode klasifikasi dalam data mining yang didasarkan pada pendekatan statistik dan probabilitas [17]. SVM merupakan metode machine learning yang bertujuan menemukan hyperplane optimal untuk memisahkan dua kelas dalam ruang input [18]. Langkah ini bertujuan mengidentifikasi kategori atau kelas dari suatu objek berdasarkan karakteristik atau fitur-fiturnya.

### b. Visualisasi Wordcloud

*WordCloud* merupakan bentuk visualisasi data yang menampilkan sekumpulan kata-kata penting berdasarkan tingkat frekuensi kemunculannya dalam dataset. Visualisasi merupakan metode yang efektif untuk menganalisis dan menyajikan kata kunci dengan cara yang lebih intuitif dan mudah dipahami oleh pengguna [19]. Visualisasi wordcloud dapat memperkaya metode pengajaran dan meningkatkan efektivitas pembelajaran[20] Supriyanto dkk mengetahui frekuensi kata yang paling banyak muncul pada program merdeka belajar [21]

### c. Analisis Sentimen

Analisis sentimen merupakan proses komputasi yang bertujuan menilai opini dan emosi seseorang terhadap suatu entitas, peristiwa, maupun atribut yang berkaitan [22]. Analisis sentimen dimanfaatkan untuk menganalisis beragam jenis teks, termasuk ulasan produk, unggahan di media sosial, artikel berita, dan sejenisnya [23], sehingga diperoleh gambaran yang lebih komprehensif mengenai persepsi masyarakat terhadap kebijakan ekspor pasir laut.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Pengumpulan Data

Data dalam penelitian ini diperoleh dari komentar pengguna terkait kebijakan ekspor pasir laut yang tersebar di platform twitter. Pengumpulan data dilakukan secara otomatis melalui metode crawling di platform tersebut dengan bantuan bahasa pemrograman Python. Tweet yang dikumpulkan menggunakan tagar #EksporPasirLaut. Secara keseluruhan, jumlah data yang berhasil dikumpulkan mencapai 2765 ulasan. Ulasan yang dikumpulkan berasal dari tanggal 16 Mei 2023 sampai tanggal 31 Oktober 2024.

#### 3.2 Proses Pre-Processing

Setelah data dikumpulkan, dilakukan pre-processing yang mencakup pembersihan karakter tidak relevan (*cleaning*), pengubahan teks ke huruf kecil (*case folding*), pemecahan teks menjadi kata-kata (*tokenizing*), penghapusan kata kurang bermakna (*stopword removal*), dan pengembalian kata ke bentuk dasarnya (*stemming*). Hasil dari pre-processing dapat dilihat pada Tabel 1.

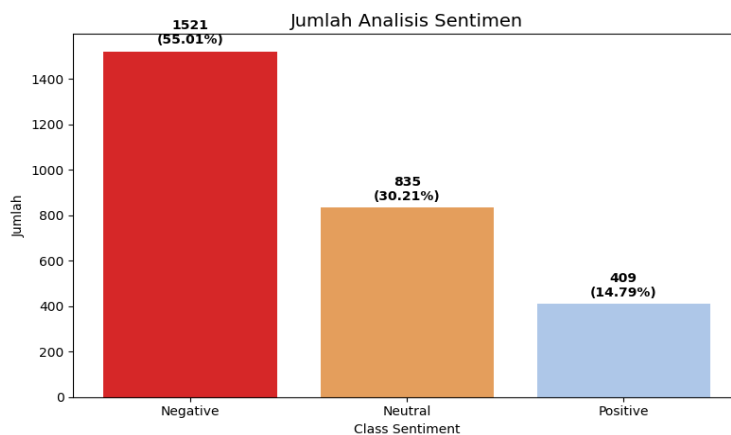
Tabel 1. Hasil Pre-Processing Data Twitter

Proses	Hasil
Data awal	Alhamdulillah akhirnya berakhir juga masa si Mulyono sebagai presiden. semoga presiden Prabowo segera mencabut semua keputusan kontroversial si Mulyono paling pertama tentang ekspor pasir laut itu.
Cleaning	Alhamdulillah akhirnya berakhir juga masa si Mulyono sebagai presiden semoga presiden Prabowo segera mencabut semua keputusan kontroversial si Mulyono paling pertama tentang ekspor pasir laut itu
Case folding	alhamdulillah akhirnya berakhir juga masa si mulyono sebagai presiden semoga presiden prabowo segera mencabut semua keputusan kontroversial si mulyono paling pertama tentang ekspor pasir laut itu
Tokeizing	['alhamdulillah', 'akhirnya', 'berakhir', 'juga', 'masa', 'si', 'mulyono', 'sebagai', 'presiden', 'semoga', 'presiden', 'prabowo', 'segera', 'mencabut', 'semua', 'keputusan', 'kontroversial', 'si', 'mulyono', 'paling', 'pertama', 'tentang', 'ekspor', 'pasir', 'laut', 'itu']
Stopword	['alhamdulillah', 'mulyono', 'presiden', 'semoga', 'presiden', 'prabowo', 'mencabut', 'keputusan', 'kontroversial', 'mulyono', 'ekspor', 'pasir', 'laut']
Stemming	alhamdulillah mulyono presiden moga presiden prabowo cabut putus kontroversial mulyono ekspor pasir laut

Proses yang dilakukan pada Tabel 1 dimulai dari data awal hingga proses stemming memiliki tujuan untuk menyederhanakan teks, hanya menyisakan kata penting untuk analisis sentimen yang akurat.

#### 3.3 Proses Pelabelan

Dari hasil pelabelan manual oleh pakar terhadap 2765 data, diperoleh bahwa komentar pada platform Twitter terdiri atas 1521 sentimen negatif, 835 sentimen positif, dan 409 sentimen netral. Persentase keberhasilan pelabelan untuk masing-masing kategori dapat dilihat secara lebih rinci pada Gambar 2.



Gambar 2. Persentasi Hasil Pelabelan Data Twitter

Grafik pada Gambar 2 menunjukkan dominasi sentimen negatif (1521 kasus, 55.01%), diikuti netral (835 kasus, 30.21%), dan paling sedikit positif (409 kasus, 14.79%). Secara keseluruhan, data menunjukkan bahwa sentimen yang dianalisis cenderung bersifat negatif.

### 3.4 TF-IDF

Tahap TF-IDF digunakan untuk menghitung bobot dari setiap kata dalam data. Hasil dari proses ini ditampilkan pada Tabel 2.

**Tabel 2.** Hasil TF-IDF

<i>Term</i>	<i>Term</i>	<i>Term</i>
Laut	0.0663084	183
Pasir	0.0602880	166
Ekspor	0.0549045	151
Sedimentasi	0.0270291	74
Jual	0.0187818	51
Jokowi	0.0185373	51
Buka	0.0148003	40
Air	0.0141343	39
Indonesia	0.0132662	36

Tabel 2 menampilkan kata-kata paling relevan dari sebuah teks, diurutkan berdasarkan nilai TF-IDF mereka. Kolom pertama adalah kata (*Term*), kolom kedua adalah nilai TF-IDF yang menunjukkan seberapa penting kata tersebut, dan kolom ketiga adalah frekuensi kemunculan kata. Semakin tinggi nilai TF-IDF, semakin unik dan signifikan kata tersebut dalam konteks data, seperti "Laut" yang memiliki nilai tertinggi.

### 3.5 Pengelompokan Kategori Variabel

Pengelompokan data dikategorikan pada 4 aspek, yaitu ekonomi, lingkungan, sosial, dan kebijakan geologis. Setelah data sentimen diberi label, hasil pelabelan ini kemudian dibandingkan antara keempat aspek tersebut untuk mengidentifikasi perbedaan dalam kecenderungan opini publik. Hasil pengelompokan berdasarkan kategori ini disajikan dalam Tabel 3.

**Tabel 3.** Tabel Kategori Variabel

Kategori Variabel	Negatif	Netral	Positif
Ekonomi	1011	384	231
Lingkungan	409	358	135
Sosial	53	42	18
Kebijakan Geologis	48	53	21

Tabel 3 menunjukkan distribusi sentimen berdasarkan empat kategori variabel. Kategori Ekonomi didominasi sentimen negatif (1011), lingkungan (409), begitu pula Sosial (53). Pada Kebijakan Geologis didominasi sentiment netral (58).

### 3.6 Data Balancing

Sebelum data dianalisis menggunakan algoritma NBC dan SVM, terlebih dahulu dilakukan proses *Random Oversampling* guna mengatasi ketidakseimbangan kelas pada dataset. Hasil dari proses tersebut disajikan pada Tabel 4 berikut.

**Tabel 4.** Hasil Penerapan SMOTE

Kategori Variabel	Data Balancing	Negatif	Netral	Positif
Ekonomi	Tanpa SMOTE	1011	384	231
	Dengan SMOTE	1011	1011	1011
Lingkungan	Tanpa SMOTE	409	358	135
	Dengan SMOTE	409	409	409
Sosial	Tanpa SMOTE	53	42	18
	Dengan SMOTE	53	53	53
Kebijakan Geologis	Tanpa SMOTE	48	53	21
	Dengan SMOTE	53	53	53

Dari Tabel 4 terlihat bahwa jumlah data pada kategori positif dan netral jauh lebih sedikit dibandingkan dengan kategori negatif. Namun, setelah diterapkan metode SMOTE, jumlah data pada ketiga kategori positif, negatif, dan netral berhasil diseimbangkan, masing-masing mengikuti data tertinggi.

### 3.7 Klasifikasi Algoritma NBC dan SVM

#### 3.7.1 Pembagian Data NBC

Setelah melalui tahap pra-pemrosesan, pelabelan, dan perhitungan bobot TF-IDF, langkah selanjutnya adalah melakukan klasifikasi dengan menggunakan model NBC. Sebelum proses klasifikasi dimulai, dataset dibagi menjadi dua bagian, yaitu data latih dan data uji. Pada penelitian ini, pembagian dilakukan dengan rasio 80:20, di mana 80%

data digunakan untuk melatih model, sementara 20% sisanya digunakan untuk menguji dan mengevaluasi kinerja model.

### 3.7.2 Hasil Algoritma Naïve Bayes Classifier

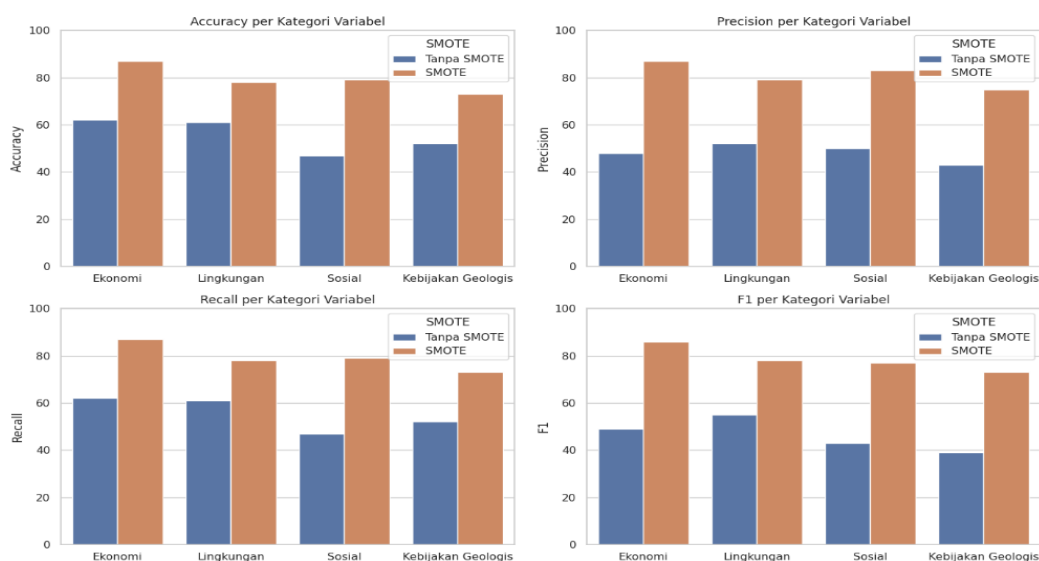
Model Naive Bayes tidak memerlukan parameter khusus dalam proses pengujian, sehingga pengujian dilakukan secara acak oleh sistem sebanyak satu kali. Proses ini dijalankan menggunakan Google Colaboratory dengan bahasa pemrograman Python. Rincian hasil pengujian tersebut disajikan pada Tabel 5.

**Tabel 5.** Penerapan SMOTE NBC

Kategori Variabel	Performa	Tanpa SMOTE	SMOTE
Ekonomi	<i>Accuracy</i>	62%	87%
	<i>Precision</i>	48%	87%
	<i>Recall</i>	62%	87%
	<i>F1-Score</i>	49%	86%
Lingkungan	<i>Accuracy</i>	61%	78%
	<i>Precision</i>	52%	79%
	<i>Recall</i>	61%	78%
	<i>F1-Score</i>	55%	78%
Sosial	<i>Accuracy</i>	47%	79%
	<i>Precision</i>	50%	83%
	<i>Recall</i>	47%	79%
	<i>F1-Score</i>	43%	77%
Kebijakan Geologis	<i>Accuracy</i>	52%	73%
	<i>Precision</i>	43%	75%
	<i>Recall</i>	52%	73%
	<i>F1-Score</i>	39%	73%

Berdasarkan Tabel 5, penggunaan SMOTE secara signifikan meningkatkan performa model di semua kategori dan metrik evaluasi (*Accuracy*, *Precision*, *Recall*, dan *F1-Score*). Peningkatan paling mencolok terlihat pada kategori Ekonomi, di mana akurasi meningkat dari 62% menjadi 87%, serta pada kategori Kebijakan Geologis, di mana *F1-Score* naik dari 39% menjadi 73%. Ini menunjukkan bahwa SMOTE sangat efektif dalam menangani ketidakseimbangan data. Adapun tampilan visualisasi dari performa NBC tanpa SMOTE dan dengan penerapan SMOTE pada Gambar 3.

Perbandingan Performa Algoritma (Tanpa SMOTE vs SMOTE)



**Gambar 3.** Hasil Pengujian NBC

Grafik pada Gambar 3 Grafik menunjukkan bahwa penerapan SMOTE secara konsisten meningkatkan performa algoritma di semua kategori dan metrik (*Accuracy*, *Precision*, *Recall*, dan *F1-Score*). Peningkatan paling signifikan terlihat pada kategori Ekonomi dan Kebijakan Geologis, di mana semua metrik mengalami lonjakan tajam setelah menggunakan SMOTE. Hal ini membuktikan bahwa SMOTE efektif dalam mengatasi ketidakseimbangan data dan meningkatkan kualitas prediksi model.

### 3.7.3 Pembagian Data SVM

Setelah melalui tahap pra-pemrosesan, pelabelan, dan perhitungan bobot TF-IDF, langkah selanjutnya adalah melakukan klasifikasi dengan menggunakan model SVM. Sebelum proses klasifikasi dimulai, dataset dibagi menjadi dua bagian, yaitu data latih dan data uji. Pada penelitian ini, pembagian dilakukan dengan rasio 80:20, di mana 90% data digunakan untuk melatih model, sementara 10% sisanya digunakan untuk menguji dan mengevaluasi kinerja model.

### 3.7.4 Hasil Algoritma Support Vector Machine

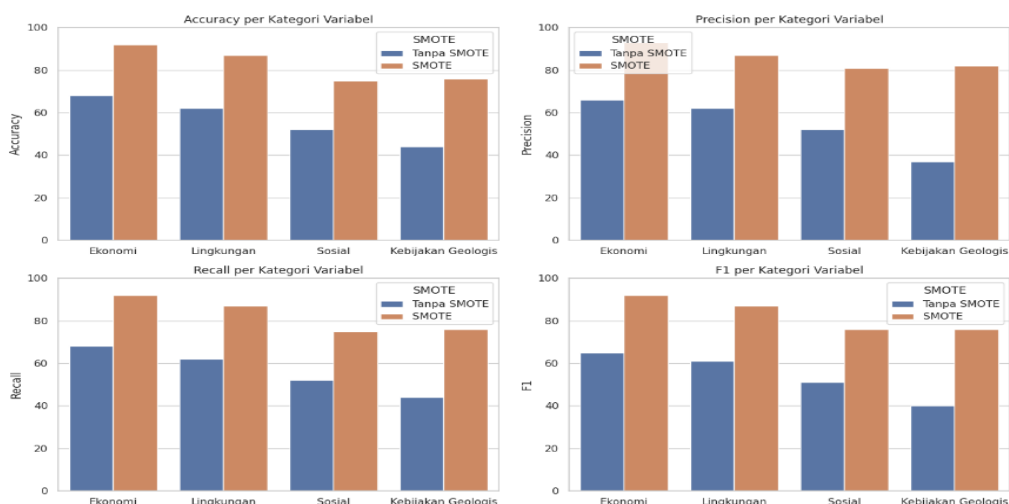
Pada algoritma Support Vector Machine (SVM), evaluasi model dilakukan berdasarkan hasil prediksi tanpa melakukan eksplorasi terhadap parameter seperti kernel, regulasi (C), maupun gamma. Fokus pengujian hanya ditujukan untuk menilai performa model yang telah tersedia. Proses pengujian menggunakan parameter default dari library yang digunakan, dan pembagian data untuk pelatihan serta pengujian dilakukan secara acak oleh sistem. Rincian lengkap dari hasil pengujian ini dapat dilihat pada Tabel 6.

**Tabel 6.** Penerapan SMOTE SVM

Kategori Variabel	Performa	Tanpa SMOTE	SMOTE
Ekonomi	<i>Accuracy</i>	68%	92%
	<i>Precision</i>	66%	93%
	<i>Recall</i>	68%	92%
	<i>F1-Score</i>	65%	92%
Lingkungan	<i>Accuracy</i>	62%	87%
	<i>Precision</i>	62%	87%
	<i>Recall</i>	62%	87%
	<i>F1-Score</i>	61%	87%
Sosial	<i>Accuracy</i>	52%	75%
	<i>Precision</i>	52%	81%
	<i>Recall</i>	52%	75%
	<i>F1-Score</i>	51%	76%
Kebijakan Geologis	<i>Accuracy</i>	44%	76%
	<i>Precision</i>	37%	82%
	<i>Recall</i>	44%	76%
	<i>F1-Score</i>	40%	76%

Berdasarkan Tabel 6, penggunaan SMOTE secara signifikan meningkatkan performa model pada semua kategori variabel. Peningkatan paling menonjol terjadi pada kategori Kebijakan Geologis, di mana Accuracy naik dari 44% menjadi 76% dan F1-Score dari 40% menjadi 76%. Demikian pula pada kategori Ekonomi, semua metrik meningkat drastis hingga mencapai sekitar 92–93%. Hal ini menunjukkan bahwa SMOTE sangat efektif dalam memperbaiki performa model, khususnya pada data yang sebelumnya kurang seimbang. Adapun tampilan visualisasi dari performa NBC tanpa SMOTE dan dengan penerapan SMOTE pada Gambar 4.

Perbandingan Performa Tanpa SMOTE vs SMOTE



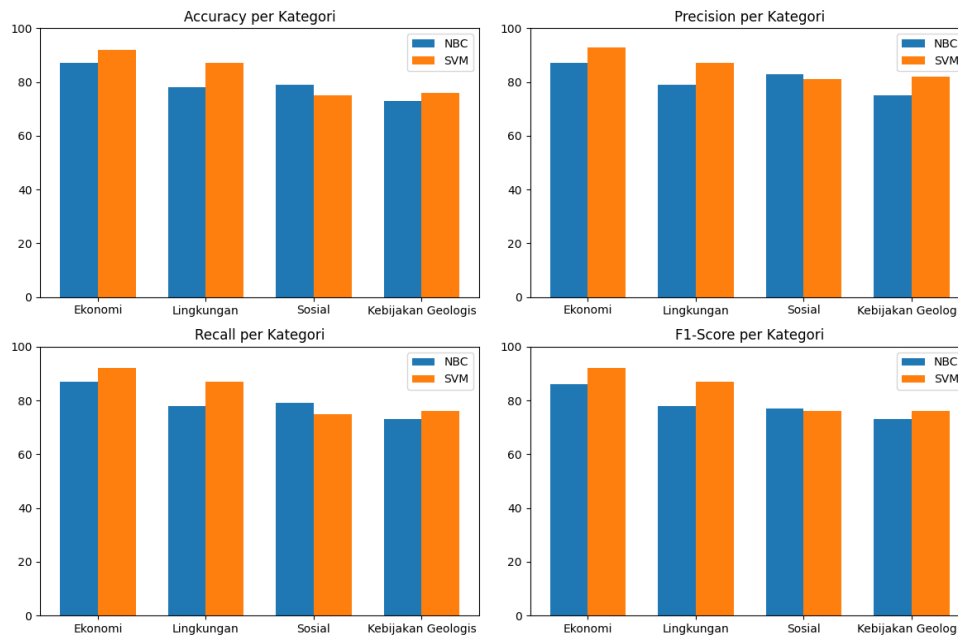
**Gambar 4.** Hasil Pengujian SVM

Grafik pada Gambar 4 Grafik menunjukkan bahwa penerapan SMOTE secara konsisten meningkatkan performa model di semua metrik dan kategori. Peningkatan paling signifikan terlihat pada kategori Kebijakan

Geologis, dengan lonjakan pada Precision dan F1-Score. Kategori Ekonomi juga mengalami peningkatan tajam, terutama pada Accuracy dan Recall yang mencapai lebih dari 90%. Secara keseluruhan, SMOTE terbukti efektif dalam mengatasi ketidakseimbangan data dan meningkatkan kualitas prediksi model.

### 3.7.5 Hasil Perbandingan Akurasi Algoritma

Perbandingan kinerja algoritma Naive Bayes Classifier (NBC) dan Support Vector Machine (SVM) berdasarkan empat metrik evaluasi utama, yaitu akurasi, precision, recall, dan F1-score, visualisasi dari perbandingan dapat dilihat pada Gambar 5.



Gambar 5. Perbandingan Performa NBC dan SVM

Grafik pada Gambar 5 menunjukkan bahwa SVM memiliki performa paling tinggi di kategori Ekonomi, dengan nilai Accuracy, Precision, Recall, dan F1-Score yang melampaui Naive Bayes (NBC) secara konsisten. Pada kategori Lingkungan dan Sosial, selisih performa antara SVM dan NBC relatif kecil dan bersaing seimbang. Sementara di kategori Kebijakan Geologis, NBC sedikit tertinggal, namun performanya tidak jauh berbeda dari SVM. Secara keseluruhan, SVM lebih unggul dalam akurasi dan konsistensi performa, tetapi NBC tetap kompetitif tergantung pada jenis kategori yang dianalisis.

### 3.8 Visualisasi Data Sentimen

Visualisasi data bertujuan untuk mengidentifikasi topik-topik yang paling sering dibicarakan oleh pengguna Twitter. Dengan menggunakan teknik ini, kata-kata yang terkandung dalam ribuan komentar dapat diungkap dengan lebih jelas dan padat, sehingga memberikan informasi yang relevan terkait kebijakan ekspor pasir laut. Visualisasi dari kata-kata yang paling sering muncul dalam ulasan Twitter tentang kebijakan ekspor pasir laut pada setiap kategori sentimen dapat dilihat pada Gambar 6, Gambar 7, dan Gambar 8.



Gambar 6. Hasil Visualisasi Sentimen Positif

Word Cloud pada Gambar 6 merupakan visualisasi Word Cloud untuk sentimen positif terhadap kebijakan ekspor pasir laut. Kata-kata yang paling dominan antara lain "untung", "jual", "manfaat", dan "nilai", yang menunjukkan bahwa sebagian masyarakat melihat kebijakan ini dari sisi ekonomi dan potensi keuntungan finansial. Kata seperti "triliun", "investasi", dan "ekonomi" juga mencerminkan harapan terhadap peningkatan pendapatan negara. Sementara itu, munculnya kata "jaga", "kelola", dan "lingkung" menandakan bahwa dukungan terhadap



ekspor pasir laut, baik dari sisi dukungan maupun penolakan, serta menyediakan landasan berbasis data yang kuat sebagai bahan evaluasi bagi para pembuat kebijakan.

Penelitian ini memberikan kontribusi berupa hasil klasifikasi dan visualisasi sentimen melalui wordcloud membantu mengungkap isu-isu yang paling banyak dibicarakan masyarakat, seperti kekhawatiran terhadap kerusakan lingkungan atau potensi manfaat ekonomi. Hal ini penting untuk memperkuat keterlibatan publik dalam pengawasan kebijakan berbasis data.

#### 4. KESIMPULAN

Penelitian ini menganalisis sentimen masyarakat terhadap kebijakan ekspor pasir laut menggunakan pendekatan pembelajaran mesin dengan algoritma Naïve Bayes (NBC) dan Support Vector Machine (SVM), serta menerapkan teknik SMOTE untuk mengatasi ketidakseimbangan data. Berdasarkan analisis terhadap 2.765 data ulasan Twitter, ditemukan bahwa sentimen negatif mendominasi (55%), diikuti sentimen netral (30%) dan positif (15%). Visualisasi Word Cloud menunjukkan bahwa sentimen negatif banyak menyoroti kekhawatiran terhadap dampak lingkungan dan kerugian ekonomi, sedangkan sentimen positif menekankan potensi manfaat dan dukungan terhadap kebijakan jika dikelola dengan baik. Hasil evaluasi performa algoritma menunjukkan bahwa SVM unggul atas Naïve Bayes (NBC) terutama pada kategori Ekonomi, dengan nilai Accuracy, Precision, Recall, dan F1-Score yang lebih tinggi secara konsisten. Pada kategori Lingkungan dan Sosial, performa kedua algoritma relatif seimbang, dengan selisih yang tidak terlalu besar. Sementara itu, pada kategori Kebijakan Geologis, performa keduanya cenderung lebih rendah dibanding kategori lain, meskipun masih menunjukkan hasil yang cukup baik. Secara umum, SVM memberikan hasil yang lebih stabil dan akurat, namun efektivitas tiap algoritma tetap bergantung pada karakteristik masing-masing kategori. Dengan demikian, dapat disimpulkan bahwa SVM dengan teknik SMOTE merupakan pendekatan yang lebih efektif dan stabil dalam mengklasifikasikan sentimen publik terhadap isu-isu kebijakan kompleks seperti ekspor pasir laut. Temuan ini diharapkan dapat memberikan kontribusi dalam pengambilan kebijakan berbasis data serta memperkaya kajian akademik di bidang analisis sentimen sosial dan pengolahan data tidak seimbang.

#### REFERENCES

- [1] D. Lestari and I. Sari, “Kebijakan Ekspor Pasir Laut Pasca Diundangkan Peraturan Pemerintah Nomor 26 Tahun 2023 Berdasarkan Prespektif Teori Sistem Hukum,” *Jurnal Hukum Samudra Keadilan*, vol. 18, no. 2, 2023, doi: <https://doi.org/10.33059/jhsk.v18i2.8322>.
- [2] S. Surlanti, A. Asrim, and R. Wardana, “Analisis Dampak Penambangan Pasir Laut Terhadap Lingkungan Dan Sosial-Ekonomi Di Desa Kamelanta Kecamatan Kapontori Kabupaten Buton,” *Jurnal Media Inovasi Teknik Sipil UNIDAYAN*, vol. 12, no. 2, pp. 59–64, Oct. 2023, doi: 10.55340/jmi.v12i2.1433.
- [3] C. Rahardjo, “Analisis Pemenuhan Ganti Kerugian Nelayan Akibat dari Kebijakan Ekspor Pasir Laut Melalui Gugatan Perdata,” *Dialogia Iuridica*, vol. 17, no. 1, pp. 082–110, Apr. 2025, doi: 10.28932/di.v17i1.11286.
- [4] D. Lestari and I. Sari, “Kebijakan Ekspor Pasir Laut Pasca Diundangkan Peraturan Pemerintah Nomor 26 Tahun 2023 Berdasarkan Prespektif Teori Sistem Hukum,” *Jurnal Hukum Samudra Keadilan*, vol. 18, no. 2, 2023, doi: <https://doi.org/10.33059/jhsk.v18i2.8322>.
- [5] D. Anggariyani, S. Sahar, and M. Sayful, “Tambang Pasir dan Dampak Sosial Ekonomi Masyarakat di Pesisir Pantai,” *SIGN Journal of Social Science*, vol. 1, no. 1, pp. 15–29, Feb. 2021, doi: 10.37276/sjss.v1i1.96.
- [6] Muammar Gomareuzzaman, “Pemanfaatan Sumber Daya Alam Pasir Laut sebagai Komoditas Tambang Masyarakat di Perairan Kabupaten Karimunjaya, Kepulauan Riau,” *Syntax Admiration*, vol. 5, p. 8, Aug. 2024, doi: <https://doi.org/10.46799/jsa.v5i8.1350>.
- [7] H. H. Mubaroroh, H. Yasin, and A. Rusgiyono, “Analisis Sentimen Data Ulasan Aplikasi Ruangguru Pada Situs Google Play Menggunakan Algoritma Naïve Bayes Classifier Dengan Normalisasi Kata Levenshtein Distance,” *Jurnal Gaussian*, vol. 11, no. 2, pp. 248–257, 2022, doi: <https://doi.org/10.14710/j.gauss.v11i2.35472>.
- [8] T. Hidayat, R. Cahyana, and I. T. Julianto, “Analisis Sentimen Layanan Sistem Informasi Akademik Mahasiswa Menggunakan Algoritma Naïve Bayes,” *Jurnal Algoritma*, vol. 21, no. 1, pp. 119–130, May 2024, doi: 10.33364/algoritma/v.21-1.1514.
- [9] S. Catur Sakti Kendari and D. Panjaitan, “Penerapan Algoritma Support Vector Machine (SVM) untuk Deteksi Penyakit Mata Katarak Yonal Supit,” *IJIDS*, vol. 2, no. 2, 2024, doi: <https://doi.org/10.30989/ijds.v2i2.1449>.
- [10] R. A. Hasibuan, D. E. Ratnawati, and R. S. Perdana, “Analisis Sentimen Kebijakan Ekspor Pasir Laut Pada Sosial Media Twitter Menggunakan Algoritma Support Vector Machine,” *Justsi*, vol. 5, no. 1, 2024. doi: <https://doi.org/10.25126/justsi.v5i1.373>.
- [11] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, “A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data,” *Frontiers Media S.A.*, Mar. 29, 2021, doi: 10.3389/fenrg.2021.652801.
- [12] Q. Huang and T. Zhao, “Data Collection and Labeling Techniques for Machine Learning,” arxiv, Jun. 2024, [Online]. Available: <http://arxiv.org/abs/2407.12793>
- [13] I. Anbi Fahrezi, and N. Azmi Verdikha, “Analisis Sentimen Twitter Atas Isu Hak Angket Menggunakan Pembobotan TF-IDF dan Algoritma SVM,” *Sci-Tech Journal*, vol. 3, no. 2, 2024, doi: <https://doi.org/10.56709/stj.v3i2.526>.
- [14] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, “Learning from class-imbalanced data: Review of methods and applications,” *Elsevier Ltd.*, May 01, 2017, doi: 10.1016/j.eswa.2016.12.035.
- [15] N. P. Y. T. Wijayanti, E. N. Kencana, and I. W. Sumarjaya, “SMOTE: Potensi Dan Kekurangannya Pada Survei,” *E-Jurnal Matematika*, vol. 10, no. 4, p. 235, Nov. 2021, doi: 10.24843/mtk.2021.v10.i04.p348.



- [16] M. Mujahid *et al.*, “Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering,” *J Big Data*, vol. 11, no. 1, Dec. 2024, doi: 10.1186/s40537-024-00943-4.
- [17] S. Busono, “Optimasi Naive Bayes Menggunakan Algoritma Genetika Sebagai Seleksi Fitur Untuk Memprediksi Performa Siswa,” *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 14, no. 1, 2020, doi: <https://doi.org/10.32815/jitika.v14i1.400>.
- [18] M. Wasil and M. Mahpuz, “Analisis Pengaruh Faktor Kemiskinan Terhadap Tingkat Kesehatan Dan Gaya Hidup Masyarakat Desa Suralaga, Lombok Timur, Menggunakan Algoritma Support Vector Machine (SVM),” *Infotek: Jurnal Informatika dan Teknologi*, vol. 4, no. 1, pp. 11–19, Jan. 2021, doi: 10.29408/jit.v4i1.2978.
- [19] M. Chandra and R. Yusuf, “Visualization Of Keywords In The 2024 Election News Using Spacy And Wordcloud,” *TEKNIMEDIA: Teknologi Informasi Dan Multimedia*, vol. 5, pp. 41–46, Jun. 2024, doi: <https://doi.org/10.46764/teknimedia.v5i1.187>.
- [20] Y. Huang, Y. Wang, and F. Ye, “A Study of the application of word cloud visualization in college english teaching,” *International Journal of Information and Education Technology*, vol. 9, no. 2, pp. 119–122, Feb. 2019, doi: 10.18178/ijiet.2019.9.2.1185.
- [21] B. F. S. Supriyanto and S. Rosalin, “Analisis Sentimen Program Merdeka Belajar dengan Text Analysis Wordcloud & Word Frequency,” *Jurnal Minfo Polgan*, vol. 12, no. 1, pp. 25–32, Mar. 2023, doi: 10.33395/jmp.v12i1.12312.
- [22] C. Steven, “The Right Sentiment Analysis Method of Indonesian Tourism in Social Media Twitter Case Study: The City of Bali,” *102 IJNMT*, vol. VII, no. 2, 2020, doi: <http://dx.doi.org/10.31937/ijnmt.v7i2.1732>.
- [23] J. K. Kim, A. Khondker, M. E. Chua, M. Rickard, and A. Lorenzo, “Sentiment analysis of U.S. News & World Report Best Children’s Hospital urology rankings: A difference in positivity between the public and academic worlds,” *J Pediatr Urol*, vol. 20, pp. S81–S85, Jan. 2024, doi: 10.1016/j.jpuro.2024.06.001.