BAB II
TINJAUAN PUSTAKA

II.1 Analisis Sentimen

Analisis sentimen merupakan salah satu bidang pengolahan data teks yang berfokus pada kajian opini, sentimen, penilaian, perilaku, serta emosi individu, sehingga hasilnya dapat dimanfaatkan sebagai bahan evaluasi (Satya Marga *et al*., 2021). Analisis sentimen, yang juga dikenal dengan istilah *Opinion Mining*, merupakan metode yang digunakan untuk mengidentifikasi dan menganalisis opini dari kumpulan teks yang diperoleh melalui berbagai sumber, sehingga peneliti dapat memahami kecenderungan perasaan terhadap suatu topik yang dibahas. Proses ini sangat berkaitan dengan penerapan teknik dalam *Natural Language Processing* (NLP). NLP sendiri merupakan cabang dari kecerdasan buatan yang dirancang agar komputer mampu memahami dan berinteraksi dengan manusia menggunakan bahasa alami. Tujuannya adalah memahami dan mengambil informasi dari data tidak terstruktur, seperti teks dalam berbagai bahasa, termasuk Bahasa Indonesia. NLP membantu komputer membaca, mendengar, memahami makna, menganalisis sentimen, dan menemukan informasi penting dari teks (Sarker *et al*., 2021).

Menurut Gouthami & Hegde (2021), Analisis sentimen merupakan salah satu metode dalam analitik teks yang digunakan untuk mengekstraksi opini dari pengguna media sosial serta mengelompokkan polaritas sentimen yang relevan dan dapat diterapkan pada berbagai bidang. Pada umumnya, analisis terhadap big data menjadi tantangan tersendiri karena dipengaruhi oleh karakteristik utama big data, seperti volume data yang besar, keberagaman data, kecepatan pertumbuhan data, variabilitas, serta tingkat keakuratan data.

Analisis sentimen memiliki beberapa jenis pendekatan. Salah satunya adalah *Fine-Grained Sentiment Analysis*, yaitu metode yang paling umum digunakan dengan fokus pada tingkat polaritas opini. Pendekatan ini mengelompokkan tanggapan ke dalam kategori positif, netral, dan negatif. Jenis berikutnya adalah *Intent Sentiment Analysis* yang bertujuan untuk mengidentifikasi

motivasi di balik pesan pengguna, sehingga dapat diketahui apakah pesan tersebut berupa saran, keluhan, pertanyaan, opini, ataupun bentuk apresiasi terhadap suatu produk maupun layanan. Selain itu, terdapat *Aspect-Based Sentiment Analysis* yang menitikberatkan pada analisis aspek-aspek tertentu dari sebuah produk atau layanan secara lebih spesifik, sehingga mampu memberikan pemahaman yang lebih mendalam terhadap elemen yang menjadi perhatian pengguna. (Namira Miftah, 2023). Fokus penelitian ini adalah tipe *Fine-Grained Sentiment Analysis* yaitu mengelompokkan teks atau kalimat berupa sentimen positif, netral atau negatif, sehingga hasil dari pengelompokkan yang diperoleh dapat mengetahui respon dari masyarakat terhadap program efisiensi anggaran.

Metode analisis sentimen terdiri dari: (1) *Machine Learning Based Approach* yaitu pendekatan berbasis *machine learning* yang melatih *classifier* dari data yang diberi label secara manual. (2) *Lexicon Based Approach* yaitu pendekatan yang menggunakan leksikon sentimen untuk mendeskripsikan polaritas (positif, negatif, dan netral) dari sebuah isi teks. (3) *Hybrid Approach* yaitu pendekatan yang menggabungkan *machine learning* dan metode berbasis leksikon (Sadia A *et al*., 2018).

II.2 X (Twitter)

Twitter diluncurkan pada bulan Juli 2006 sebagai *platform microblogging* yaitu tulisan pendek yang disiarkan ke publik dan dirancang untuk memberi tahu orang lain tentang "status" pengguna. Twitter banyak mengambil inspirasi dari pesan teks SMS dan memprioritaskan antarmuka yang ramping seperti teks perpesanan. Tweet dibatasi hingga 140 karakter kemudian diperpanjang menjadi 280 karakter pada bulan November 2017. Tidak seperti SMS, pengguna dapat terhubung dengan individu secara lebih luas. Pengguna dapat berkomunikasi dengan pengikut dan publik dengan menggunakan simbol *at* (@) untuk mengarahkan percakapan ke orang tertentu dan dapat menyertakan simbol *hashtag* (#) untuk menautkan postingan ke kumpulan tweet yang lebih besar di bawah hashtag tertentu. Twitter mempelopori tagar untuk mengatur wacana, dan *platform* lain mengikutinya (Murthy, 2025).

Peran Twitter sebagai platform data untuk analisis opini publik telah berkembang pesat dalam beberapa tahun terakhir, berfungsi sebagai alat bagi para peneliti untuk menganalisis dialog politik dan sentimen publik. Alasan memilih Twitter sebagai platform untuk penelitian kebijakan publik dan politik sangat beragam. Pertama, Twitter memberikan kesempatan unik untuk berinteraksi secara *real-time* dengan demografi yang aktif secara politik, terutama generasi muda (Murić *et al*., 2021). Demografi ini ditandai dengan tingkat keterlibatan politik yang lebih tinggi dibandingkan populasi umum, memberikan gambaran tentang sentimen publik yang sedang berkembang dan dapat mempengaruhi pembentukan kebijakan. Selain itu, sifat platform yang relatif terbuka memungkinkan peneliti mengakses pandangan populasi yang beragam, meningkatkan validitas analisis mereka (Payson *et al*., 2022). Para akademisi telah mendokumentasikan bahwa Twitter bertindak sebagai "media elit" dalam beberapa konteks, di mana figur politik dan pejabat publik berinteraksi secara aktif, membentuk diskursus dan momentum untuk berbagai isu (Kotkaniemi *et al*., 2024). Oleh karena itu, jangkauan luas dan kecepatan Twitter memainkan peran sentral dalam penelitian politik kontemporer.

II.3 Efisiensi Anggaran

Efisiensi mencerminkan suatu keberhasilan seseorang atau organisasi dalam usaha melakukan kebijakan yang diambil. Efisiensi digunakan sebagai tolak ukur (indikator) untuk membandingkan input/modal dengan output/hasil (Putra, 2021). Dikatakan efisien, jika input/modal dapat mendapatkan output/hasil yang maksimal, dan dikatakan tidak efisien, jika input/modal melebihi dari jumlah yang diperlukan (Antohi *et al*., 2022). Penelitian yang dilakukan (Abqa *et al*., 2024), bahwa prinsip efisiensi dituangkan dalam beberapa peraturan hukum yang berlaku. Artinya harus ada penghematan dan kehati-hatian dalam mengeluarkan anggaran, terlebih jika anggaran yang dikeluarkan dalam jumlah yang sangat besar.

Untuk mengukur efisiensi anggaran, indikator dan parameter yang spesifik sangat penting. Metrik ini dapat mencakup evaluasi efektivitas biaya, analisis varians anggaran, dan metrik berbasis kinerja yang menilai apakah hasil yang dicapai sesuai dengan anggaran yang dialokasikan (Nawawi dkk., 2023). Misalnya,

penerapan prinsip-prinsip *"Value for Money"* (VfM) yang melibatkan perbandingan antara pengeluaran dengan manfaat sebenarnya yang diperoleh dari investasi publik, sehingga memberikan gambaran yang komprehensif mengenai efektivitas anggaran (Sam *et al*., 2024).

Upaya meningkatkan efisiensi anggaran birokrasi merupakan tantangan yang kompleks, namun memiliki peran penting dalam mendukung pembangunan serta peningkatan kualitas pelayanan publik. Berbagai kendala yang dihadapi meliputi tingginya biaya operasional, rendahnya tingkat transparansi dan akuntabilitas, ketidakpastian dalam penganggaran, serta kebijakan yang belum berjalan secara optimal. Untuk mengatasi permasalahan tersebut, beberapa langkah strategis dapat diterapkan, antara lain optimalisasi pemanfaatan teknologi, peningkatan transparansi pengelolaan anggaran, penyusunan perencanaan anggaran jangka panjang, evaluasi terhadap program dan kebijakan, reformasi birokrasi, serta penguatan kerja sama dengan sektor swasta maupun Lembaga Swadaya Masyarakat (LSM). Melalui penerapan berbagai strategi tersebut, pemerintah diharapkan mampu meningkatkan efisiensi penggunaan anggaran, meminimalkan pemborosan, serta memperbaiki kualitas pelayanan publik kepada masyarakat (Matondang dkk., 2024).

II.4 Algoritma *Naïve Bayes*

Algoritma *Naïve Bayes* telah menjadi lazim dalam tugas-tugas klasifikasi teks karena strukturnya yang jelas dan kesederhanaannya yang didasarkan pada *Teorema Bayes*. Prinsip dasar dari *Naïve Bayes* adalah mengasumsikan independensi antara fitur-fitur yang diberi label kelas. Pendekatan probabilistik ini menghitung probabilitas posterior kelas berdasarkan pengetahuan sebelumnya dan kemungkinan fitur yang muncul dalam setiap kondisi kelas. Secara khusus, algoritma ini menerapkan perhitungan sederhana yang melibatkan perkalian probabilitas kemunculan setiap fitur yang diberikan kelas, dibagi dengan faktor normalisasi untuk memastikan bahwa probabilitasnya berjumlah satu (Zulfikar dkk., 2023). Khususnya, ada berbagai variasi *Naïve Bayes*, seperti *Bernoulli* dan *Multinomial Naïve Bayes* menjadi sangat efektif dalam skenario di mana data input

didasarkan pada jumlah atau frekuensi yang khas dalam aplikasi teks (Ridho dkk., 2022).

Dalam klasifikasi teks, *Naïve Bayes* memiliki kelebihan dan kekurangan. Salah satu kelebihan utamanya adalah efisiensi komputasi yang memungkinkan klasifikasi cepat dalam aplikasi waktu nyata dengan set data yang besar. Penelitian menunjukkan bahwa klasifikasi *Naïve Bayes* dapat mencapai tingkat akurasi yang kompetitif dibandingkan dengan algoritma yang lebih kompleks seperti *Support Vector Machine* (SVM) atau Jaringan Saraf Tiruan (Pradana & Sugiharti, 2023). Namun, kelemahan utama *Naïve Bayes* adalah asumsinya bahwa fitur-fitur independen yang dapat menyebabkan ketidakakuratan dalam aplikasi dunia nyata di mana fitur-fitur mungkin saling terkait. Selain itu, kinerja dapat menurun pada dataset yang tidak seimbang atau ketika distribusi kelas sangat tidak merata (Munawaroh & Alamsyah, 2023).

Dalam algoritma ini perhitungan ditunjukkan dengan menggunakan persamaan II.1 (Jannah & Kusnawi, 2024) :

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \qquad \text{..... (1)}$$

Dimana:

X : *Class* data yang belum diketahui.

H : Hipotesis merupakan suatu kelas spesifik

P(H|X): Probabilitas hipotesis nilai H berdasarkan kondisi nilai X (posteriori probabilitas)

P(H) : Probabilitas hipotesis H (prior probabilitas)

P(X|H) : Probabilitas nalai X dari kondisi pada hipotesis nilai H

P(X) : Probabilitas nilai X

Dalam penerapan klasifikasi teks, varian yang paling umum digunakan adalah *Multinomial Naïve Bayes,* karena mampu memodelkan frekuensi kemunculan kata pada dokumen dengan baik. Pada model ini terdapat parameter *alpha* (α) yang berfungsi sebagai teknik *smoothing* untuk mengatasi kemungkinan

munculnya probabilitas nol ketika suatu kata tidak ditemukan pada data pelatihan dalam kelas tertentu. Penerapan parameter *smoothing* tersebut membantu menjaga stabilitas distribusi probabilitas sehingga model dapat melakukan proses klasifikasi dengan lebih konsisten (Mursyid *et al.*, 2025).

II.5 Algoritma *Support Vector Machine* (SVM)

Teori dasar dan prinsip kerja *Support Vector Machine* (SVM) berakar pada konsep teori pembelajaran statistik, khususnya prinsip minimalisasi risiko struktural, yang bertujuan untuk menyeimbangkan kompleksitas model dan kemampuan generalisasi (Chen, 2024). SVM beroperasi dengan menemukan *hyperplane* yang paling baik memisahkan titik-titik data dari kelas yang berbeda dalam ruang berdimensi tinggi, menekankan pentingnya support vector, atau titik-titik data yang paling dekat dengan *hyperplane* yang mempengaruhi posisinya. Metode ini meminimalkan kesalahan klasifikasi sambil memaksimalkan margin antar kelas, meningkatkan ketahanan model terhadap *overfitting*. Efektivitas SVM telah diakui secara luas dalam berbagai penelitian yang menunjukkan kemampuannya dalam berbagai aplikasi, termasuk klasifikasi biomedis (Li *et al.*, 2021).

Fungsi kernel memainkan peran penting dalam kinerja SVM dengan mentransformasi data yang tidak dapat dipisahkan secara linier ke dalam ruang berdimensi lebih tinggi di mana pemisah linier dapat diidentifikasi (Alkesaiberi *et al*., 2022). Berbagai fungsi kernel, seperti *kernel linear, polinomial*, dan *radial basis function (RBF),* secara signifikan mempengaruhi hasil klasifikasi. Pemilihan kernel mempengaruhi kompleksitas model dan kemampuannya untuk menggeneralisasi data yang tidak terlihat (Wang *et al*., 2019).

Support Vector Machine menunjukkan beberapa keunggulan dalam bidang pemrosesan data teks, sehingga sangat cocok untuk tugas-tugas seperti analisis sentimen dan klasifikasi dokumen. Kemampuannya untuk mengelola ruang dimensi tinggi secara efisien memungkinkan SVM bekerja dengan baik bahkan ketika jumlah fitur melebihi jumlah sampel, sebuah skenario yang umum terjadi pada data teks (Bachri *et al*., 2019). Selain itu, ketahanan SVM terhadap *overfitting*,

terutama disebabkan oleh prinsip minimalisasi risiko struktural, memungkinkan kinerja yang andal bahkan dalam kasus data pelatihan yang terbatas (Zheng *et al*., 2021). Penelitian telah menunjukkan bahwa SVM secara konsisten mengungguli pengklasifikasi tradisional termasuk pengklasifikasi linier dan *Naïve Bayes* dalam hal akurasi dan kemampuan generalisasi dalam berbagai tugas berbasis teks (Fairclough *et al*., 2021).

Hyperplane tersusun atas parameter berupa vektor bobot yang direpresentasikan sebagai vektor kolom (w), serta parameter posisi bidang terhadap titik pusat koordinat (b), yangdinyatakan dalam persamaan II.2 (Amini, 2023) :

$$f(x) = wx + b$$

……..……..… (2)

Dimana:

w = vektor yang tegak lurus dengan *hyperplane*

x = data

b = nilai nias

f(x) = nilai *hyperplane*

Performa SVM dipengaruhi oleh beberapa *hyperparameter* penting yaitu parameter *C* (cost), parameter *gamma* (γ), dan parameter *degree* (Mursyid *et al*., 2025). Parameter utama yang digunakan adalah C, berperan sebagai parameter regularisasi dalam mengatur keseimbangan antara lebar margin dan tingkat kesalahan klasifikasi pada data pelatihan. Nilai C yang tinggi menyebabkan model lebih sensitif terhadap kesalahan klasifikasi, sedangkan nilai C yang rendah cenderung menghasilkan margin yang lebih luas meskipun masih memungkinkan terjadinya beberapa kesalahan prediksi.

Pada kernel *non-linear* juga terdapat parameter tambahan. Pada kernel *Radial Basis Function* (RBF) digunakan parameter *gamma* (γ) yang menentukan sejauh mana pengaruh satu data latih terhadap pembentukan batas keputusan model. Nilai gamma yang besar menghasilkan batas keputusan yang lebih kompleks, sedangkan nilai yang kecil menghasilkan pemisahan yang lebih halus. Selain itu, pada kernel *polynomial* terdapat parameter *degree* yang mengatur tingkat

kompleksitas fungsi polinomial dalam memetakan data ke ruang fitur yang lebih tinggi. Pemilihan nilai parameter yang tepat menjadi faktor penting dalam memperoleh performa klasifikasi yang optimal.

II.6 *Particle Swarm Optimization* (PSO)

Optimalisasi algoritma *machine learning*, khususnya dalam konteks penyetelan parameter merupakan area penelitian penting yang telah mendapatkan kontribusi signifikan dari *Particle Swarm Optimization* (PSO). PSO adalah algoritma metaheuristik yang terinspirasi oleh perilaku sosial burung yang berkelompok atau kawanan ikan yang beroperasi dengan mempertahankan populasi kandidat solusi (partikel) yang mengeksplorasi ruang solusi dan berbagi informasi tentang posisi dan keberhasilan mereka. Algoritma ini mengevaluasi solusi potensial melalui serangkaian pembaruan berdasarkan pengetahuan individu dan kolektif secara iteratif meningkatkan kinerja hingga mencapai tingkat optimasi yang optimal (Shami *et al*., 2022).

Dalam bidang tuning algoritma *machine learning*, PSO telah memantapkan dirinya sebagai teknik yang kuat untuk mengoptimalkan hiperparameter sehingga meningkatkan kinerja model di berbagai tugas. Dengan menerapkan PSO, para peneliti dapat secara efektif menemukan pengaturan optimal untuk parameter algoritma yang penting seperti yang ada di pengklasifikasi *Support Vector Machine* dan *Naïve Bayes* (Manik dkk., 2022). Secara khusus, kemampuan algoritma yang melekat dalam menavigasi ruang pencarian yang kompleks memungkinkannya untuk secara efektif menyeimbangkan antara eksplorasi konfigurasi baru dan eksploitasi konfigurasi yang baik yang telah diketahui yang mengarah pada peningkatan akurasi dalam model prediktif (Xue *et al*., 2019). Penyetelan parameter menggunakan PSO ini berkontribusi secara signifikan dalam mengurangi waktu dan sumber daya komputasi yang diperlukan untuk mencapai model berkinerja tinggi yang sangat penting dalam aplikasi dunia nyata (Arukonda & Cheruku, 2023).

II.7 *Text Mining*

Text mining yang juga dikenal sebagai analisis teks merupakan metode kunci dalam analisis data teks di berbagai bidang dengan memanfaatkan teknik komputasi untuk mengekstrak wawasan yang bermakna dari data tidak terstruktur. *Text mining* mencakup proses mengubah dataset teks yang besar menjadi informasi terstruktur dengan menggunakan algoritma canggih dan teknologi kecerdasan buatan atau AI (Gnanavel *et al*., 2022). Definisi ini menggambarkan esensi *text mining* bukan sekadar teknik manipulasi data, melainkan pendekatan transformatif yang memfasilitasi penemuan pola, tren, dan hubungan dalam informasi teks yang krusial bagi proses pengambilan keputusan di bidang penelitian dan industri (Carey *et al*., 2022).

Literatur terkini secara konsisten menyoroti keunggulan *text mining* dalam berbagai bidang aplikasi. Misalnya, pada konteks pendidikan, *text mining* memfasilitasi evaluasi metode pengajaran melalui agregasi umpan balik siswa yang memungkinkan sintesis dan penilaian data pedagogis yang beragam (Okoye *et al*., 2023). Selain itu, penerapan *text mining* di lingkungan korporat menunjukkan efektivitasnya dalam menganalisis sentimen dan keterlibatan pelanggan melalui ulasan dan analisis media sosial, sehingga memberikan wawasan yang dapat ditindaklanjuti yang mencerminkan perilaku konsumen (Li *et al*., 2023).

Metodologi yang digunakan dalam *text mining* melibatkan berbagai teknik seperti *Natural Language Processing* (NLP), analisis sentimen, dan *machine learning*. Teknik-teknik ini memungkinkan peneliti untuk mengotomatisasi ekstraksi dan kategorisasi informasi dari volume teks yang besar yang semakin penting di era big data (Souza *et al*., 2024). Misalnya, teknik seperti *Latent Dirichlet allocation* (LDA) telah digunakan untuk mengklasifikasikan data teks ke dalam topik yang kohesif, meningkatkan pemahaman tentang tema-tema utama yang terdapat dalam dataset yang luas (Park *et al*., 2021).

II.8 *Text Preprocessing*

Text preprocessing merupakan langkah awal dalam pengolahan teks yang bertujuan untuk mempersiapkan teks menjadi bentuk data yang siap diolah lebih

lanjut. Dalam analisis sentimen tahap *preprocessing* menjadi sangat penting karena bertujuan untuk membersihkan dan menormalkan data tekstual untuk meningkatkan kinerja algoritma *machine learning*. Di ranah twitter, dimana fitur linguistik sering kali tidak terstruktur dan mengandung noise seperti singkatan, bahasa gaul, dan emoji sehingga teknik pemrosesan awal yang efisien sangat penting. Berbagai penelitian menekankan bahwa teknik *preprocessing* yang berbeda dapat secara signifikan mempengaruhi hasil model analisis sentimen, menganjurkan pendekatan sistematis untuk mengevaluasi dan mengimplementasikan teknik-teknik tersebut untuk meningkatkan akurasi dan keandalan model (Chandrasekaran *et al*., 2021).

Dalam *preprocessing* terdiri atas beberapa proses yaitu *cleaning, case folding, normalization, tokenization, stopword removal* dan *stemming*. Secara umum proses yang dilakukan adalah sebagai berikut (Kulkarni & Shivananda, 2019):

1. *Cleaning*

   Proses *cleaning* merupakan tahap awal dalam pra-pemrosesan data, khususnya untuk data teks. Tahap ini memiliki peran penting guna memastikan bahwa data yang akan diproses sudah bersih dan minim gangguan (*noisy*), sehingga dapat memudahkan sistem dalam melakukan pemrosesan tahap selanjutnya. Kegiatan pada tahap ini bisa mencakup penghapusan tanda baca (seperti: , . ? !), angka, karakter khusus non-ASCII (contoh: @, $, #, &), tautan URL, serta spasi yang berlebihan. Selain itu, pemilihan atribut atau fitur juga merupakan bagian dari proses ini

2. *Case Folding*

   Case folding adalah proses menyamakan bentuk huruf dalam sebuah dokumen, terutama untuk mendukung proses pencarian. Karena tidak semua dokumen memiliki konsistensi dalam penggunaan huruf kapital, maka *case folding* diperlukan untuk mengubah semua teks dalam dokumen menjadi format yang seragam, yaitu menggunakan huruf kecil atau *lowecase.*

3. *Normalization*

   Normalization merupakan proses untuk mengubah kata-kata tidak baku seperti slang atau singkatan menjadi bentuk kata yang sesuai dengan ejaan baku menurut Kamus Besar Bahasa Indonesia (KBBI). Normalisasi penting dilakukan terutama pada data yang diperoleh dari media sosial, dimana pengguna sering kali menggunakan bentuk singkatan atau penulisan yang tidak sesuai dengan kaidah. Proses ini bertujuan agar teks memiliki penulisan yang konsisten dan seragam.

4. *Tokenization*

   Tokenization adalah tahapan memecah teks yang awalnya berupa kalimat menjadi bagian-bagian yang terdiri dari kata-kata. Proses ini dimulai dengan menghapus tanda pemisah (*delimiter*) seperti simbol atau tanda baca yang terdapat dalam teks. *Tokenize* (juga dikenal sebagai analisis leksikal atau segmentasi kata) bertujuan untuk membagi teks menjadi unit-unit kecil yang disebut token, agar dapat dilakukan proses analisis lebih lanjut. Secara umum, *tokenization* mengacu pada pemisahan kata dari suatu kalimat dengan cara mengambil setiap token sebagai urutan karakter yang biasanya dipisahkan oleh spasi.

5. *Stopword Removal*

   Stopword removal adalah proses penyaringan kata-kata yang sering muncul dalam teks, namun dianggap tidak membawa makna penting dalam konteks analisis data. Kata-kata tersebut umumnya bersifat umum atau fungsional dan tidak memberikan kontribusi signifikan terhadap pemahaman isi, sehingga perlu dihapus oleh sistem. Contoh dari *stopword* meliputi kata “yang” “dan”, “ke”, “atau”, dan lain sebagainya.

6. *Stemming*

   Stemming merupakan proses untuk mengolah kata dengan tujuan memperoleh bentuk dasar dari kata yang telah mengalami penambahan imbuhan, dengan asumsi bahwa kata tersebut memiliki makna yang serupa dengan bentuk dasarnya. Proses ini menggunakan pendekatan morfologi berdasarkan struktur

bahasa Indonesia, yang mencakup imbuhan seperti awalan, akhiran, sisipan, maupun gabungan awalan dan akhiran.

II.9 *Lexicon Based*

Pendekatan *Lexicon-Based* dikenal sebagai metode klasik dan paling sederhana dalam ekstraksi sentimen dari teks atau ulasan pengguna. Teknik ini menggunakan kamus sentimen, yang memuat kata-kata dengan polaritas tertentu (positif, negatif, netral), untuk menganalisis makna emosional dari sebuah teks. Namun, banyak metode ini hanya menetapkan satu nilai polaritas per kata, tanpa mempertimbangkan nuansa emosi yang kompleks.

Kamus atau leksikon ini merupakan kumpulan kata yang umum digunakan untuk menyampaikan emosi tertentu, dan setiap kata diberi bobot sentimen. Pendekatan ini terbukti memiliki performa yang baik terutama pada analisis lintas domain karena fleksibilitasnya dalam memperbarui basis pengetahuan dengan menambahkan entri baru ke dalam kamus. Metode leksikal juga populer dalam analisis sentimen karena kemampuannya mengevaluasi orientasi semantik kata dan menghitung sentimen berdasarkan nilai yang ditetapkan untuk setiap istilah (Prakash & Aloysius, 2021).

Sebagai perbandingan, pendekatan *machine learning* memerlukan data latih berlabel, proses pemilihan fitur, pembentukan model, dan pengujian pada data baru. Sementara itu, metode *Lexicon-Based* hanya membutuhkan kamus dan aturan interpretasi untuk langsung menentukan sentimen dari teks yang dianalisis.

Kelebihan metode ini adalah ketahanannya dalam berbagai domain tanpa perlu banyak modifikasi. Selain itu, menerapkan kamus ke bahasa atau konteks baru dinilai lebih mudah dibandingkan dengan pelabelan ulang data. Dengan memanfaatkan struktur linguistik dalam teks, pendekatan ini menunjukkan kombinasi yang baik antara linguistik dan teknik komputasi.

II.10 *Python* dan *Google Colaboratory*

Bahasa pemrograman *Python* telah berkembang pesat selama bertahun-tahun dan kini menjadi salah satu bahasa pemrograman paling populer di dunia.

Sejarahnya bermula pada akhir tahun 1980-an, ketika Guido van Rossum merancang bahasa ini sebagai pengganti bahasa pemrograman ABC, dengan tujuan menciptakan bahasa yang menekankan keterbacaan kode dan kesederhanaan. Sejak dirilis secara publik pada tahun 1991, *Python* telah mengalami peningkatan dan perluasan yang signifikan, dengan komunitas yang kuat mengembangkan ribuan perpustakaan dan kerangka kerja di berbagai bidang, termasuk pengembangan web, analisis data, kecerdasan buatan, dan komputasi ilmiah (Nag, 2023).

Salah satu fitur menonjol *Python* adalah sintaksnya yang mudah dipelajari, yang sangat dipengaruhi oleh prinsip-prinsip keterbacaan dan kesederhanaan. Hal ini membuat *Python* dapat diakses oleh pemula sambil tetap cukup powerful untuk programmer tingkat lanjut. Sebagai bahasa pemrograman berorientasi objek, *Python* memungkinkan enkapsulasi fungsionalitas dalam objek, sehingga memudahkan pendekatan yang lebih terorganisir dalam pemrograman (Maharazu Mamman *et al*., 2023).

Google Colaboratory yang lebih dikenal sebagai *Colab* adalah *platform* berbasis *cloud* yang sangat penting untuk pemrograman *Python*, terutama populer dalam bidang *machine learning* dan analisis data. Keunggulan *platform* ini terletak pada integrasi sumber daya komputasi yang besar, fitur aksesibilitas, dan alat kolaboratif yang meningkatkan pengalaman pengguna serta peluang pendidikan dalam pemrograman dan ilmu data. Popularitas platform ini terutama berasal dari penyediaan akses gratis ke sumber daya komputasi yang besar, termasuk GPU seperti Nvidia Tesla K80 dan V100. Aksesibilitas ini memungkinkan pengguna untuk melakukan tugas komputasi intensif yang biasanya memerlukan investasi hardware yang signifikan (Z. Li, 2022).

Selain itu, sifat interaktif *Google Colab* mendukung kerja kolaboratif, memungkinkan beberapa pengguna untuk mengedit dan menjalankan kode secara real-time. Fitur ini terbukti sangat bermanfaat dalam lingkungan pendidikan, di mana instruktur dapat mendemonstrasikan praktik pemrograman sementara siswa dapat mengikuti tanpa memerlukan pengaturan lokal (Vera & Carpio, 2024).

Keunggulan utama *Google Colab* adalah arsitekturnya yang ramah pengguna, yang mempermudah proses onboarding bagi pengguna yang tidak

memiliki latar belakang pemrograman yang luas. Menurut Mclean, integrasi alat kecerdasan buatan canggih dalam *Google Colab* menurunkan hambatan masuk untuk analisis tingkat lanjut di bidang seperti biologi komputasional (McLean, 2024). Penelitian menunjukkan bahwa penggunaan *Colab* untuk tujuan ini tidak hanya mempercepat kecepatan komputasi tetapi juga memungkinkan penanganan dataset besar yang diperlukan untuk mendapatkan wawasan yang berarti (Fagundes-Júnior *et al.*, 2023).

II.11 Penelitian Sejenis

Penelitian sejenis atau literatur sejenis merupakan tinjauan pustaka di mana penulis mengumpulkan berbagai penelitian terdahulu yang berkaitan dengan topik yang sedang peneliti lakukan, kemudian melakukan perbandingan antara peneliti dengan penelitian sebelumnya serta sebagai referensi dalam penelitian. Pada Tabel II.1 merupakan kumpulan literatur sejenis dalam analisis sentimen dengan algoritma *Naïve Bayes* dan *Support Vector Machine* menggunakan optimasi PSO dalam berbagai penelitian.

Tabel II.1 Penelitian Sejenis yang relevan

No	Nama Penulis	Publish	Judul	Metodologi	Dataset	Target	Evaluasi	Hasil
1	(Rousyati *et al*., 2023)	*Journal of Applied Informatics and Computing (JAIC), Vol.7, No.2, Desember 2023*	Sentiment Analysis on Fuel Purchase Policy Through MyPertamina Application Using NB and SVM Methods Optimized by PSO as Weight Optimization	- Penelitian ini menggunakan Naive Bayes dan SVM dengan optimasi PSO untuk analisis sentimen pada aplikasi MyPertamina. - Preprocessing: cleansing data untuk mengurangi data redundan, tokenizing, stopword removal dan stemming menggunakan RapidMiner.	- Dataset diperoleh dari ulasan aplikasi MyPertamina di Google Play Store sebanyak 1.999 data.	Mengklasifikasikan sentimen ulasan pengguna berupa Positif dan Negatif.	Evaluasi mengguna kan Akurasi, Precision, Recall dan AUC.	- Hasil: Algoritma NB + PSO lebih unggul dari pada SVM + PSO yaitu menghasilkan akurasi sebesar 79.49%, precision tertinggi sebesar 79.57%, recall sebesar 79.38%, dan AUC tertinggi sebesar 95.30%.
2	(Kurniawan *et al*., 2019)	*Jurnal Resti (Rekayasa Sistem dan Teknologi Informasi), Sinta 2, Vol.*	Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik pada	- Penelitian ini menggunakan algoritma NB dan SVM yang dioptimasikan dengan PSO. - Penelitian ini menngunakan metode CRISP-DM - Preprocessing: case folding, tokenizing, tagging, filtering,	- Dataset diperoleh dari 6 media online Indonesia periode 16 Februari-31	Menentuka n algoritma terbaik dalam mengklasifikasikan komentar	Evaluasi mengguna kan akurasi dan AUC	- Hasil: Algoritma SVM + PSO mendapat akurasi terbaik dengan akurasi 78.40% dan AUC 0.850

No	Nama Penulis	Publish	Judul	Metodologi	Dataset	Target	Evaluasi	Hasil
		3, No. 2, Tahun 2019	Komentar Media Berita Online.	stemming menggunakan Gata Framework Textmining dan proses pengolahan data menggunakan RapidMiner. - Deployment: VB.NET dan MySQL.	Mei 2018 dengan jumlah data yang sudah dibersihkan berjumlah 1.459 komentar.	masyarakat terkait tokoh politik.		sedangkan NB + PSO dengan akurasi 74.98% dan AUC 0.708.
3	(Anggita Dyah & Ikmah, 2020)	*Jurnal Resti (Rekayasa Sistem dan Teknologi Informasi), Sinta 2, Vol. 4, No. 2, Tahun 2020*	Komparasi Algoritma Klasifikasi Berbasis Particle Swarm Optimization pada Analisis Sentimen Ekspedisi Barang	- Penelitian dilakukan dengan algoritma NB dan SVM yang dioptimasi dengan PSO. - Preprocessing: tokenize, transform case dan stem porter, pengolahan data menggunakan RapidMiner.	- Dataset menggunakan jasa ekspedisi Sicepat berjumlah 272 data.	Membandingkan performa klasifikasi Naïve Bayes dan SVM tanpa optimasi maupun setelah optimasi.	Evaluasi menggunakan confusion matrix dan cross validation	- Hasil: algoritma NB+PSO memiliki tingkat akurasi lebih tinggi yaitu 80.81% dibandingkan dengan SVM+PSO dengan akurasi 80.3%.
4	(Puspitaningtyas *et al*., 2025)	*Jurnal Riset Informatika, Sinta 4, Vol.*	Implementation of Support Vector Machine,	- Penelitian menggunakan algoritma NB dan SVM dengan PSO.	- Dataset diperoleh dari Kaggle (ulasan	Menganalisis dan mengklasifikasikan	Evaluasi menggunakan Confusion	- Hasil: Algoritma SVM + PSO terbukti memiliki akurasi

No	Nama Penulis	Publish	Judul	Metodologi	Dataset	Target	Evaluasi	Hasil
		7, No. 2, Maret 2025	Particle Swarm Optimization, and Naïve Bayes Algorithms in Sentiment Analysis of Product Reviews: A Case Study of E-Commerce Lazada	- Preprocessing: menghapus hastag, URL, retweet dan mention menggunakan tools RapidMiner.	produk lazada) sebanyak 810 data setelah proses pembersihan data.	sentimen pengguna berupa positif/ negatif.	Matrix, Akurasi dan AUC.	lebih unggul sebesar 84.84% dan AUC 0.898 dari pada algoritma SVM tanpa optimasi dan algoritma NB
5	(Wardhani *et al.*, 2018)	*Journal of Theoretical and Applied Information Technology (JATIT), Vol. 96, No. 24, Desember 2018.*	Sentiment Analysis Article News Coordinator Minister of Maritime Affairs Using Algorithm Naive Bayes and Support	- Penelitian menggunakan algoritma NB dan SVM dengan PSO. - Menggunakan kerangka CRISP-DM. - Preprocessing: Tokenize, filter tokens, stopwords removal, transform cases menggunakan tools Rapidminer	- Dataset diperoleh dari 6 situs berita Indonesia, periode 28 Juli 2016 – 23 November 2017 yang berjumlah 200 data.	Melakukan analisis dan klasifikasi sentimen positif/ negatif dari artikel berita terkait figur politik	Evaluasi mengguna kan akurasi dan AUC.	- Hasil: model NB+PSO mendapat akurasi tertinggi yaitu 92% sedangkan model SVM+PSO 90.50%. Namun untuk AUC

No	Nama Penulis	Publish	Judul	Metodologi	Dataset	Target	Evaluasi	Hasil
			Vector Machine with Particle Swarm Optimization			(Menko Kemaritiman)		tertinggi didapat dari model SVM tanpa optimasi sebesar 0.979
6	(Saepudin *et al*., 2024)	*Computer Science (CO-SCIENCE), Vol. 4 No. 1, Januari 2024*	Analisis Sentimen Pemanfaatan Artificial Intelligence di Dunia Pendidikan menggunakan SVM berbasis Particle Swarm Optimization	- Penelitian ini menggunakan algoritma SVM yang dioptimalkan dengan pemilihan fitur PSO. - Proses pre-processing meliputi proses tokenize, transform case, stopword removal, filter token by length dan n-grams. - Pengujian dilakukan dengan menggunakan tools RapidMiner.	- Dataset diperoleh dari crawling data X/Twitter menggunakan python melalui google colab sebanyak 400 data.	Meningkatkan akurasi klasifikasi sentimen dan mengklasifikasikan sentimen (Positif/ Negatif).	Evaluasi menggunakan akurasi, presisi, recall dan AUC	- Hasil algoritma SVM + PSO menghasilkan nilai *accuracy* 89.50%, *precision* 86.98%, *recall 93%* dan AUC 0.964. Sedangkan algoritma SVM mimiliki nilai *accuracy* 87.50%, *precision* 85.46%, *recall 90.50%* dan AUC 0.956.

No	Nama Penulis	Publish	Judul	Metodologi	Dataset	Target	Evaluasi	Hasil
7	(Sasi *et al*., 2023)	*International Journal of Research Publication and Reviews, Vol. 4 No. 1, Januari 2023*	Analysis of Twitter Sentiments About the Russian-Ukraine War Using Naive Bayes Based on Particle Swarm Optimization	- Penelitian ini menggunakan algortima NB yang di optimasi dengan PSO. - Preprocessing: Cleaning, Labeling, Transform Case, Tokenizing, Stopword removal, Stemming, Filter Token, TF-IDF, Split data 70:30, prosesnya menggunakan RapidMiner.	- Dataset diperoleh dari Twitter API sejumlah 1.260 tweet setelah dilakukan proses pembersihan.	Melakukan klasifikasi sentimen (positf/ negatif) dan membandingkan performa algoritma	Evaluasi mengguna kan confusion matrix dan error rate.	- Hasil: Model NB+PSO lebih unggul dengan akurasi 73,48% dengan error rate 50.36%. NB tanpa optimasi yaitu dengan akurasi 67,72% dengan error rate 57.14%.

Melalui tabel studi literatur yang telah dipaparkan di atas, penulis membandingkan berbagai penelitian terdahulu dengan penelitian yang dilakukan saat ini guna mengidentifikasi perbedaan dan kontribusi yang dihasilkan. Rangkuman perbandingan tersebut ditampilkan pada tabel berikut.

No.	Penelitian	Dataset		Polaritas		Ekstraksi Fitur		Algoritma		Optimasi PSO	
		Twitter	Lainnya	Binary Class	Multi Class	TF-IDF	Lainnya	NB + SVM	Lainnya	Feature Selection	Hyperparameter Tuning
1	(Rousyati *et al*., 2023)		√		√	√		√			√
2	(Kurniawan *et al*., 2019)		√	√			√	√			√
3	(Anggita Dyah & Ikmah, 2020)	√		√			√	√			√
4	(Puspitaningtyas *et al*., 2025)		√	√				√			√
5	(Wardhani *et al*., 2018)		√	√			√	√		√	
6	(Saepudin *et al*., 2024)	√		√			√		√	√	
7	(Sasi *et al*., 2023)	√		√		√			√		√
8	(Nurhayati, 2026)	√			√	√		√			√

Berdasarkan perbandingan terhadap penelitian terdahulu, dapat disimpulkan bahwa sebagian besar studi analisis sentimen masih berfokus pada domain aplikasi digital dan layanan berbasis *platform* dengan pendekatan klasifikasi biner serta penggunaan algoritma *Naïve Bayes* dan *Support Vector Machine* yang dioptimasi menggunakan *Particle Swarm Optimization*. Namun, penelitian-penelitian tersebut umumnya hanya menampilkan hasil akhir model setelah optimasi tanpa menyajikan analisis komparatif yang sistematis antara performa *baseline* dan model teroptimasi. Selain itu, evaluasi kinerja model masih didominasi oleh metrik *accuracy* dan AUC, sehingga belum sepenuhnya menggambarkan stabilitas dan keseimbangan klasifikasi antar kelas.

Penelitian ini memberikan kebaruan pada perluasan domain kajian ke isu kebijakan publik strategis, yaitu sentimen terhadap kebijakan efisiensi anggaran pemerintahan, serta menyajikan analisis komprehensif terhadap perbandingan performa model sebelum dan sesudah optimasi PSO pada algoritma *Naïve Bayes* dan *Support Vector Machine*. Dengan pendekatan evaluasi *multidimensional* menggunakan *accuracy*, *precision*, r*ecall*, *F1-score*, dan *confusion matrix*, penelitian ini diharapkan memberikan kontribusi metodologis yang lebih robust sekaligus relevansi praktis dalam memahami respons publik terhadap kebijakan nasional.