Jurnal Informatika Terpadu

https://journal.nurulfikri.ac.id/index.php/JIT
ISSN ONLINE : 2460-8998

ANALISIS SENTIMEN KONTAMINASI RADIOAKTIF DI KAWASAN INDUSTRI CIKANDE MENGGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE*

Taufik Ramlan Alfiansyah[1], Audy Abdillah Hidayat[2], Alfarezi Hidayat Pratama[3], Agil Aqshol Mahenda[4], Muhammad Rafly[5]

[1, 2, 3, 4, 5] Teknik Informatika, Universitas Bina Sarana Informatika
Depok, Jawa Barat, Indonesia 16424
taupikramlan3590@gmail.com, audyhidayat31@gmail.com, alfarezihidayat4@gmail.com,
agilaqsholmahenda@gmail.com, muhammadrfy919@gmail.com

Abstract

The suspicion of radioactive contamination in the Cikande industrial area has prompted a strong public reaction, as shown by the many comments on TikTok. This research aims to understand how people feel about this issue and to measure how well the Support Vector Machine (SVM) algorithm classifies opinions. The data used comes from 3,160 comments collected via web scraping, then processed through several steps, including cleaning text, normalizing, separating words, deleting common words, and summarizing words, before using the TF-IDF method for representation. Comments were then labelled using a lexicon-based method, which showed that 70.79% were negative and 29.21% were positive. Modelling was carried out using SVM on training and test data in an 80:20 ratio. The results show that the model achieved an accuracy of 89% and recognized both sentiment types well. In general, negative comments expressed greater concern about health and environmental impacts, and a lack of confidence in how waste is managed, while positive comments emphasized the importance of scientific verification and official monitoring. These findings indicate the need for clearer, more consistent, and data-based communication about risks to reduce public concerns and increase public trust.

Keywords: *Cikande, Radioactive Contamination, Sentiment Analysis, Support Vector Machine, TikTok*

Abstrak

Dugaan ada kontaminasi radioaktif yang mencemari kawasan industri Cikande membuat masyarakat bereaksi keras, yang terlihat dari banyaknya komentar di TikTok. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap isu tersebut serta mengevaluasi kinerja algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan opini secara otomatis. Data yang digunakan berasal dari 3.160 komentar yang dikumpulkan dengan cara *web scraping*, lalu diproses melalui beberapa langkah seperti membersihkan teks, normalisasi, pemisahan kata, menghapus kata umum, dan merangkum kata sebelum digunakan metode TF-IDF untuk representasi. Komentar kemudian diberi label menggunakan cara berdasarkan leksikon, yang menunjukkan 70.79% komentar bersifat negatif dan 29.21% bersifat positif. Pemodelan dilakukan dengan SVM menggunakan data latih dan uji dengan perbandingan 80:20. Hasilnya menunjukkan bahwa model berhasil mendapatkan akurasi sebesar 89% dan bisa dengan baik mengenali kedua jenis sentimen. Secara umum, komentar negatif lebih banyak mengungkapkan kekhawatiran terhadap kesehatan, dampak lingkungan, dan kurangnya kepercayaan terhadap cara pengelolaan limbah, sementara komentar positif menekankan pentingnya verifikasi secara ilmiah dan pemantauan resmi. Temuan ini menunjukkan perlunya komunikasi tentang risiko yang lebih jelas, konsisten, dan berdasarkan data agar bisa mengurangi kekhawatiran masyarakat dan meningkatkan kepercayaan publik.

Kata kunci: Analisis Sentimen, Cikande, Kontaminasi Radioaktif, *Support Vector Machine*, TikTok

1. PENDAHULUAN

Perkembangan industri modern telah memberikan keuntungan bagi ekonomi, tetapi di sisi lain juga bisa menyebabkan masalah besar bagi lingkungan, salah satunya adalah bahaya dari radiasi. Daerah industri yang sangat aktif, seperti Cikande di Banten, bisa terpapar limbah radioaktif karena aktivitas pabrik, penelitian, dan pengangkutan bahan berbahaya. Masalah seperti radiasi dan

Diterima 14 November 2025, Direvisi 22 November 2025, Diterima untuk publikasi 14 Januari 2026

pencemaran ini sering membuat masyarakat khawatir, terutama ketika informasi resmi dianggap tidak jelas. Dalam hal ini, media sosial seperti TikTok menjadi tempat bagi masyarakat untuk berbagi pendapat, sentimen, dan pandangan mereka tentang isu lingkungan dan kesehatan. Menganalisis komentar-komentar ini menjadi cara baru untuk lebih cepat dan lebih baik memahami pandangan publik.

Analisis sentimen adalah cara menggunakan komputer untuk mengetahui apa pendapat orang tentang masalah atau hal tertentu. Dalam penelitian yang telah dilakukan sebelumnya, analisis sentimen sering dipakai untuk melihat bagaimana pandangan masyarakat tentang kebijakan dari pemerintah dan layanan publik [1].

Pendekatan analisis sentimen juga diterapkan untuk menilai sistem informasi pendidikan. Seperti pada situs SISKA, memakai algoritma *Naive Bayes* yang diperbaiki dengan metode *Particle Swarm Optimization* (PSO). Hasil dari penelitian ini menunjukkan bahwa gabungan tersebut dapat membuat ketepatan dalam mengklasifikasikan pendapat pengguna tentang kinerja sistem mencapai 90%, membuktikan bahwa teknik pembobotan TF-IDF dan pengolahan awal data teks efektif [2].

Analisis sentimen di media sosial sangat berguna untuk mengetahui bagaimana reaksi orang-orang terhadap masalah sosial dan keputusan pemerintah. Penelitian ini menunjukkan bahwa banyak orang lebih suka menggunakan media digital untuk mengungkapkan pendapat mereka dengan cepat, yang mencerminkan bagaimana mereka merasa dan seberapa percaya mereka terhadap lembaga yang bersangkutan [3].

Sampah radioaktif memiliki zat berbahaya seperti *stronsium* dan *cesium*. Zat-zat ini bisa menyebabkan ionisasi yang merugikan pada tanah dan air, serta dapat menyebabkan kanker [4].

Pemantauan terbaru di salah satu daerah aliran sungai di kota menunjukkan bahwa *Cesium*-137, *Cobalt*-60, dan *Iodine*-131 terdeteksi dalam jumlah yang sangat kecil dan lebih rendah dari standar yang ditetapkan. Contohnya, *Cesium*-137 mudah menyebar dan *Cobalt*-60/*Iodine*-131 biasanya digunakan dalam layanan kesehatan. Penjelasan lebih lanjut tentang radiasi menjelaskan sifat-sifat utama (seperti energi, waktu paruh, dan bahaya) serta dampak kesehatan jika terpapar (misalnya, risiko kanker akibat paparan *Cesium*-137) [5].

Sebuah studi [6] menggunakan analisis sentimen untuk melihat bagaimana orang Indonesia di Twitter merespons munculnya aplikasi Threads. Penelitian ini memberikan pandangan umum tentang reaksi publik terhadap aplikasi yang baru saja diluncurkan, dan hasilnya menunjukkan bahwa sentimen negatif lebih banyak muncul (71,2%) dibandingkan sentimen positif (28,8%).

Algoritma Support Vector Machine (SVM) adalah salah satu algoritma yang bisa dipakai untuk meramalkan harga saham, sebab pergerakan harga saham umumnya tidak mengikuti pola linier. Metode ini memiliki kelebihan karena tidak mengalami masalah *overfitting* dan bertujuan untuk menemukan *hyperplane* yang terbaik untuk memisahkan dua kelompok data [7]. Dalam konteks klasifikasi teks, SVM memiliki keunggulan signifikan karena kemampuannya menangani dimensi fitur yang sangat tinggi *(high-dimensional data)* yang dihasilkan dari representasi kata seperti TF-IDF. Selain itu, SVM terbukti efektif dalam memisahkan kelas yang tidak linier melalui penggunaan fungsi *kernel*, sehingga sangat andal dalam mengenali pola sentimen yang kompleks dalam bahasa alami meskipun dengan jumlah *dataset* yang terbatas.

Penelitian yang berbeda [8] yang memeriksa pendapat tentang aplikasi TikTok di *Google Playstore*, menyatakan bahwa analisis pendapat bisa digunakan dengan baik untuk melihat cara pandang pengguna. Hasil dari analisis ini bisa digunakan sebagai dasar untuk membuat keputusan yang penting dan menjadi acuan untuk meningkatkan kualitas aplikasi berdasarkan saran yang ditemukan dari pendapat pengguna.

Berdasarkan kajian dari berbagai riset sebelumnya, bisa dinyatakan bahwa studi tentang pencemaran radioaktif di daerah pabrik masih jarang dihubungkan dengan analisis pandangan masyarakat di sosial media. Penelitian ini bertujuan untuk mengisi celah literatur tersebut dengan menganalisis sentimen publik terhadap dugaan pencemaran radioaktif di kawasan industri Cikande menggunakan algoritma *Support Vector Machine* (SVM). Selain itu, penelitian ini secara khusus bertujuan untuk mengevaluasi efektivitas algoritma SVM dalam mengklasifikasikan opini dari media sosial TikTok. Melalui pendekatan ini, diharapkan hasil penelitian dapat memberikan kontribusi ilmiah mengenai persepsi risiko lingkungan di era digital serta sumbangan praktis bagi perbaikan strategi komunikasi risiko dan transparansi informasi publik di Indonesia.

2. METODE PENELITIAN

Penelitian ini mengadopsi pendekatan deskriptif kualitatif dalam usaha memahami cara pandang warganet mengenai isu dugaan pencemaran radioaktif dengan menganalisis konteks pembicaraan di kolom komentar TikTok. Dalam kerangka ini, data komentar dianalisis dengan seksama, meliputi intonasi, alasan, dan pola reaksi untuk memetakan kecenderungan yang bersifat positif dan negatif serta mengartikan argumen yang muncul dalam diskusi. Sebagai pendukung, sebuah penelitian [9] juga mengklasifikasikan studi percakapan warganet sebagai bentuk penelitian kualitatif deskriptif. Penelitian tersebut menjelaskan pendekatannya sebagai "penelitian kualitatif deskriptif yang menggunakan pendekatan netnografi".

2.1 Metode Pengumpulan Data, Instrumen Penelitian, dan Metode Pengujian

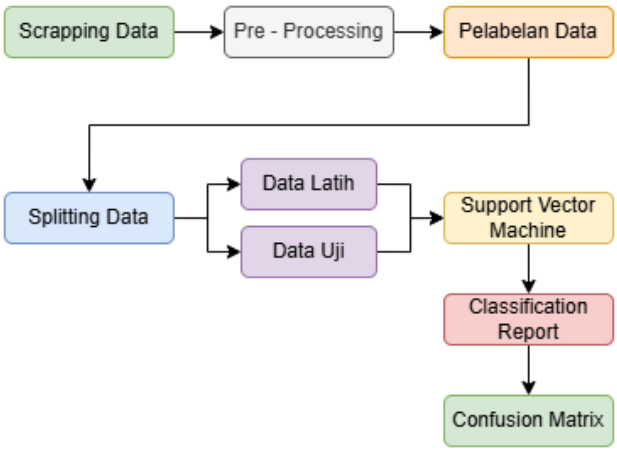
Data penelitian ini dikumpulkan dari kolom komentar di TikTok menggunakan metode pengambilan data berbasis *Python* selama periode 1 Oktober 2025 hingga 8 Oktober 2025. Kumpulan awal terdiri dari 3.184 komentar, kemudian disaring dengan menghapus duplikasi dan *spam*/penyebutan massal, memastikan relevansi topik, serta merapikan format hingga tersisalah 3.160 entri yang dapat dianalisis. Proses pengelolaan data mematuhi prinsip etika penelitian, identitas individu tidak diungkapkan dan laporan disajikan dalam bentuk agregat. Pemilihan metode pengambilan data sejalan dengan penjelasan bahwa pengambilan data dari *web* adalah tindakan mengambil data atau informasi dari situs *web* dan diterapkan ketika data tidak tersedia melalui RSS atau API [10].

Penelitian ini memanfaatkan perangkat lunak *Python* versi 3.11.0 yang dilengkapi dengan pustaka *pandas*, *numpy*, dan *scikit-learn* untuk melakukan vektorisasi dan pemodelan, serta Sastrawi dalam proses *stemming*. *Dataset* yang digunakan terdiri dari 3.160 entri. Pada tahap kedua, yang dikenal sebagai pra-proses teks, meliputi langkah-langkah seperti tokenisasi, penghilangan *stopwords*, dan penyaringan *token* berdasarkan panjangnya sebelum melanjutkan ke proses pembobotan TF-IDF.

Setelah proses *pre-processing*, teks diubah menjadi vektor TF-IDF sebagai bentuk pengukuran. Selanjutnya, data dibagi menjadi set pelatihan dan pengujian untuk mempertahankan proporsi masing-masing kelas. Model klasifikasi dilatih menggunakan data pelatihan dan kemudian diuji. Kinerja dinilai berdasarkan matriks kebingungan, dengan menggunakan metrik akurasi, presisi, dan *recall* yang diperoleh dari nilai TP, FP, TN, dan FN di dalam matriks kebingungan, yang mencerminkan ketepatan keseluruhan serta kemampuan model dalam mengenali setiap kelas. Struktur ini sejalan dengan referensi yang menempatkan TF-IDF dalam tahap *pre-processing* dan menjelaskan evaluasi yang mengandalkan akurasi, presisi, dan *recall* berdasarkan *confusion matrix* [11].

2.2 Tahapan Penelitian

Pada tahap penelitian ini diorganisir secara berurutan agar proses kerjanya bisa diulang dengan jelas. Singkatnya, penelitian dimulai dengan pengambilan data, diikuti oleh proses awal, penandaan opini, pembagian data, penerapan model (*Support Vector Machine*), dan evaluasi model.



Gambar 1. Tahapan Penelitian

Urutan penelitian dari mengumpulkan data sampai menilai model SVM ditunjukkan pada Gambar 1.

2.3 *Scraping Data*

Scraping data adalah tindakan menarik informasi atau konten dari halaman *web*, dan dilakukan ketika informasi yang diinginkan tidak dapat diakses melalui RSS atau API. Metode ini digunakan sebagai cara untuk mendapatkan data yang berbasis situs agar mendapatkan informasi yang terorganisasi dengan baik [10].

2.4 *Pre-Processing*

Pra-pemrosesan teks merupakan langkah penting sebelum melakukan penambangan data atau *machine learning* untuk membersihkan serta menstandarisasi korpus agar dapat diproses. Di fase ini (pembersihan data, normalisasi kata, penghapusan *stopwords*, serta pelaksanaan *stemming*/normalisasi) memiliki dampak langsung pada ketepatan model analisis sentimen. Studi dalam jurnal yang serupa mengulas berbagai metode pra-pemrosesan (dasar tanpa pengolahan, penghapusan kata-kata umum, dan pengurangan bentuk kata) dan menemukan adanya variasi dalam tingkat akurasi antara metode tersebut, menunjukkan bahwa perancangan saluran pra-pemrosesan perlu dianalisis secara empiris [12].

a) *Cleaning*

Pembersihan adalah titik awal dalam pra-proses teks yang ditujukan untuk meminimalisir gangguan dan merapikan korpus agar siap untuk dianalisis. Dalam tahapan ini, bagian-bagian yang tidak memberikan arti sentimen, seperti tautan *web*, tagar, simbol, nama pengguna, angka, emoji, dan juga tanda baca khusus, dibuang, untuk membuat representasi kata menjadi lebih jernih dan seragam. Sebuah penelitian [13] juga menjelaskan bahwa kata-kata yang dibersihkan selama proses pembersihan, seperti *Hastag* (#), *URL* atau tautan, tanda baca, simbol, dan nama pengguna.

b) *Case Folding*

Case folding merupakan langkah awal dalam proses yang bertujuan untuk menstandarkan penulisan karakter dengan mengubah seluruh konten menjadi huruf kecil. Tujuan dari langkah ini adalah untuk mengurangi perbedaan yang tidak memberikan arti sehingga daftar istilah menjadi lebih sederhana. Sebuah studi [14] juga menunjukkan bahwa salah satu langkah penting dalam mempersiapkan teks adalah *case folding*, yaitu "mengubah seluruh tulisan menjadi huruf kecil untuk memastikan data tetap konsisten".

c) Normalisasi Kata

Normalisasi kata merupakan tahap yang mengatur bentuk tidak standar seperti kata *slang*, istilah ganti atau penulisan yang tidak sesuai dengan kaidah menjadi versi yang lebih baku atau umum sebelum proses pemodelan. Ini bertujuan untuk meminimalkan ketidakteraturan yang tidak perlu, mengurangi kata yang tidak dikenal, serta menciptakan representasi fitur yang lebih konsisten sehingga meningkatkan akurasi pada klasifikasi sentimen. Istilah *slang* sering kali memiliki arti yang tidak jelas atau dapat ditafsirkan dengan cara berbeda, yang bisa bikin pemrosesan bahasa alami jadi salah dan menghasilkan hasil yang tidak tepat dalam tugas seperti mengklasifikasikan teks atau menganalisis sentimen. Maka dari itu, sebuah penelitian [15] menyoroti betapa pentingnya untuk menormalkan istilah slang tersebut sebelum teks dimasukkan ke dalam proses Pengolahan Bahasa Alami.

d) *Tokenizing*

Tokenisasi merupakan langkah awal dalam pengolahan yang membagi teks asli menjadi unit-unit yang memiliki arti (*token*) seperti kata-kata atau kelompok kata. Kumpulan *token* ini berfungsi sebagai data *input* untuk tahap berikutnya, sembari membantu mengurangi kompleksitas dalam representasi teks. Tokenisasi merupakan langkah dasar dalam hampir setiap aplikasi Pemrosesan Bahasa Alami. Cara umum dalam tokenisasi kata adalah dengan membagi teks menjadi kata-kata, biasanya memakai spasi sebagai batasnya [16].

e) *Stopword Removal*

Stopword Removal adalah langkah awal dalam pemrosesan yang menyingkirkan kata-kata yang sangat umum atau kurang penting agar kumpulan data lebih ringkas, fitur lebih berguna, dan penilaian tidak terpengaruh oleh kata-kata yang muncul terlalu sering tapi tidak membantu. Sebuah studi [17] menjelaskan bahwa penghapusan kata umum adalah langkah dalam mengolah teks yang bertujuan untuk menyingkirkan kata-kata yang dianggap tidak penting atau tidak memberikan kontribusi berarti dalam analisis. Kata-kata ini biasanya adalah kata-kata sehari-hari (seperti "dan," "atau," "di") yang tidak membawa informasi khusus dan sering muncul berulang kali di berbagai dokumen.

f) *Stemming*

Stemming adalah langkah awal yang membantu mengubah kata-kata dengan imbuhan seperti awalan, sisipan, dan akhiran kembali ke bentuk aslinya. Ini berguna supaya variasi kata tidak menambah fitur yang tidak penting. Dengan menggabungkan bentuk kata-kata, cara kita menyatakan fitur seperti TF-IDF menjadi lebih jelas dan teratur. *Stemming* juga didefinisikan sebagai “proses awal dalam analisis sentimen yang menemukan dan menghapus imbuhan” [18].

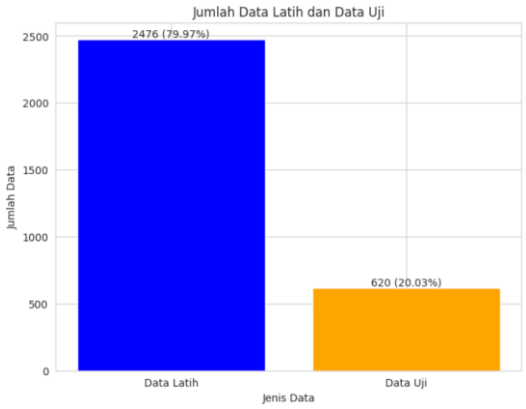
2.5 Pelabelan Data *(Lexicon Based)*

Pelabelan sentimen yang menggunakan leksikon memberikan label positif dan negatif pada sebuah komentar tanpa perlu melatih model terlebih dahulu. Ini dilakukan dengan menjumlahkan skor dari kata-kata yang ada, menggunakan kamus sentimen. Metode analisis sentimen berbasis leksikon bekerja dengan mengukur makna dari kata-kata dalam leksikon. Dalam analisis ini, leksikon yang sudah ada dipakai untuk menilai dokumen dengan cara menggabungkan nilai sentimen dari semua kata. Beberapa aturan sederhana bisa dipakai untuk mengatasi kata-kata yang menunjukkan negasi [19].

2.6 *Splitting Data* (80:20)

Pembagian data adalah langkah awal untuk melihat seberapa baik model bisa bekerja secara umum, sebagian dari data digunakan untuk melatih *(training)* dan sisanya untuk menguji *(testing)* supaya hasil yang didapat tidak terpengaruh oleh data yang digunakan untuk melatih. Rasio 80:20 dipilih karena memberikan cukup banyak data untuk belajar pola, sambil menyisakan bagian yang cukup untuk

melihat seberapa baik model di data yang belum dikenalnya. “Pembagian antara data latihan dan data uji adalah salah satu hal penting yang mempengaruhi seberapa akurat, *dataset* dibagi dengan beberapa rasio (90:10, 80:20, 70:30, 60:40, 50:50). Model dinilai menggunakan *confusion matrix* (akurasi, presisi, *recall*, *F1-score*). Dalam percobaan yang dilakukan, SVM dengan rasio 80:20 memperoleh akurasi paling tinggi dibandingkan dengan pengaturan lainnya” [20].



Gambar 2. *Splitting* Data

Pada Gambar 2, proses pembagian data memisahkan kumpulan data menjadi dua bagian, yaitu data yang digunakan untuk pelatihan dan data yang digunakan untuk pengujian. Data pelatihan berfungsi untuk mengajari model *Support Vector Machine*, sementara data pengujian dipakai untuk menilai seberapa baik kinerja model dengan menggunakan ukuran-ukuran evaluasi seperti akurasi, presisi, dan *recall*.

2.7 Penerapan Algoritma *Support Vector Machine*

Support Vector Machine adalah metode untuk mengelompokkan data yang berusaha menemukan pemisah terbaik dengan jarak paling lebar untuk memisahkan dua kelompok, sehingga dapat bekerja dengan baik bahkan pada data yang banyak fiturnya seperti yang diwakili oleh TF-IDF dalam teks. Algoritma pembelajaran mesin ini biasanya dipakai karena ia bisa mengelola data teks dan memberikan hasil yang tepat. *Support Vector Machine* adalah algoritma yang mencari garis pemisah terbaik untuk memisahkan data di ruang fitur yang memiliki banyak dimensi, dengan tujuan untuk memperbesar jarak antara dua kelompok data [21].

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil analisis sentimen yang didapat dari pengolahan komentar di TikTok tentang masalah dugaan pencemaran radioaktif di daerah industri Cikande. Setiap langkah mulai dari mengumpulkan data, mempersiapkan data, memberi label, membuat model, hingga menilai hasil dijelaskan satu per satu untuk menunjukkan alur kerja penelitian dan hasil dari model klasifikasi yang digunakan

3.1 Pengumpulan Data

Data dari penelitian ini diambil dari kolom komentar di TikTok, terutama pada beberapa saluran berita yang membahas tentang dugaan pencemaran radioaktif di area industri Cikande. Setiap komentar yang diambil mencakup informasi publik seperti nama akun dan isi komentar. Pengambilan data dilakukan secara otomatis menggunakan bahasa pemrograman *Python* dengan cara web *scraping*, lalu hasilnya diproses untuk menghilangkan komentar yang sama dan memastikan bahwa data yang disimpan berhubungan dengan topik penelitian.

Tabel 1. Hasil *Scraping* Data dari Tiktok

Author	*Comment*
@violova.id	ternyata amerika benar.. 😔 udangnya mengandung radioaktif
@ridhojulian11	Yg gak tau baca ini ya “Radioaktif adalah sifat... hati-hati makan hasil laut ya
@ceritafilm8587	ngeri lo kalau kena radiasi radio aktif.. bahaya
@rully_993	padahal orang dulu termasuk kakek saya sering deket area situ tapi sehat-sehat aja

Hasil *scrapping* data dapat dilihat pada tabel 1 berikut, kolom *author* berfungsi untuk mengidentifikasi akun pengguna TikTok yang memberikan komentar tanpa menampilkan informasi pribadi, sedangkan kolom *comment* merupakan sumber utama teks yang digunakan dalam analisis sentimen. Data komentar inilah yang menjadi dasar dalam proses pra-pemrosesan dan pelabelan sentimen, sehingga setiap opini publik dapat dianalisis berdasarkan isi tuturan yang sebenarnya.

3.2 Hasil *Pre-processing*

Tahap pra-pemrosesan dilakukan setelah kita mengumpulkan semua komentar agar teks bisa dianalisis secara otomatis. Tujuan dari langkah ini adalah untuk menyusun komentar yang masih mentah menjadi bentuk yang bersih, seragam, dan memiliki makna dengan menghilangkan bagian yang bukan teks dan mengubah kata-kata kembali ke bentuk dasarnya. Dengan langkah ini, setiap komentar akan diubah menjadi format kata yang lebih mudah untuk diproses oleh model pembelajaran mesin.

Berikut adalah hasil penerapan tahap pra-pemrosesan pada komentar:

Tabel 2. Hasil *Pre-Processing*

Tahapan	Hasil Komentar
Data Mentah	ternyata amerika benar.. 😔 udangnya mengandung radioaktif
Cleaning	ternyata amerika benar udangnya mengandung radioaktif
Case Folding	ternyata amerika benar udangnya mengandung radioaktif
Normalisasi Kata	ternyata amerika benar udang mengandung radioaktif
Tokenizing	[ternyata, amerika, benar, udang, mengandung, radioaktif]
Stopword Removal	[ternyata, amerika, benar, udang, radioaktif]
Stemming	[nyata, amerika, benar, udang, radioaktif]

Hasil *pre-processing* data dapat dilihat pada tabel 2, teks asli diubah menjadi kumpulan kata yang lebih pendek dan memiliki arti. Proses pembersihan menghapus tanda baca dan emoji yang tidak perlu, penyeragaman huruf membuat semua huruf menjadi kecil, normalisasi menyamakan bentuk kata seperti "udangnya" menjadi "udang", pemecahan kalimat memisahkan kalimat menjadi kata-kata kecil, penghilangan kata umum menghapus kata yang sering muncul dan tidak penting, dan pengembalian kata mengubah kata kembali ke bentuk aslinya. Hasil akhir ini digunakan untuk menghitung bobot TF-IDF dan melatih model *Support Vector Machine* (SVM) di tahap berikutnya.

3.3 Hasil Pelabelan Data *(Lexicon Based)*

Proses memberi label dilakukan dengan cara menggunakan pendekatan berbasis kamus, yaitu dengan mencocokkan setiap kata dalam komentar dengan daftar kata positif dan negatif yang ada di kamus sentimen. Kata-kata yang memiliki makna baik akan mendapat nilai +1, sedangkan

kata-kata yang mempunyai makna buruk akan diberi nilai -1. Total nilai dalam satu komentar akan menentukan kategorinya, di mana nilai lebih dari nol dianggap positif dan nilai kurang dari nol dianggap negatif.

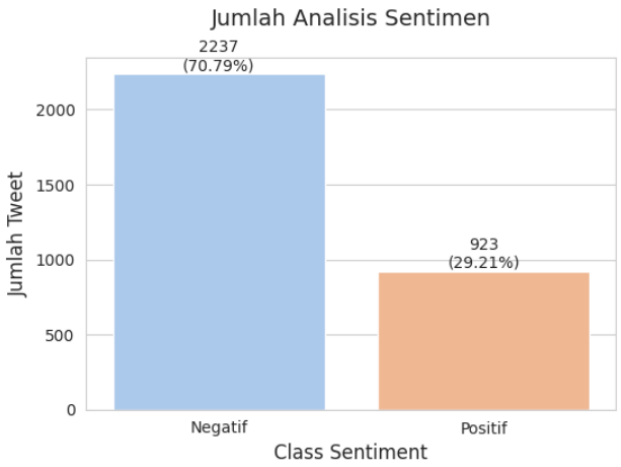
Berikut adalah contoh hasil pemberian label sentimen menggunakan metode berbasis kamus:

Tabel 3. Hasil Pelabelan Data Komentar TikTok

Author	*Comment*	Sentimen
violova.id	ternyata amerika benar...	Negatif
ridhojulian11	Yg gak tau baca ini ya...	Negatif
ceritafilm8587	ngeri lo kalau kena...	Negatif
rully_993	padahal orang dulu...	Positif

Hasil Pelabelan data dapat dilihat pada tabel 3, kita dapat melihat bahwa banyak komentar memiliki sentimen negatif. Ini menunjukkan bahwa orang-orang khawatir tentang bahaya dari paparan radiasi dan efeknya bagi kesehatan serta lingkungan. Di sisi lain, komentar yang positif menunjukkan pandangan yang lebih tenang, ragu-ragu, atau meyakini bahwa situasinya tidak seburuk yang diberitakan.

Visualisasi hasil pelabelan dapat dilihat pada Gambar di bawah, yang menunjukkan perbandingan antara komentar baik dan komentar buruk di TikTok.



Gambar 3. Hasil Pelabelan Data

Pada gambar 3 tersebut, kita bisa lihat bahwa komentar yang memiliki sentimen negatif lebih banyak dalam analisis ini. Ada total 2.237 komentar atau sekitar 70,79% yang dianggap negatif, sementara 923 komentar sekitar 29,21% dianggap positif. Ini menunjukkan bahwa mayoritas pengguna TikTok menunjukkan rasa khawatir, curiga, atau mengkritik masalah yang berkaitan dengan dugaan radioaktif, sedangkan sedikit yang memberikan pandangan positif atau penghiburan.

3.4 Penerapan dan Evaluasi Algoritma *Support Vector Machine*

Setelah semua langkah pembersihan dan pelabelan selesai, langkah selanjutnya adalah menggunakan algoritma *Support Vector Machine* (SVM) untuk mengklasifikasikan sentimen dari komentar di TikTok. Pada tahap ini, teks yang sudah dibersihkan dan distandarkan akan diwakili dengan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) supaya setiap kata mendapatkan nilai sesuai dengan pentingnya dalam dokumen.

Kumpulan data dibagi menjadi dua bagian, yaitu 80% untuk data latihan dan 20% untuk data pengujian. Data latihan digunakan untuk membangun model SVM, sedangkan data pengujian digunakan untuk mengukur seberapa baik model dapat mengenali pola sentimen yang baru. Hasil evaluasi didapatkan dengan melihat empat ukuran utama, yaitu *precision*, *recall*, *F1-score*, dan akurasi.

Classification Report for SVM:

	precision	recall	f1-score	support
Negatif	0.897	0.956	0.925	427.000
Positif	0.885	0.756	0.816	193.000
accuracy	0.894	0.894	0.894	0.894
macro avg	0.891	0.856	0.870	620.000
weighted avg	0.893	0.894	0.891	620.000

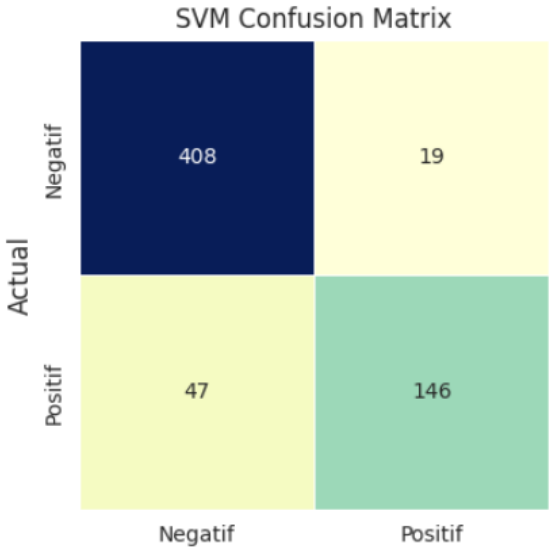
Gambar 4. *Classification Report*

Berdasarkan hasil *classification report* yang terdapat di gambar 4, model SVM mendapatkan akurasi sebesar 89%. Ini berarti model ini dapat mengenali pola sentimen dengan baik. Untuk nilai *precision* dan *recall*, kelas negatif sedikit lebih tinggi daripada kelas positif, yang menunjukkan bahwa model ini lebih peka terhadap komentar negatif. Namun, hasil untuk kelas positif tetap stabil dan menunjukkan kinerja yang seimbang.

Secara keseluruhan, algoritma SVM dengan metode pembobotan TF-IDF dapat menghasilkan klasifikasi yang tepat untuk analisis sentimen pada komentar TikTok. Kinerja dari model ini bisa ditingkatkan di masa mendatang dengan menambahkan berbagai jenis data serta menggunakan fitur *n-gram* untuk memperluas pemahaman konteks kata.

3.5 *Confusion Matrix*

Untuk mengetahui seberapa banyak kesalahan dalam masing-masing jenis sentimen, kita pakai analisis *Confusion Matrix*. Matriks ini memperlihatkan perbandingan antara hasil prediksi yang dihasilkan model dengan data sebenarnya pada dua kategori utama, yaitu positif dan negatif.



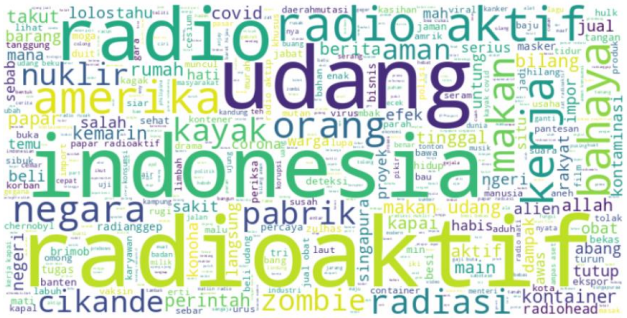
Gambar 5. *Confusion Matrix*

Visualisasi *confusion matrix* dapat dilihat pada gambar 5 tersebut, model *Support Vector Machine* (SVM) berhasil mengelompokkan sebagian besar komentar dengan akurat. Ada 408 komentar negatif yang teridentifikasi dengan benar, sedangkan 19 komentar negatif salah diidentifikasi sebagai positif. Di sisi lain, dari komentar positif, 146 berhasil terprediksi dengan benar, tetapi ada 47 komentar positif yang salah dikategorikan sebagai negatif.

Secara keseluruhan, hasil ini menunjukkan bahwa model memiliki tingkat kesalahan yang cukup rendah dan lebih baik dalam mengenali komentar bernada negatif. Ini sejalan dengan jumlah data yang lebih banyak pada kategori negatif, sehingga model memiliki lebih banyak variasi teks untuk dipelajari. Jadi, SVM dapat dikatakan memiliki kinerja yang stabil dan cukup efektif dalam mengidentifikasi sentimen komentar TikTok berdasarkan konteks topik radioaktif.

3.6 Visualisasi *Word Cloud*

Untuk menunjukkan penyebaran kata yang paling sering muncul dalam komentar, digunakan visualisasi *Word Cloud*. Teknik ini menampilkan seberapa sering sebuah kata muncul berdasarkan ukuran kata semakin besar ukuran kata tersebut, semakin sering kata itu muncul dalam kumpulan komentar TikTok.



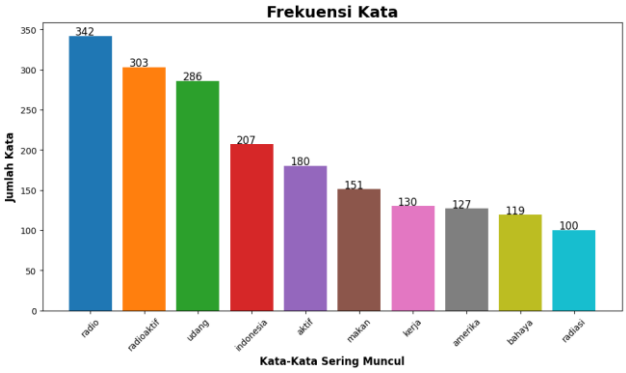
Gambar 6. *Word Cloud*

Pada gambar 6 tersebut, terlihat kata-kata seperti “radioaktif”, “indonesia”, “udang”, “bahaya” dan “cikande” muncul paling banyak. Ini menunjukkan bahwa isu limbah radioaktif yang terkait dengan udang dan lokasi Cikande menjadi topik utama yang sering dibicarakan pengguna TikTok.

Beberapa kata lain seperti “amerika”, “radiasi”, “kerja” dan “pabrik” juga sering muncul, menunjukkan adanya campuran antara kecemasan, dugaan, dan pembicaraan mengenai penyebab serta dampak dari kemungkinan kontaminasi tersebut. Visualisasi ini membantu menampilkan secara umum topik yang sedang dibahas dan persepsi masyarakat yang muncul dari komentar di media sosial.

3.7 Visualisasi Frekuensi Kata

Selain membuat visualisasi *Word Cloud*, juga dilakukan analisis menghitung berapa kali suatu kata muncul dalam komentar. Metode ini berguna untuk mengetahui kata-kata yang sering muncul dan penting dalam pembicaraan pengguna TikTok tentang topik radioaktif.



Gambar 7. Frekuensi kata

Visualisasi frekuensi kata dapat dilihat pada gambar 7 tersebut, kata "radio" muncul paling sering sebanyak 342 kali, diikuti oleh kata "radioaktif" (303 kali), "udang" (286 kali), dan "indonesia" (207 kali). Kemunculan kata-kata ini menunjukkan bahwa topik utama yang dibicarakan oleh masyarakat terpusat pada isu radioaktif di Indonesia, terutama yang berkaitan dengan produk udang.

Selain itu, kata-kata seperti "aktif", "makan", "kerja", "amerika", "bahaya" dan "radiasi" juga muncul dengan frekuensi tinggi. Hal ini menunjukkan adanya campuran antara pendapat kritis, kekhawatiran, dan juga pembicaraan umum tentang penyebab dan dampak dari isu radioaktif tersebut. Analisis ini mendukung temuan sebelumnya bahwa persepsi masyarakat di media sosial lebih banyak fokus pada bahaya dan dampak dari radioaktif, dibandingkan dengan narasi positif atau upaya menenangkan situasi.

4. KESIMPULAN

Hasil penelitian ini menunjukkan bahwa implementasi analisis sentimen efektif dalam memetakan persepsi publik

terkait isu dugaan kontaminasi radioaktif di kawasan industri Cikande melalui ekstraksi opini dari *platform* TikTok. Dengan mengikuti beberapa langkah seperti memproses teks terlebih dahulu, memberikan bobot dengan TF-IDF, dan menggunakan algoritma *Support Vector Machine* (SVM), penelitian ini berhasil menggambarkan pola pendapat publik secara teratur. Sebagian besar komentar menunjukkan pandangan yang negatif, terutama terkait kekhawatiran tentang kesehatan, dampak terhadap lingkungan, dan rendahnya kepercayaan pada keterbukaan informasi.

Hasil dari pengujian model menunjukkan bahwa SVM sangat baik dalam mengklasifikasikan sentimen dengan tingkat ketepatan 89%. *Confusion matrix* mencatat bahwa 408 komentar negatif dan 146 komentar positif berhasil dikenali dengan benar, sedangkan terdapat kesalahan pada 19 komentar negatif yang dianggap positif dan 47 komentar positif yang dianggap negatif. Distribusi data hasil penandaan berdasarkan leksikon memperlihatkan bahwa 70,79% komentar memiliki sentimen negatif dan 29,21% positif, yang sesuai dengan pola yang diprediksi oleh model. Temuan ini menunjukkan bahwa SVM bekerja dengan baik pada data teks yang tidak seimbang dan mampu menangkap pola sentimen yang lebih kuat dalam percakapan publik.

Berdasarkan temuan dari penelitian ini, disarankan agar studi berikutnya memperhatikan penggunaan kosakata yang lebih bervariasi, termasuk kata-kata informal dan *slang* yang biasa dipakai di media sosial, agar pelabelan bisa lebih akurat. Selain itu, bisa juga dilakukan pengujian dengan algoritma lain seperti *Naive Bayes* atau LSTM untuk melihat perbandingan kinerjanya. Pemerintah dan pihak-pihak yang berwenang juga harus meningkatkan cara berkomunikasi tentang risiko dengan lebih jelas dan konsisten agar bisa mengurangi kekhawatiran masyarakat, terutama pada masalah lingkungan yang sensitif.

Ucapan Terima Kasih

Penulis ingin mengucapkan terima kasih kepada Universitas Bina Sarana Informatika, khususnya Program Studi Teknik Informatika di Fakultas Teknik, untuk semua bimbingan, dukungan, dan sarana yang telah diberikan selama penelitian ini. Penulis juga ingin berterima kasih kepada teman-teman satu tim peneliti atas kerja sama yang baik dan kepada semua pihak yang telah membantu dalam menyelesaikan penelitian ini. Penulis menyadari bahwa penelitian ini belum sempurna dan berharap mendapatkan kritik dan saran yang berguna untuk perbaikan ke depan.

DAFTAR PUSTAKA

[1] F. A. Ryandi, D. Pratiwi, dan S. Sari, “Analisis Sentimen Masyarakat di Media Sosial X Terhadap Kemenkes Dengan Naive Bayes Dan SVM,” *Jurnal Sains dan Teknologi*, vol. 7, no. 1, hal. 1–6, 2025, doi:10.55338/saintek.v7i1.

[2] K. R. Amaliah, U. Enri, dan S. Defiyanti, “Analisis Sentimen Pengguna Website Siska Menggunakan Algoritma Naive Bayes Dengan Optimasi Particle Swarm Optimization,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 3, hal. 365–369, 2025, doi:10.23960/jitet.v13i3.6839.

[3] Z. Alhaq, A. Mustopa, S. Mulyatun dan J. D. Santoso “Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter," *Journal of Information System Management*, vol. 3, no. 1, hal. 16-21, 2021, doi:10.24076/joism.2021v3i2.558.

[4] A. G. Husna, H. Susanto, dan Purwono, “Penyisihan Unsur Stronsium (Sr) dan Cesium (Cs) Pada Limbah Radioaktif Dengan Kandungan Organik Polycyclic Aromatic Hydrocarbon (Pah) Menggunakan Membran Nanofiltrasi,” *Jurnal Teknik Lingkungan*, vol. 6, no. 1, hal. 1-20, 2017.

[5] D. Sukandar, G. F. Amri, Dasumiati, dan D. Fauzih, “Pemantauan Konsentrasi Radionuklida Cs-137 , Co-60 , I-131 Serta Kualitas Air Daerah Aliran Sungai Bekasi,” *Jurnal Ilmu Lingkungan*, vol. 23, no. 1, hal. 23–34, 2025, doi: 10.14710/jil.23.1.23-34.

[6] F. Lim dan A. T. Ayunda, “Analisis Sentimen Masyarakat Indonesia Terhadap Aplikasi Threads di Twitter Menggunakan Naïve Baiyes,” *Jurnal Sistem Informasi*, vol. 9, no. 2, hal. 55-64, 2023, doi:10.19109/jusifo.v9i2.19965.

[7] R. F. T. Wulandari, dan D. Anubhakti, “Implementasi Algoritma Support Vector Machine (SVM) dalam Memprediksi Harga Saham Pt. Garuda Indonesia Tbk,” *Indonesia Journal Information System*, vol. 4, no. 2, hal. 250–256, 2021, doi:10.36080/idealis.v4i2.2847.

[8] M. T. F. Maulana, B. I. Nugroho, dan E. U. S. Utami, “Analisis Sentimen Ulasan Aplikasi TikTok di Google Playstore Menggunakan Algoritma Naive Bayes,” *Journal of Articial Intelligence and Digital Business*, vol. 4, no. 3, hal. 6627–6635, 2025, doi:10.31004/riggs.v4i3.2962.

[9] R. Abidin dan A. Herawati, “Analisis Sentimen Publik Terhadap Kebijakan Program Tabungan Rakyat (Tapera),” *Journal of Information System and Computer*, vol. 4, no. 1, hal. 13–19, 2024, doi:10.34001/jister.v4i1.1002.

[10] N. Adila, F. Sembiring, dan W. Jatmiko, “Implementation of Web Scraping for Journal Data Collection on the Sinta Website,” *Jurnal dan Penelitian Teknik Informatika*, vol. 7, no. 4, hal. 2478–2485, 2022, doi:10.33395/sinkron.v7i4.11576.

[11] E. Apriani, I. F. Hanif, F. Oktavianalisti, L. D. H. Monasari, dan I. Winarni, “Analisis Sentimen Penggunaan TikTok Sebagai Media Pembelajaran Menggunakan Algoritma Naïve Bayes Classifier,” *Indonesian Journal of Machine Learning and*

Computer Science, vol. 4, no. July, hal. 1160–1168, 2024, doi:10.57152/malcom.v4i3.1482.

[12] B. Hakim, "Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning," *Journal of Business and Audit Information System*, vol. 4, no. 2, hal. 16–22, 2021, doi:10.30813/jbase.v4i2.3000.

[13] D. F. M. Dina, T. Haryanti, dan M. A. Haq, "Analisis Sentimen Terhadap Komentar Pada Media Sosial Tiktok Yang Berpotensi Menyebabkan Depresi Menggunakan Metode Naive Bayes," *Jurnal Ilmiah Computing Insight*, vol. 7, no. 1, hal. 1–9, 2025, doi:10.30651/comp_insight.v7i1.26327.

[14] Junaedi, A. H. Gunawan, V. Kuswanto, dan Jonathan, "Eksplorasi Algoritma Support Vector Machine untuk Analisis Sentimen Destinasi Wisata di Indonesia," *Journal Bit-Tech*, vol. 7, no. 2, hal.323-330, 2024, doi: 10.32877/bt.v7i2.1810.

[15] P. M. S. Ardinata, A. A. J. Permana, dan I. N. S. W. Wijaya, "Identifikasi Dan Normalisasi Teks Slang Dengan Fasttext Pada Twitter Dalam Bahasa Indonesia," *Jurnal Pendidikan Teknologi dan Kejuruan*, vol. 21, no. 1, hal. 34-44, 2024, doi:10.23887/jptkundiksha.v21i1.66381.

[16] J. Petrus, Ermatita, Sukemi, dan Erwin, "A Novel Approach : Tokenization Framework based on Sentence Structure in Indonesian Language," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 2, hal. 541–549, 2023, doi:10.14569/IJACSA.2023.0140264.

[17] R. Rinandyaswara, Y. A. Sari, dan M. T. Furqon, "Pembentukan Daftar Stopword Menggunakan Term Based Random Sampling Pada Analisis Sentimen Dengan Menggunakan Metode Naive Bayes," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 4, hal. 717–724, 2022, doi: 10.25126/jtiik.202294707.

[18] I. M. A. Agastya, "Pengaruh Stemmer Bahasa Indonesia Terhadap Peforma Analisis Sentimen Terjemahan Ulasan Film," *Jurnal Tekno Kompak*, vol. 12, no. 1, hal. 18–23, 2018, doi:10.33365/jtk.v12i1.70.

[19] F. A. J. Nur, A. Romadhony, dan Hasmawati, "Analisis Sentimen Pada Media Sosial Universitas Dengan Metode Berbasis Leksikon," *e-Proceeding of Engineering*, vol. 10, no. 2, hal. 1691–1697, 2023.

[20] Y. A. Prasetyo, E. Utami, dan A. Yaqin, "Pengaruh Komposisi Split Data Terhadap Performa Akurasi Analisis Sentimen Algoritma Naïve Bayes dan SVM," *Journal of Electrical Engineering and Computer*, vol. 6, no. 2, hal. 382–390, 2024, doi:10.33650/jeecom.v4i2.

[21] R. Ramlan, N. Satyahadewi, dan W. Andani, "Analisis Sentimen Pengguna Twitter Menggunakan Support Vector Machine Pada Kasus Kenaikan Harga Bbm," *Jambura Journal of Mathematics*, vol. 5, no. 2, hal. 431–445, 2023, doi:10.34312/jjom.v5i2.20860.