



## ANALISIS SENTIMEN PUBLIK KRISIS EKONOMI UNTUK INOVASI PRODUK ASURANSI: PERBANDINGAN *LEXICON* DAN SVM PADA MEDIA SOSIAL

Nalar Istiqomah<sup>1</sup>, Fanny Novika<sup>2</sup>

<sup>1,2</sup>Sekolah Tinggi Manajemen Asuransi Trisakti

Program Studi Aktuaria

Kampus C Trisakti, Gedung A Jl. Jend. Ahmad Yani Kav. 85, Jakarta Timur 13210

e-mail: <sup>1</sup>[nalar.istiqomah23@gmail.com](mailto:nalar.istiqomah23@gmail.com), <sup>2</sup>[novikafanny@gmail.com](mailto:novikafanny@gmail.com)

### Abstrak

Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap isu ekonomi dan asuransi sebagai dasar dalam mengidentifikasi peluang inovasi produk asuransi yang sesuai dengan kebutuhan masyarakat. Data dikumpulkan dari dua platform media sosial, yaitu Twitter dan YouTube, melalui metode web scraping menggunakan bahasa pemrograman Python, dengan total 18.530 tweet dan 233.782 komentar. Analisis sentimen dilakukan menggunakan dua pendekatan, yakni berbasis lexicon dan machine learning dengan algoritma Support Vector Machine (SVM). Hasil analisis berbasis lexicon menunjukkan dominasi sentimen positif, namun disertai tingkat akurasi dan presisi yang rendah, bahkan setelah dilakukan perluasan kamus manual yang disesuaikan dengan konteks bahasa media sosial. Temuan ini menunjukkan keterbatasan metode lexicon dalam menangkap konteks semantik secara menyeluruh. Sebaliknya, pendekatan berbasis SVM menghasilkan performa yang jauh lebih baik, dengan tingkat akurasi sebesar 98% untuk data Twitter dan 96,7% untuk data YouTube. Berdasarkan hasil confusion matrix, model paling akurat dalam mengidentifikasi sentimen negatif, sedangkan akurasi untuk sentimen positif dan netral relatif lebih rendah akibat ketidakseimbangan kelas (class imbalance). Secara umum, hasil penelitian menunjukkan bahwa sentimen negatif mendominasi kedua platform, mencerminkan adanya ketidakpuasan dan kekhawatiran publik terhadap kondisi ekonomi saat ini. Meskipun demikian, sebagian kecil sentimen positif mengindikasikan apresiasi dan dukungan terhadap upaya di sektor asuransi. Temuan ini memberikan implikasi penting bagi pengembangan inovasi produk asuransi yang lebih berorientasi pada kebutuhan konsumen. Dengan demikian, pemahaman terhadap persepsi publik di media sosial dapat menjadi landasan strategis dalam merancang produk asuransi yang lebih inklusif, transparan, dan relevan bagi masyarakat digital.

**Kata kunci:** analisis sentimen, inovasi produk asuransi, lexicon-based, SVM, media sosial

### Abstract

This study aims to analyze public sentiment toward economic and insurance-related issues as a basis for identifying opportunities for insurance product innovation that align with public needs. Data were collected from two social media platforms, Twitter and YouTube, using a web scraping technique implemented in Python, resulting in a total of 18,530 tweets and 233,782 comments. Sentiment analysis was conducted using two approaches: a lexicon-based method and a machine learning model employing the Support Vector Machine (SVM) algorithm. The lexicon-based analysis revealed a predominance of positive sentiment; however, it exhibited low accuracy and precision, even after incorporating a manually expanded dictionary adapted to the linguistic characteristics of social media. This indicates the lexicon method's limitation in capturing contextual semantics comprehensively. In contrast, the SVM-based approach demonstrated significantly improved performance, achieving an accuracy of 98% for Twitter data and 96.7% for YouTube data. According to the confusion matrix, the model was most accurate in identifying negative sentiment, while the accuracy for positive and neutral classes was relatively lower due to class imbalance in the dataset. Overall, the findings show that negative sentiment dominates both platforms, reflecting public dissatisfaction and concern about

*current economic conditions. Nevertheless, a small portion of positive sentiment indicates appreciation and support for developments in the insurance sector. These insights provide valuable implications for the development of consumer-oriented insurance product innovations. Thus, understanding public perceptions on social media can serve as a strategic foundation for designing insurance products that are more inclusive, transparent, and relevant to the digital society.*

**Keywords:** *sentiment analysis, insurance product innovation, lexicon-based, SVM, social media*

## 1. PENDAHULUAN

Indonesia tengah menghadapi beragam tantangan krisis ekonomi di tahun 2025 ini. Faktor-faktor tersebut disebabkan oleh melambatnya ekonomi global, berkurangnya daya beli masyarakat, dan kemungkinan kenaikan harga energi, yang menjadi isu utama [1]. Krisis ekonomi tidak hanya menimbulkan tekanan pada indikator makroekonomi seperti inflasi dan nilai tukar, tetapi juga berpengaruh langsung pada kehidupan masyarakat, terutama di sektor pekerjaan. Saat daya beli menurun dan sektor industri terpengaruh, masyarakat dihadapkan pada masalah penghasilan, PHK dan ketidakpastian masa depan. Dalam konteks ini, menyadari pandangan masyarakat terhadap krisis menjadi krusial, karena hal ini dapat mencerminkan tingkat kepercayaan dan kepanikan yang muncul di kalangan warga.

Menurut laporan Badan Pusat Statistik (BPS), total penduduk usia kerja di Indonesia pada Februari 2025 mencapai 216,79 juta orang, meningkat 2,79 juta dari Februari 2024. Sementara itu, total angkatan kerja bertambah sebesar 2,46 juta individu, mencapai 153 juta individu. Namun, jumlah pengangguran justru meningkat dari 7,20 juta di Februari 2024 menjadi 7,28 juta orang pada Februari 2025, dimana Tingkat Pengangguran Terbuka (TPT) mencapai 4,76%. Peningkatan ini menunjukkan bahwa pasar kerja belum sanggup menyerap tambahan tenaga kerja [2].

Keadaan itu diperburuk oleh tingkat inflasi yang tinggi pada tahun-tahun sebelumnya. BPS juga mencatat bahwa Indonesia menghadapi inflasi yang signifikan, terutama pada kategori makanan dan energi, yang memperburuk kemampuan beli masyarakat, khususnya kelompok yang rentan [3].

Dalam era digital, tanggapan masyarakat mengenai situasi ini banyak terlihat di media sosial seperti Twitter dan YouTube, yang berfungsi sebagai *platform* untuk menyampaikan kritik, keluhan, atau dukungan terhadap kebijakan ekonomi. Analisis sentimen masyarakat mengenai krisis ekonomi perlu dilakukan tidak hanya untuk mempelajari pandangan publik, tetapi juga untuk menjadi landasan dalam merancang inovasi produk dan layanan asuransi digital. Dengan mengetahui kecenderungan opini publik, sektor asuransi dapat mengadaptasi strategi komunikasi, produk mikro, serta layanan digital yang lebih peka terhadap kebutuhan konsumen saat krisis.

Dalam penelitian ini, istilah asuransi digital tidak hanya mengacu pada penjualan produk asuransi melalui saluran digital, tetapi juga mencakup transformasi menyeluruh industri asuransi berbasis teknologi. Asuransi digital mencakup proses digitalisasi dari hulu ke hilir, termasuk pemasaran produk, proses klaim, layanan pelanggan berbasis aplikasi, hingga pengembangan produk berbasis data (*data-driven insurance*) [4]. Konsep ini sejalan dengan perkembangan *digital banking* seperti Bank Jago atau SeaBank, yang beroperasi sepenuhnya dalam ekosistem digital tanpa ketergantungan pada cabang fisik.

Dalam konteks ini, inovasi asuransi digital mencakup dua dimensi utama yaitu digitalisasi layanan dan *digital-native insurance*. Dengan digitalisasi layanan dapat meningkatkan akses, efisiensi, dan transparansi dalam pembelian maupun pengelolaan polis dengan penggunaan teknologi digital. Adapun *digital-native insurance* adalah pengembangan model bisnis baru yang sejak awal dirancang berbasis teknologi (misalnya asuransi mikro berbasis aplikasi atau *insurtech*) [5]. Dengan demikian, hasil analisis sentimen publik dalam penelitian ini diharapkan dapat memberikan masukan strategis bagi pengembangan produk, layanan, dan komunikasi asuransi digital, baik oleh perusahaan asuransi konvensional yang melakukan transformasi digital maupun oleh perusahaan asuransi digital murni (*fully digital insurer*).

Pemilihan Twitter dan YouTube dalam studi ini didasarkan pada sifat khas kedua *platform* yang saling mendukung dalam menggambarkan pendapat masyarakat. Twitter terkenal sebagai *platform* media sosial yang cepat dan responsif, sering dimanfaatkan untuk menyampaikan pendapat dengan singkat, emosional, dan spontan, terutama dalam keadaan krisis atau tuduhan nasional [6]. Sementara itu, YouTube tidak hanya berfungsi sebagai *platform* untuk berbagi video, tetapi juga sebagai tempat diskusi publik melalui kolom komentar yang lebih panjang dan naratif, yang memberikan konteks yang

lebih mendalam untuk analisis sentiment [7]. Kedua *platform* tersebut memiliki jumlah pengguna yang besar dan aktif di Indonesia, serta menawarkan data publik yang dapat diakses secara sah untuk penelitian [8]. Walaupun *platform* media sosial lain seperti Facebook, Instagram, dan TikTok juga banyak dipakai, kendala akses data (API) dan privasi menjadikan Twitter dan YouTube pilihan yang lebih tepat untuk penelitian akademis yang mengandalkan text mining [9].

Twitter dan Youtube memberikan kesempatan kepada masyarakat untuk membangun opini dan emosi dengan cepat, dan bisa juga memengaruhi pandangan publik yang lain. Namun, sifat khas dari setiap *platform* menghadirkan tantangan yang berbeda dalam pengolahan dan interpretasi data [10]. Dalam konteks ini, analisis sentimen menjadi teknik yang krusial untuk menangkap pendapat masyarakat yang berbasis teks. Dua metode banyak yang digunakan adalah metode berbasis kamus (*lexicon-based*) dan metode pembelajaran mesin seperti *Support Vector Machine* (SVM). Pendekatan *lexicon* lebih mudah digunakan tetapi sensitif terhadap ambiguitas konteks, sedangkan akurasi SVM lebih baik tetapi memerlukan data pelatihan berkualitas [10].

Penelitian sebelumnya juga menunjukkan bahwa algoritme rekomendasi di *platform* seperti YouTube dapat memperkuat kecenderungan opini publik, terutama mengenai isu politik atau ekonomi [11]. Dengan kata lain, selain merefleksikan pendapat, media sosial juga bisa membentuk pandangan masyarakat melalui algoritme yang memperkuat eksposur terhadap konten tertentu. Selain itu, *text mining* dan *machine learning* menjadi alat yang semakin krusial dalam menganalisis data opini publik, terutama dalam situasi krisis ekonomi dan ramalan pasar [12] [13].

Beberapa studi sebelumnya telah membandingkan efektivitas metode berbasis *lexicon* dan *Support Vector Machine* (SVM) dalam analisis sentimen. Hasil penelitian Ratnaswari, *et al.* [14] menunjukkan bahwa metode SVM mampu memberikan performa yang lebih akurat pada analisis sentimen pada isu Pilpres di Twitter, dengan tingkat akurasi mencapai 93%. Ahmad dan rekan-rekan [15] mengembangkan sistem analisis sentimen Bahasa Indonesia dengan pendekatan *lexicon* dan *Naïve Bayes*. Dari hasil penelitian tersebut diketahui bahwa pendekatan *Naïve Bayes* memiliki hasil yang lebih akurat. Rintyarna dkk. [16] juga mengembangkan *Ina-SASet*, sebuah leksikon sentimen Bahasa Indonesia yang disusun dari lebih dari seratus ribu cuitan Twitter dan dirancang untuk menghasilkan analisis sentimen yang lebih halus serta relevan terhadap dinamika bahasa di media sosial. Meskipun demikian, pendekatan berbasis leksikon ini tetap memiliki keterbatasan dalam menghadapi variasi bahasa tidak baku, sarkasme, dan perubahan makna yang sangat dipengaruhi oleh konteks kalimat. Di sisi lain, SVM memberikan kinerja yang lebih baik, sebagaimana dilaporkan oleh Pamungkas *et al.* [17], yang mencatat bahwa akurasi SVM sebesar 84,21% untuk klasifikasi sentimen berita kasus benih lobster 2020 di *platform* Twitter. Zhang *et al.* [13] juga menjadikan SVM sebagai acuan yang handal dalam penelitian analisis sentimen, khususnya untuk dataset menengah. Walaupun metode *machine learning* biasanya lebih baik dalam akurasi, metode *lexicon* tetap penting, terutama dalam penelitian eksploratif, studi dengan sumber daya terbatas, atau saat interpretabilitas sangat dibutuhkan [14].

Sehingga, studi ini bertujuan untuk menganalisis sentimen masyarakat terhadap krisis ekonomi Indonesia dengan membandingkan dua metode, yaitu berbasis *lexicon* dan SVM, pada data yang diperoleh dari Twitter dan komentar di YouTube. Hasil analisis sentimen ini tidak sekadar menggambarkan opini publik, tetapi menjadi landasan penting bagi perusahaan asuransi untuk merancang produk dan layanan digital yang inovatif, relevan, dan responsif terhadap kebutuhan nyata konsumen, sekaligus meningkatkan kepercayaan masyarakat terhadap industri asuransi.

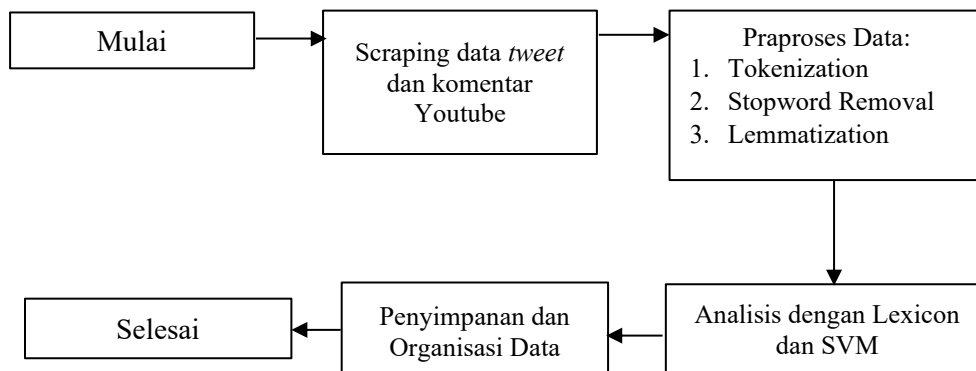
## 2. METODE PENELITIAN

Penelitian ini menggunakan data primer berbasis teks yang dikumpulkan dari dua *platform* media sosial utama, yaitu Twitter dan YouTube, melalui teknik *web scraping*. Teknik ini dipilih karena memungkinkan peneliti mengakses data publik dalam jumlah besar dan *real-time*, yang merepresentasikan opini dan sentimen masyarakat terhadap isu krisis ekonomi. Adapun detail pengumpulan data dijelaskan pada Gambar 1 [18].

Data yang digunakan penelitian ini berasal dari dua sumber, yaitu:

1. Data tweet dari Twitter dimanfaatkan sebagai sumber data karena *platform* ini memiliki sifat percakapan yang cepat dan luas, dengan banyak topik yang menjadi viral dalam waktu singkat. Cuitan-cuitan tentang situasi ekonomi, pemutusan hubungan kerja (PHK), harga kebutuhan pokok, dan inflasi sering kali menjadi topik yang hangat di *platform* ini.

2. Data komentar video YouTube dipilih karena mencerminkan konten yang ditonton, dan sering kali mengandung pandangan panjang yang menggambarkan sikap atau emosi penonton terkait isu-isu yang dibahas, termasuk keadaan ekonomi nasional.



Gambar 1. Langkah Analisis Data

Pengumpulan data dari Twitter dilakukan melalui aplikasi Tweet Crawler yang berbasis pemrograman Python. Aplikasi ini tidak memerlukan akses API resmi, sehingga lebih fleksibel untuk mengumpulkan cuitan yang berhubungan dengan kata kunci tertentu seperti “krisis ekonomi”, “PHK”, “inflasi”, “utang negara”, dan “resesi”. Data yang dihimpun mencakup konten cuitan dan tanggal publikasi dari Januari 2022 sampai Juli 2025. Periode ini dipilih karena mencerminkan keadaan ekonomi setelah pandemi dan mencakup beragam dinamika kebijakan ekonomi baik nasional maupun global. Adapun data dari YouTube diambil dari komentar video yang membahas isu ekonomi dengan memanfaatkan aplikasi youtube-comment-scraper. Video yang dipilih adalah video populer dan relevan berdasarkan kata kunci yang serupa di Twitter, dengan jangka waktu unggahan yang mencakup tahun 2022 hingga Juli 2025. Komentar yang diperoleh mencakup konten teks dan waktu rilis, tanpa mencantumkan nama akun pengguna untuk menjaga etika dan privasi. Data tersebut kemudian disimpan dalam format CSV.

Setelah pengumpulan data, tahap praproses data dilakukan untuk memastikan kualitas teks yang akan dianalisis [19]. Proses ini melibatkan pembersihan teks dari elemen yang tidak penting seperti URL, sebutan (*mention*), tagar (*hashtag*), emoji, angka, dan tanda baca. Selanjutnya, seluruh naskah diubah menjadi huruf kecil (*lowercasing*) dan dibagi menjadi token (kata-kata terpisah). Kata-kata umum yang tidak memiliki arti penting seperti “yang”, “dari”, dan “itu” dihilangkan melalui proses penghapusan *stopword* menggunakan *library* kata Bahasa Indonesia. Proses *stemming* juga dilakukan dengan memanfaatkan *library* Sastrawi, sehingga setiap kata dapat dikembalikan ke bentuk awalnya (tanpa imbuhan). Kemudian, dilakukan penyaringan untuk menghapus data ganda, komentar *spam*, atau teks yang terlalu singkat. Validasi manual juga dilakukan secara acak untuk memastikan bahwa data benar-benar relevan dan mencerminkan pandangan publik terhadap masalah ekonomi.

Untuk keperluan klasifikasi menggunakan pendekatan *Support Vector Machine* (SVM), data yang telah menjalani proses praproses selanjutnya diberi label sentimen. Pelabelan ini dilakukan secara manual, dengan kategori positif, negatif, dan netral. Label manual ditetapkan berdasarkan pemahaman konteks isi teks oleh peneliti. Proses penandaan ini penting agar SVM dapat dilatih untuk mengidentifikasi pola sentimen dalam teks. Adapun metode berbasis *lexicon* tidak memerlukan data yang sudah diberi label karena analisis dilakukan dengan membandingkan kata-kata dalam teks dengan daftar kata yang telah ditentukan sebelumnya berdasarkan sentimen.

Analisis data dalam studi ini dilakukan dengan dua pendekatan utama, yaitu pendekatan berbasis *lexicon* dan pendekatan *machine learning* menggunakan algoritme *Support Vector Machine* (SVM). Pendekatan berbasis *lexicon* dilakukan dengan mencocokkan setiap kata dalam teks dengan daftar kata yang memiliki orientasi sentimen (kamus sentimen) bahasa Indonesia, seperti SentiStrength-ID atau kamus sentimen Bahasa Indonesia karya Rintyarna dkk [16]. Jumlah skor sentimen untuk setiap kata dihimpun untuk mendapatkan nilai akhir sentimen dari setiap teks. Apabila total skor lebih besar dari nol, maka dianggap sebagai sentimen positif; jika kurang dari nol, dianggap negatif; dan jika nol,

diklasifikasikan sebagai netral. Metode ini tidak membutuhkan proses pelatihan model dan sesuai untuk analisis dasar.

Adapun pendekatan kedua adalah algoritme *Support Vector Machine* (SVM), yang memerlukan data berlabel. Setelah proses pelabelan selesai, data dibagi menjadi data latih dan data uji, dengan perbandingan 80:20. Model SVM selanjutnya dilatih untuk mengidentifikasi sentimen berdasarkan pola fitur yang ada. Setelah pelatihan selesai, data uji digunakan untuk menguji model dalam menilai kinerjanya melalui metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1-score*.

Lebih lanjut, pendekatan *lexicon* tidak memerlukan pelabelan data karena bersifat *unsupervised*, prosesnya lebih cepat, dan mudah digunakan untuk eksplorasi awal. Akan tetapi, tingkat akurasinya sangat bergantung pada keterlengkapan dan mutu kamus sentimen. Sebaliknya, SVM yang merupakan metode pembelajaran terawasi membutuhkan data berlabel, tetapi biasanya menghasilkan hasil yang lebih akurat karena mampu mempelajari pola linguistik rumit dalam teks. SVM lebih mampu menangani variasi bahasa, istilah gaul, dan konteks kalimat, terutama di media sosial yang seringkali bersifat santai. Secara singkat, perbedaan antara keduanya terlihat dalam tabel di bawah ini:

**Tabel 1.** Perbedaan metode *lexicon*-based dan SVM

Aspek	<i>Lexicon</i> -Based	Support Vector Machine (SVM)
Jenis Pendekatan	<i>Unsupervised</i>	<i>Supervised</i>
Kebutuhan Label Data	Tidak perlu	Perlu data berlabel (positif, negatif, netral)
Kamus/Model	Kamus Sentimen Bahasa Indonesia	Model SVM + fitur TF-IDF
Waktu Pemrosesan	Cepat (langsung skor per teks)	Lebih lama (butuh <i>training</i> model)
Akurasi	Tergantung kelengkapan kamus	Umumnya lebih tinggi jika dilatih dengan baik
Fleksibilitas	Terbatas pada kata yang ada di kamus	Dapat menangkap pola kalimat kompleks
Cocok diterapkan untuk	Baseline, eksplorasi awal	Analisis mendalam dan prediksi sentimen otomatis
Contoh Tools	SentiStrength-ID, kamus In-Set	Scikit-learn, TF-IDF, SVM <i>classifier</i>

Setelah mendapatkan hasil sentimen dari kedua metode, dilakukan analisis kinerja dan distribusi sentimen berdasarkan *platform* Twitter dan YouTube. Visualisasi disajikan dalam bentuk grafik batang, dan *confusion matrix* untuk memberikan pemahaman yang lebih jelas mengenai kecenderungan opini masyarakat terhadap krisis ekonomi. Perbandingan ini diharapkan dapat memberikan wawasan yang mendalam mengenai efektivitas setiap metode dan dinamika perasaan publik di dua *platform* media sosial yang berbeda.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Hasil Penelitian

##### a. Pengumpulan Data

Analisis sentimen dalam penelitian ini dilakukan terhadap dua sumber utama, yaitu Twitter dan YouTube. Data dari kedua *platform* dikumpulkan melalui proses *web scraping* menggunakan bahasa pemrograman Python dengan memanfaatkan aplikasi *Tweet Crawler* untuk Twitter dan YouTube Data API untuk YouTube. Proses ini memungkinkan pengambilan data komentar dan cuitan secara otomatis berdasarkan kata kunci yang berkaitan dengan topik ekonomi, asuransi, dan literasi keuangan.

Dari hasil pengumpulan data, diperoleh total 18.530 *tweet* dan 233.827 komentar YouTube. Perbedaan jumlah data yang cukup signifikan antara kedua *platform* ini mencerminkan perbedaan karakteristik pengguna dan pola interaksi di masing-masing media sosial. YouTube, yang berbasis konten video, cenderung mendorong pengguna untuk mengekspresikan pendapat lebih panjang di kolom komentar. Hal ini menjelaskan mengapa jumlah komentar yang terkumpul jauh lebih besar. Banyak video yang membahas topik ekonomi dan asuransi juga memancing diskusi aktif, baik berupa dukungan, kritik, maupun pengalaman pribadi pengguna terkait produk asuransi atau kondisi ekonomi.

Sebaliknya, Twitter memiliki keterbatasan panjang teks (maksimum 280 karakter) dan pola komunikasi yang lebih cepat serta ringkas. Akibatnya, jumlah *tweet* yang relevan dengan topik asuransi dan ekonomi relatif lebih sedikit dibandingkan YouTube. Selain itu, pengguna Twitter cenderung lebih tertarik membahas isu-isu aktual atau politik, sehingga diskusi mendalam mengenai literasi keuangan

dan asuransi tidak terlalu dominan. Dengan demikian, perbedaan volume data antara kedua *platform* ini juga mencerminkan perbedaan dinamika percakapan dan preferensi pengguna dalam membahas isu ekonomi dan asuransi di ruang digital.

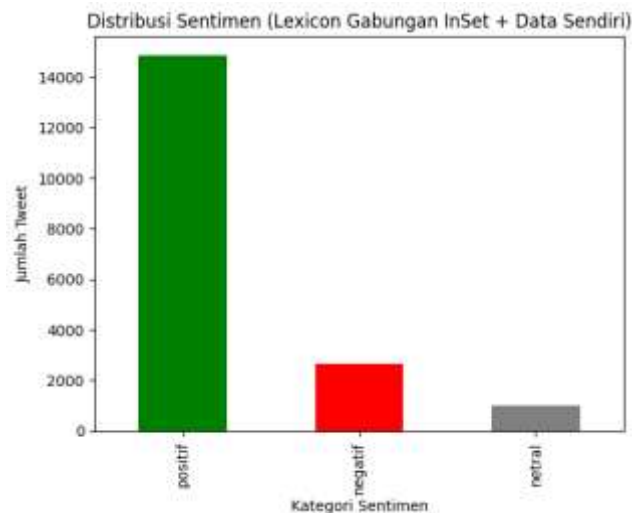
### **b. Praproses Data**

Sebelum masuk ke tahap analisis, dilakukan praproses terlebih dahulu, yaitu membuang *hashtag*, *mention*, *stopword* dan melakukan *stemming*. Setelah tahap praproses selesai, data diberi label sentimen (positif, negatif, dan netral) secara manual sebagai *gold standard* yang akan digunakan untuk pelatihan dan evaluasi model [20]. Pemberian label dilakukan oleh 3 orang peneliti secara independen untuk memastikan konsistensi. Berdasarkan hasil *labeling* manual, diketahui bahwa dari data *tweet*, terdapat 18.142 sentimen negatif, 231 sentimen netral, dan 157 sentimen positif. Sementara itu, dari data komentar YouTube diperoleh 230.150 sentimen negatif, 1.088 sentimen netral, dan 2.589 sentimen positif.

Distribusi data tersebut menunjukkan adanya ketimpangan kelas (*class imbalance*) yang cukup besar, dimana sentimen negatif mendominasi baik pada data Twitter maupun YouTube. Ketidakseimbangan ini berpotensi memengaruhi hasil analisis sentimen, karena model atau metode yang digunakan cenderung akan lebih sering mengenali kelas dengan jumlah data terbanyak. Dengan demikian, hasil analisis perlu ditafsirkan secara hati-hati, khususnya pada kategori positif dan netral yang jumlah datanya relatif sedikit.

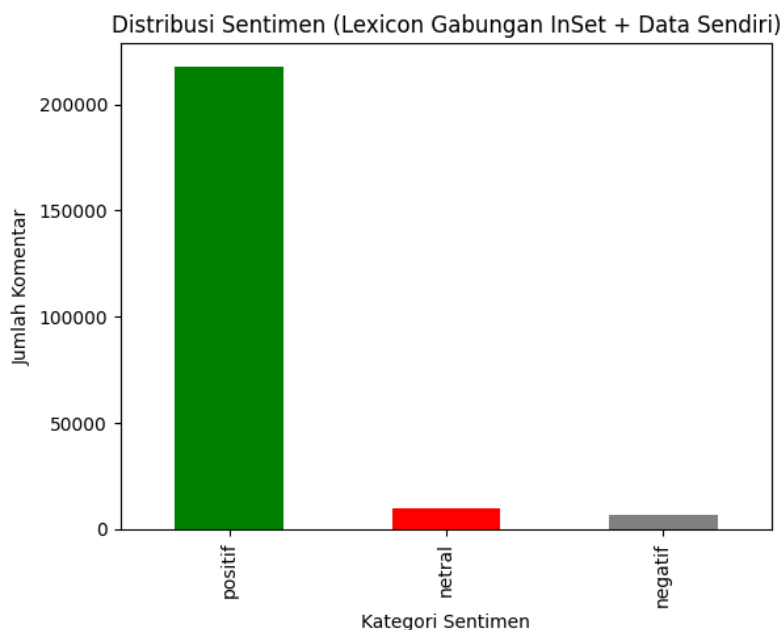
### **c. Analisis Data Menggunakan metode *Lexicon***

Tahap pertama analisis dilakukan menggunakan pendekatan *lexicon*-based dengan menggabungkan kamus InSet dan kamus tambahan yang dikembangkan sendiri. Kamus tambahan dikembangkan secara manual dengan tujuan menyesuaikan konteks bahasa yang digunakan di media sosial, khususnya pada *platform* Twitter dan YouTube yang sering kali memuat kata tidak baku, singkatan, serta ekspresi emosional yang tidak terdapat dalam kamus sentimen konvensional seperti InSet. Proses pengembangannya dilakukan melalui beberapa tahapan. Pertama, dilakukan analisis frekuensi kata menggunakan *term frequency-inverse document frequency* (TF-IDF) untuk mengidentifikasi kata-kata yang paling sering muncul dalam data. Dari daftar kata tersebut, peneliti kemudian melakukan seleksi manual terhadap kata yang memiliki muatan emosional atau opini, baik yang positif maupun negatif. Selanjutnya, setiap kata dianalisis berdasarkan konteks penggunaannya dalam kalimat. Misalnya, kata “anjir”, “parah”, “mantap”, atau “gila” dapat memiliki makna positif maupun negatif tergantung konteksnya. Penentuan polaritas sentimen (positif, negatif, netral) dilakukan dengan mempertimbangkan makna dominan dari kata tersebut dalam data yang diamati. Setelah daftar kata diperoleh, kamus tambahan ini kemudian digabungkan dengan kamus InSet untuk memperluas cakupan kata pada proses analisis sentimen. Dengan demikian, model *lexicon* tidak hanya mengandalkan kata baku dari kamus utama, tetapi juga mengenali *slang*, istilah populer, serta ekspresi khas media sosial yang sering digunakan oleh pengguna dalam percakapan daring.



Gambar 2. Hasil sentimen tweet menggunakan Lexicon

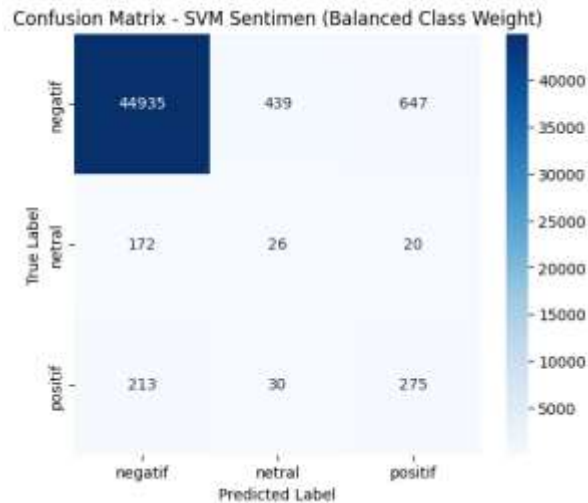
Hasil analisis pada data Twitter menggunakan kamus Inset dan kamus tambahan, ditunjukkan pada Gambar 2. Dari Gambar 2 terlihat dominasi sentimen positif yang sangat kuat, dengan 14.868 tweet termasuk dalam kategori positif, sedangkan 2.654 tweet dikategorikan negatif dan hanya 1.008 tweet yang bersifat netral. Pola serupa juga terlihat pada hasil lexicon data Youtube di Gambar 3, dimana 217.958 komentar termasuk dalam kategori positif, sementara komentar netral berjumlah 9.353 dan komentar negatif 6.471.



Gambar 3. Hasil sentimen komentar Youtube menggunakan Lexicon

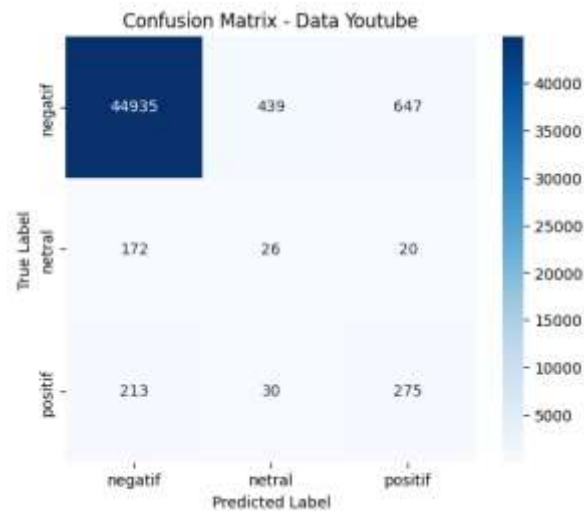
#### d. Analisis Data Menggunakan metode SVM

Evaluasi kinerja model SVM pada penelitian ini dilakukan dengan menggunakan *confusion matrix* untuk mengukur sejauh mana model mampu mengklasifikasikan sentimen (positif, negatif, dan netral) dengan benar. *Confusion matrix* menggambarkan perbandingan antara label sebenarnya hasil *labeling* manual dan label prediksi yang dihasilkan oleh model. Melalui matriks ini, dapat diidentifikasi pola kesalahan model dan tingkat keberhasilannya dalam mengenali setiap kategori sentimen. Untuk mengevaluasi kinerja model, data dibagi menjadi dua yaitu data latih dan data uji. Sebanyak 80% data dijadikan sebagai data latih dan 20% sebagai data uji untuk menguji kinerja model.



Gambar 4. Hasil sentimen tweet menggunakan SVM

Gambar 4 menunjukkan *confusion matrix* pada data Twitter. Dari gambar, dapat dilihat bahwa model SVM telah bekerja dengan baik setelah dilakukan pelatihan menggunakan data berlabel. Sebagian besar data negatif berhasil diklasifikasikan dengan benar. Namun untuk data netral dan positif masih terdapat kesalahan prediksi. Sementara itu, Gambar 5 menunjukkan *confusion matrix* pada data Youtube memperlihatkan pola yang mirip. Model cenderung sangat baik mengenali komentar negatif, tetapi masih kurang konsisten dalam membedakan antara komentar netral dan positif. Hal ini disebabkan adanya *class imbalance* pada data *tweet* dan komentar Youtube yang didominasi kelas negatif.



Gambar 5. Hasil sentimen komentar Youtube menggunakan SVM

### 3.2 Pembahasan

#### a. Analisis Data Menggunakan *Lexicon*

Kebaruan (*novelty*) penelitian ini terletak pada penggunaan dua sumber data media sosial secara simultan, yaitu Twitter dan YouTube, yang masih jarang dilakukan dalam penelitian analisis sentimen berbasis konteks ekonomi dan asuransi di Indonesia. Selain itu, penelitian ini tidak hanya membandingkan metode *lexicon* dan SVM, tetapi juga melakukan penyesuaian kamus sentimen secara manual agar sesuai dengan karakteristik bahasa media sosial. Kombinasi ini memberikan perspektif yang lebih komprehensif mengenai persepsi publik dan berkontribusi sebagai model dasar analisis sentimen untuk sektor asuransi digital di Indonesia.

Dari hasil analisis data menggunakan metode *lexicon-based*, diketahui adanya dominasi sentimen positif yang sangat kuat dibandingkan dengan kelas netral dan negatif. Dominasi yang berlebihan ini menunjukkan bahwa meskipun kamus sentimen telah diperluas dengan penambahan kamus manual

untuk menyesuaikan konteks bahasa media sosial, metode *lexicon* masih belum mampu memprediksi kelas sentimen secara akurat. Hasil evaluasi menunjukkan bahwa akurasi dan presisi metode ini relatif rendah. Hal ini disebabkan oleh sifat dasar pendekatan *lexicon* yang hanya mengandalkan pencocokan kata dengan kamus sentimen tanpa mempertimbangkan konteks kalimat secara menyeluruh. Misalnya, kalimat yang mengandung kata positif tetapi digunakan dalam konteks sarkasme atau ironi akan tetap diklasifikasikan sebagai positif. Selain itu, banyaknya kata informal, singkatan, dan campuran bahasa Indonesia-Inggris di media sosial juga berkontribusi terhadap penurunan akurasi karena tidak semuanya terdaftar dalam kamus sentimen. Contoh, pada kalimat berikut:

- Iya pada mati kelaparan soalnya yg miskin, **terimakasih** pemerintah♥♥♥
- **Terimakasih** pemerintah sudah menciptakan pengangguran
- Hidup koruptor 😏😏😏 koruptor adalah **pahlawan** nasional indonesia. tak korupsi berarti bukan indonesia 😏😏😏
- **Mantap** pajak naik ...**bravo** prabowo...
- Menteri paling **hebat** di dunia dan akhirat, ekonomi selalu tumbuh, dibawah pemerintahan jokowi presiden **hebat**, ekonomi **meroket**, meroket ndasmu

```

Akurasi: 0.1502

=== Confusion Matrix (Tabel) ===
Prediksi      negatif  netral  positif
Kelas Sebenarnya
negatif          2628    1005    14509
netral             26         1         204
positif             0         2         155
    
```

Gambar 6. Hasil Akurasi, *Precision* dan *Recall* Lexicon pada data *twee*

```

Akurasi: 0.0387

=== Confusion Matrix (Tabel) ===
Prediksi      negatif  netral  positif
Kelas Sebenarnya
negatif          6464    9328    214313
netral              4         11     1073
positif              3         14     2572
    
```

Gambar 7. Hasil Akurasi, *Precision* dan *Recall* Lexicon pada data Youtube

Akurasi yang dihasilkan dari pendekatan *lexicon* untuk data Twitter dan YouTube menunjukkan nilai yang sangat rendah, masing-masing 15,02% dan 3,87%. Nilai ini menggambarkan bahwa model *lexicon* gagal mengklasifikasikan sebagian besar data dengan benar. Berdasarkan *confusion matrix* pada Gambar 6 dan 7, mayoritas data dari kedua *platform* diklasifikasikan sebagai sentimen positif, bahkan untuk teks yang sebenarnya bersifat negatif atau netral. Dengan kata lain, model *lexicon* cenderung *overestimate* terhadap kelas positif, sehingga menurunkan tingkat *precision* dan *recall* pada kelas lain. Hal ini mengindikasikan bahwa meskipun secara visual hasil distribusi sentimen tampak positif, kesimpulan tersebut tidak sepenuhnya dapat diandalkan secara statistik karena bias klasifikasi tinggi.

Untuk mengatasi keterbatasan tersebut, analisis kemudian dilanjutkan menggunakan pendekatan berbasis pembelajaran mesin, yaitu *Support Vector Machine* (SVM). Model SVM dilatih menggunakan data hasil labeling yang seimbang agar dapat membedakan sentimen dengan lebih baik.

## b. Hasil Analisis Data Menggunakan SVM

Pada tahap awal tanpa *balancing*, akurasi model masih rendah karena distribusi data yang timpang, tetapi setelah dilakukan penyesuaian kelas, hasilnya meningkat secara signifikan. Gambar 8 dan 9 merupakan nilai akurasi, *precision* dan *recall* metode SVM pada data Tweet dan Youtube. Pada data *tweet*, nilai *precision* dan *recall* kelas negatif mencapai 0,98, menunjukkan bahwa hampir seluruh *tweet* negatif dikenali secara tepat oleh model. Namun, performa pada kelas positif dan netral relatif lebih rendah, dengan *precision* dan *recall* yaitu 0,46. Artinya, sebagian *tweet* yang sebenarnya netral atau

positif masih sering diklasifikasikan sebagai kelas lain. Kondisi ini menunjukkan bahwa model masih mengalami kesulitan dalam membedakan ekspresi emosional yang halus atau ambigu, yang umum terjadi dalam teks pendek seperti tweet. Nilai akurasi keseluruhan mencapai 98%, tetapi sebagian besar akurasi tersebut didorong oleh dominasi kelas negatif yang lebih mudah dikenali. Komentar dengan nuansa positif ringan atau netral sering kali terklasifikasi ke dalam kelas lain.

Akurasi: 0.9800323799244468				
	precision	recall	f1-score	support
negatif	0.99	0.99	0.99	3628
netral	0.46	0.48	0.47	46
positif	0.46	0.38	0.41	32
accuracy			0.98	3706
macro avg	0.64	0.61	0.62	3706
weighted avg	0.98	0.98	0.98	3706

Gambar 8. Hasil Akurasi, Precision dan Recall SVM pada data tweet

Akurasi: 0.967470111427166				
	precision	recall	f1-score	support
negatif	0.99	0.98	0.98	46021
netral	0.05	0.12	0.07	218
positif	0.29	0.53	0.38	518
accuracy			0.97	46757
macro avg	0.45	0.54	0.48	46757
weighted avg	0.98	0.97	0.97	46757

Gambar 9. Hasil Akurasi, Precision dan Recall SVM pada data Youtube

Sementara itu, pada data YouTube, model SVM menghasilkan akurasi 96,7% dengan precision tinggi pada kelas negatif (0,99). Meskipun demikian, kelas positif dan netral tetap menjadi kategori dengan kinerja rendah karena jumlah datanya relatif sedikit dibandingkan kelas negatif. Hasil *confusion matrix* memperlihatkan bahwa sebagian besar data negatif terklasifikasi dengan benar, sedangkan kesalahan klasifikasi terbesar terjadi antara kelas positif dan netral.

Kondisi ini menunjukkan bahwa meskipun nilai akurasi model tinggi (98% untuk Twitter dan 96,7% untuk YouTube), distribusi prediksi antar kelas belum sepenuhnya seimbang. Model cenderung lebih sensitif terhadap ekspresi negatif yang eksplisit dibandingkan terhadap opini positif atau netral. Hal ini dapat disebabkan oleh karakteristik data yang sangat tidak seimbang serta kompleksitas bahasa informal pada media sosial. Dengan demikian, *confusion matrix* memberikan gambaran yang lebih mendalam dibandingkan sekadar melihat nilai akurasi tunggal, karena mampu mengungkap ketimpangan performa model antar kelas sentimen serta potensi bias yang muncul akibat distribusi data yang tidak seimbang.

Secara keseluruhan, perbandingan antara pendekatan *lexicon* dan SVM menunjukkan bahwa metode *lexicon* memiliki keunggulan dalam kesederhanaan dan interpretabilitas, namun kurang mampu menangkap konteks bahasa alami yang kompleks di media sosial. Sebaliknya, model SVM mampu memberikan hasil yang jauh lebih akurat dan stabil. Hal ini menegaskan pentingnya penggunaan teknik pembelajaran mesin dalam analisis sentimen berskala besar, khususnya ketika menghadapi data dengan distribusi yang tidak seimbang.

Dari hasil tersebut, dapat disimpulkan bahwa pendekatan berbasis SVM lebih berhasil memberikan hasil yang lebih realistis dan mendekati kondisi sebenarnya dibandingkan *lexicon*, sekaligus memperkuat validitas analisis sentimen lintas *platform* yang dilakukan dalam penelitian ini. Temuan ini konsisten dengan penelitian Singgalen, Y.A [10] yang menyatakan pendekatan *lexicon* lebih mudah digunakan tetapi sensitif terhadap ambiguitas konteks, sedangkan akurasi SVM lebih baik tetapi memerlukan data pelatihan berkualitas. Selain itu, hasil ini juga sejalan dengan penelitian Ratnaswari,

*et al.* [14] yang menemukan bahwa SVM lebih baik dalam mengidentifikasi sentimen pada data media sosial. Temuan penelitian ini menguatkan teori bahwa model *machine learning* lebih adaptif terhadap variasi bahasa digital dibandingkan metode berbasis kamus yang bersifat statis.

Dengan memahami pola sentimen ini, perusahaan asuransi dapat merancang strategi komunikasi yang lebih tepat sasaran serta mengembangkan produk berbasis teknologi yang mampu menjawab kekhawatiran publik. Hasil analisis sentimen ini dapat dijadikan dasar empiris dalam merancang inovasi produk dan layanan asuransi yang berorientasi pada kebutuhan serta persepsi masyarakat, sekaligus memperkuat upaya peningkatan literasi dan inklusi keuangan di Indonesia. Selain itu, dengan memahami sentimen masyarakat, perusahaan asuransi dapat merancang strategi komunikasi yang lebih tepat sasaran serta mengembangkan produk berbasis teknologi yang mampu menjawab kekhawatiran publik.

#### 4. KESIMPULAN

Analisis sentimen pada media sosial Twitter dan YouTube menunjukkan bahwa persepsi publik terhadap isu ekonomi dan asuransi menunjukkan pola yang berbeda antarplatform namun memiliki kecenderungan umum yang serupa. Jumlah data Youtube jauh lebih besar daripada Twitter, hal ini menunjukkan bahwa pengguna Youtube lebih aktif mengekspresikan opini dan tanggapan tentang asuransi. Metode *lexicon* masih belum mampu memprediksi sentimen secara akurat, meskipun kamus sentimen telah diperluas dengan penambahan kamus manual untuk menyesuaikan konteks bahasa media sosial. Hal ini ditunjukkan oleh nilai akurasi dan presisi yang rendah. Sebaliknya, hasil analisis dengan menggunakan SVM menunjukkan performa yang jauh lebih baik, dengan akurasi 98% untuk data Twitter dan 96,7% untuk data YouTube, serta mampu mengenali sentimen negatif dengan presisi tinggi meski masih sulit membedakan antara sentimen positif dan netral.

Hasil analisis sentimen ini tidak hanya menggambarkan persepsi publik terhadap isu ekonomi dan asuransi, tetapi juga memberikan gambaran penting bagi pengembangan inovasi produk asuransi di Indonesia. Dominasi sentimen negatif yang ditemukan, terutama yang berkaitan dengan ketidakpercayaan terhadap klaim dan kerumitan prosedur, dapat menjadi masukan berharga bagi perusahaan asuransi untuk melakukan inovasi pada aspek layanan pelanggan, digitalisasi proses klaim, dan kejelasan manfaat produk. Sebaliknya, sentimen positif menunjukkan adanya penerimaan masyarakat yang cukup baik terhadap topik ekonomi dan asuransi. Terutama ketika dikaitkan dengan edukasi dan literasi keuangan. Dengan memahami pola sentimen ini, perusahaan asuransi dapat merancang strategi komunikasi yang lebih tepat sasaran serta mengembangkan produk berbasis teknologi yang mampu menjawab kekhawatiran publik.

#### Daftar Pustaka

- [1] CNBC Indonesia, "Ramalan Bank Dunia: Negara ini bisa jatuh ke jurang krisis pada 2025. CNBC Indonesia," 2025. <https://www.cnbcindonesia.com/news/20250621155135-4-642789/ramalan-bank-dunia-negara-ini-bisa-jatuh-ke-jurang-krisis-pada-2025>.
- [2] Badan Pusat Statistik, "Keadaan ketenagakerjaan Indonesia Februari 2025 (Berita Resmi Statistik No. 44/05/Th.XXVIII)," 2025. <https://www.bps.go.id/id/pressrelease/2025/05/05/2432/tingkat-pengangguran-terbuka--tpt--sebesar-4-76-persen--rata-rata-upah-buruh-sebesar-3-09-juta-rupiah-.html>.
- [3] Badan Pusat Statistik, "Inflasi terjadi pada Juni 2022 sebesar 0,61 persen: Inflasi tertinggi terjadi di Gunungsitoli sebesar 2,72 persen," 2022. <https://www.bps.go.id/id/pressrelease/2022/07/01/1862/inflasi-terjadi-pada-juni-2022-sebesar-0-61-persen--inflasi-tertinggi-terjadi-di-gunungsitoli-sebesar-2-72-persen-.html>.
- [4] M. Eling and M. Lehmann, "The impact of digitalization on the insurance value chain and the insurability of risks," *Geneva Pap. risk Insur. Pract.*, vol. 43, no. 3, pp. 359–396, 2018.
- [5] E. Kırkbeşoğlu, "How Digitalization Affects Insurance Companies: Current Status And Future Expectations," *Sivas Soft Bilisim Proje Danismanlik Egit. Sanayi ve Ticaret Ltd. Sirketi*, vol. 1, pp. 32–43, 2021, doi: 10.52898/ijif.2021.4.
- [6] K. Kristianto, A. B. Ramadhan, and F. D. Marsetyo, "Media sosial dan connective action: Studi kasus penggunaan Twitter sebagai ruang solidaritas selama pandemi COVID-19," *J. Soc. Dev. Stud.*, vol. 2, no. 1, pp. 1–13, 2021.

- [7] S. Alim and A. F. Dharma, "Youtube Sebagai Ruang Publik Alternatif Bagi Anak Muda," *Expo. J. Ilmu Komun.*, vol. 4, no. 1, pp. 1–13, 2021.
- [8] Rumah Media, "Daftar jumlah pengguna media sosial Indonesia terbaru: Siapa yang paling banyak?," 2025. <https://www.rumahmedia.com/insights/daftar-jumlah-pengguna-media-sosial-indonesia-terbaru-siapa-yang-paling-banyak>.
- [9] S. Stieglitz, M. Mirbabaie, B. Ross, and C. Neuberger, "Social media analytics—Challenges in topic discovery, data collection, and data preparation," *Int. J. Inf. Manage.*, vol. 39, pp. 156–168, 2018.
- [10] Yerik Afrianto Singgalen, "Pemilihan Metode dan Algoritma dalam Analisis Sentimen di Media Sosial: Systematic Literature Review," *Journal of Information Systems and Informatics*, Vol. 3 No. 2, 2021.
- [11] I. Srba *et al.*, "Auditing YouTube's recommendation algorithm for misinformation filter bubbles," *ACM Trans. Recomm. Syst.*, vol. 1, no. 1, pp. 1–33, 2023.
- [12] Janková, Z., "Critical review of text mining and sentiment analysis for stock market prediction," *Journal of Business Economics and Management*, vol. 24, no. 1, 2023.
- [13] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdiscip. Rev. data Min. Knowl. Discov.*, vol. 8, no. 4, p. e1253, 2018.
- [14] S. Ratnaswari, N. C. Wibowo, and D. S. Y. Kartika, "Analisis sentimen menggunakan metode lexicon-based dan Support Vector Machine pada Presiden dan Wakil Presiden Indonesia periode 2024–2029," Universitas Pembangunan Nasional Veteran Jawa Timur, 2024.
- [15] M. Ahmad, M. F. Octaviansyah, A. Kardiana, and K. F. Prasetyo, "Sentiment Analysis System of Indonesian tweets using lexicon and naïve Bayes approach," in *2019 Fourth International Conference on Informatics and Computing (ICIC)*, 2019, pp. 1–5.
- [16] B. S. Rintyarna, W. Suharso, A. Sanosra, dan I. S. Windiarti, "Ina-SASet: Dataset for Developing Indonesian Sentiment Lexicon for Extracting Consumer Preference Based on Fine-Grained Sentiment Analysis Technique," *International Journal of Intelligent Systems and Applications in Engineering (IJISAE)*, vol. 12, no. 23s, pp. 295–300, 2024.
- [17] B. Pamungkas, M. E. Purbaya, and D. J. AK, "Analisis Sentimen Twitter Menggunakan Metode Support Vector Machine (SVM) pada Kasus Benih Lobster 2020," *J. Informatics Inf. Syst. Softw. Eng. Appl.*, vol. 3, no. 2, pp. 10–20, 2021.
- [18] N. Istiqomah and F. Novika, "Sentiment Analysis Penyedia layanan Asuransi dari Media Sosial Twitter," *J. Tekno Kompak*, vol. 18, no. 1, pp. 77–89, 2024.
- [19] N. Istiqomah and F. Novika, "Extracting Post-Disaster Health Impact Information from News Reports Using Named Entity Recognition," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 7, no. 2, pp. 163–172, 2025.
- [20] N. Istiqomah, F. Novika, and S. Tinggi Manajemen Asuransi Trisakti, "Perbandingan Kinerja Model NER IndoBERT dan IndoLEM dalam Ekstraksi Informasi Kesehatan Pascabencana dari Berita Daring di Indonesia Comparative Performance of IndoBERT and IndoLEM Baseline Models for Post-Disaster Health Information Extraction from Indonesian Online News," *J. Comput. Sci. Informatics Eng.*, vol. 04, no. 3, pp. 158–174, 2025, [Online]. Available: <http://creativecommons.org/licenses/by-sa/4.0/>.



*ZONasi: Jurnal Sistem Informasi*

Is licensed under a [Creative Commons Attribution International \(CC BY-SA 4.0\)](http://creativecommons.org/licenses/by-sa/4.0/)