

BAB II

TINJAUAN PUSTAKA

A. Penelitian Terdahulu

Penelitian-penelitian yang serupa telah dilaksanakan antara lain Kasih *et al.* (2021) melakukan penelitian yang berjudul Rancang Bangun Aplikasi *Web* untuk Analisis Sentimen Pengguna Twitter terhadap Kinerja Pemerintah Provinsi Jawa Barat Menggunakan *Naïve Bayes Classifier*. Analisis sentimen pengguna *Twitter* tentang kinerja pemerintah menggunakan *Naïve Bayes Classifier* dengan *Flask*. Data teks melalui pembersihan, tokenisasi, *stopword removal*, dan *stemming*, lalu diekstraksi menggunakan *CountVectorizer*. Model dilatih pada data berlabel sentimen (positif dan negatif). *Flask* digunakan untuk membangun aplikasi *web* yang menampilkan hasil analisis dalam bentuk *pie chart* dan *bar chart*, memungkinkan pemrosesan data dan visualisasi yang terpadu.

Haerunnissa *et al.* (2022) melakukan penelitian yang berjudul Analisis Sentimen Kepuasan Pelanggan Perusahaan Telekomunikasi Seluler Telkomsel di *Twitter*. Sistem ini dibangun menggunakan *Python* dan *Flask*, dengan menu-menu seperti *dashboard*, data *tweet*, *preprocessing*, pelabelan manual, *training*, *testing*, dan klasifikasi menggunakan *Naive Bayes*. Grafik bar digunakan untuk menampilkan jumlah *tweet* berdasarkan kelas positif, negatif, dan netral. Namun, sistem ini belum mendukung pemanggilan *file* secara otomatis.

Winarno & Agung (2020) melakukan penelitian yang berjudul Visualisasi Data Untuk Analisis Sentimen Pada Penilaian Kepuasan Pelanggan. Aplikasi ini menganalisis ulasan kepuasan pelanggan menggunakan sentiment analysis SVM dan menyajikan hasilnya dalam bentuk *wordcloud* untuk menampilkan kata-kata yang sering muncul dari sentimen positif dan negatif. Visualisasi ini membantu perusahaan memahami persepsi pelanggan, mengidentifikasi kekuatan, serta menemukan area untuk perbaikan secara cepat melalui antarmuka *web*.

Handayani (2021) melakukan penelitian yang berjudul Analisis Sentimen Berita Covid-19 Berbasis *Web* Menggunakan *Python Flask* dan *TextBlob*. Aplikasi berbasis *Flask* ini mengidentifikasi sentimen berita Covid-19 menggunakan *TextBlob* untuk analisis awal dan *Naive Bayes Classifier* untuk meningkatkan akurasi. Fitur seperti TF-IDF digunakan untuk ekstraksi fitur. Aplikasi ini mengklasifikasikan sentimen menjadi positif dan negatif, serta menyajikan visualisasi seperti *wordcloud*, *pie chart*, dan grafik batang, membantu pengguna memahami pola sentimen publik terkait pandemi secara komprehensif.

Rahman (2023) melakukan penelitian yang berjudul Implementasi *Machine Learning* Pada *Application Programming Interface* Untuk Klasifikasi Komentar *Cyberbullying*. mengembangkan API berbasis *Machine Learning* untuk mendeteksi dan membatasi komentar *cyberbullying* menggunakan pendekatan NLP, mengikuti metodologi penelitian Muneer & Fati (2020) dan *framework FAST*. API ini dibangun dengan, *Python 3.9*, *Flask 2.1*, dan

MongoDB, dan akan menyediakan fitur deteksi teks *cyberbullying*, akses *API*, kontak *developer*, serta pengiriman *feedback* pengguna.

Aretha (2019) melakukan penelitian yang berjudul Rancang Bangun Aplikasi Analisis Sentimen Layanan Pdam Surya Sembada Kota Surabaya Dengan Metode *Naive Bayes Classifier*. Penelitian ini mengusulkan sistem otomatis untuk PDAM Surabaya yang menggunakan *Text Mining* untuk mengolah berita dan metode *Naive Bayes Classifier* untuk mengklasifikasikan sentimen berita menjadi positif atau negatif. Sistem ini membantu PDAM dalam menganalisis jumlah berita positif dan negatif secara efisien, sehingga dapat menilai opini publik terhadap layanan mereka.

Harita *et al.* (2022) melakukan penelitian yang berjudul Pengembangan Aplikasi Analisis Sentimen Terhadap Brand Berbasis *Web* Menggunakan Kerangka Kerja *Flask*. Aplikasi ini dibangun menggunakan *Python* dan *Flask*, yang menampilkan data *real-time* dari Twitter serta menyimpan hasil analisis sentimen terhadap tiga brand (Samsung, Apple, dan Xiaomi) ke dalam *database MySQL*. Aplikasi ini dapat diakses secara *online* melalui *platform hosting* Heroku yang mendukung *Flask*.

Effendi dan Noviana (2021) melakukan penelitian yang berjudul Perancangan *Web* Aplikasi Analisis Sentimen Media Sosial *Twitter* Dengan Metode *Valence Aware Dictionary and Sentiment Reasoner* (VADER) Menggunakan *PHP* dan *MySQL* pada Pemerintah Kota Bekasi. Penelitian ini mengembangkan aplikasi analisis sentimen untuk mengelompokkan *tweet* ke dalam kategori positif, netral, dan negatif menggunakan metode VADER.

Aplikasi ini memanfaatkan *API Twitter* untuk *crawling data* dan menggunakan *PHP* serta *MySQL*. Tujuannya adalah membantu analisis media dalam memantau isu publik terkait Pemerintah Kota Bekasi.

Hilal *et al.* (2024) melakukan penelitian yang berjudul Analisis Sentimen Berbasis Aspek Untuk Pengguna PLN *Mobile* Pada *Google Playstore* Menggunakan Metode *Support Vector Machine* (SVM). Penelitian ini menganalisis ulasan pengguna berbahasa Indonesia dari Juli 2023 hingga April 2024 menggunakan LDA untuk ekstraksi aspek dan SVM untuk klasifikasi sentimen. Data dikumpulkan lewat *web scraping* dari *Google PlayStore* dan diproses dengan normalisasi, tokenisasi, dan *stemming*. Evaluasi dilakukan dengan *Confusion Matrix*, *K-Fold Cross-Validation*, dan hasilnya divisualisasikan menggunakan *Wordcloud*.

Putra *et al.* (2023) melakukan penelitian yang berjudul Penerapan Metode *Extreme Programming* Pada Rancang Bangun Sistem Analisis Sentimen Portal Berita. Sistem analisis sentimen ini dikembangkan dengan mengintegrasikan layanan *Google NLP* untuk menentukan skor sentimen setiap kalimat dan menggunakan teknik *web scraping* untuk pengambilan data. Sistem ini dibangun dengan *framework Laravel* dan mengadopsi metode *Extreme Programming* untuk pengembangan cepat. *Website* dipilih sebagai *platform* untuk memastikan aksesibilitas dari berbagai perangkat, baik *mobile* maupun *desktop*. Sistem ini menjadi solusi alternatif bagi individu dan organisasi dalam melakukan analisis sentimen, membantu proses pengambilan keputusan dan evaluasi kinerja.

Tabel 2.1 Penelitian Terdahulu

No	Judul (Penulis)	Metode	Hasil
1	Rancang Bangun Aplikasi Web untuk Analisis Sentimen Pengguna Twitter terhadap Kinerja Pemerintah Provinsi Jawa Barat Menggunakan <i>Naive Bayes Classifier</i> (Kasih <i>et al.</i> , 2021)	<i>Naive Bayes</i>	Aplikasi ini menggunakan <i>Naive Bayes Classifier</i> untuk menganalisis sentimen pengguna <i>Twitter</i> mengenai kinerja pemerintah, dengan menggunakan <i>framework Flask</i> dan <i>CountVectorizer</i> .
2	Analisis Sentimen Kepuasan Pelanggan Perusahaan Telekomunikasi Seluler Telkomsel di Twitter (Haerunnissa <i>et al.</i> , 2022)	<i>Naive Bayes</i>	Sistem ini menggunakan <i>Python</i> dan <i>Flask</i> serta klasifikasi <i>Naive Bayes</i> . Grafik bar menampilkan jumlah <i>tweet</i> berdasarkan kelas positif, negatif, dan netral, namun belum mendukung pemanggilan <i>file</i> otomatis.
3	Visualisasi Data Untuk Analisis Sentimen Pada Penilaian Kepuasan Pelanggan (Winarno & Agung, 2020)	SVM	Menggunakan visualisasi <i>wordcloud</i> untuk menampilkan sentimen pada ulasan kepuasan pelanggan, aplikasi ini membantu perusahaan dalam memahami kepuasan pelanggan dari data ulasan yang dikumpulkan
4	Analisis Sentimen Berita Covid-19 Berbasis Web Menggunakan <i>Python Flask</i> dan <i>TextBlob</i> (Handayani, 2021)	<i>Naive Bayes</i>	Mengembangkan aplikasi berbasis <i>Flask</i> yang mengidentifikasi sentimen berita Covid-19 menggunakan <i>TextBlob</i> . Hasil menunjukkan bahwa aplikasi membantu pengguna memahami persepsi publik terhadap pandemi.
5	Implementasi <i>Machine Learning</i> Pada Application Programming Interface Untuk Klasifikasi Komentar Cyberbullying. (Rahman, 2023)	<i>Random Forest</i>	Pembatasan <i>cyberbullying</i> dapat dilakukan dengan <i>model Supervised Machine Learning</i> menggunakan RFC, yang memiliki akurasi 91.25% pada <i>testing</i> dan performa baik (presisi 93.50%, <i>recall</i> 93.81%). Model dapat diintegrasikan ke dalam API untuk klasifikasi teks.
6	Rancang Bangun Aplikasi Analisis Sentimen Layanan Pdam Surya Sembada Kota Surabaya Dengan Metode <i>Naive Bayes Classifier</i> . (Aretha, 2019)	<i>Naive Bayes</i>	Aplikasi ini mengumpulkan berita menggunakan <i>crawling CRUL</i> , mengklasifikasikan sentimen dengan <i>Naive Bayes</i> , dan menampilkan jumlah berita positif dan negatif. Evaluasi menunjukkan akurasi 77%.
7	Pengembangan Aplikasi Analisis Sentimen Terhadap Brand Berbasis Web Menggunakan Kerangka Kerja <i>Flask</i> (Harita <i>et al.</i> , 2022)	<i>Naive Bayes</i>	merupakan aplikasi berbasis <i>web</i> untuk memvisualisasikan hasil klasifikasi analisis sentiment perbandingan <i>brand</i> . Aplikasi ini dibuat menggunakan kerangka kerja <i>Flask</i> untuk <i>back-end</i> dan kerangka kerja <i>Bootstrap 4</i> untuk <i>front-end</i> .

Tabel 2.1 Penelitian Terdahulu (lanjutan)

8	Perancangan <i>Web</i> Aplikasi Analisis Sentimen Media Sosial Twitter Dengan Metode <i>Valence Aware Dictionary and Sentiment Reasoner</i> (VADER) Menggunakan PHP dan <i>MySQL</i> pada Pemerintah Kota Bekasi. (Effendi & Noviana, 2021)	<i>Naive Bayes</i> dan SVM	Aplikasi analisis sentimen Twitter menggunakan VADER dapat mendeteksi <i>tweet</i> positif, negatif, atau netral dengan cepat. Aplikasi ini membantu memonitor isu publik terkait kebijakan pemerintah Kota Bekasi. Untuk mengakses data tanpa batasan waktu, diperlukan API Premium Twitter. Uji <i>black box</i> menunjukkan fungsi aplikasi berjalan dengan baik.
9	Analisis Sentimen Berbasis Aspek Untuk Pengguna <i>Pln Mobile</i> Pada <i>Google Playstore</i> Menggunakan Metode <i>Support Vector Machine</i> (SVM). (Hilal <i>et al.</i> , 2024)	LDA dan SVM	LDA berhasil mengekstrak 2 topik, yaitu <i>Application Service</i> dan <i>Customer Service</i> , dengan <i>Coherence Score</i> 0.52. Algoritma SVM mencapai akurasi 93.53% dalam <i>cross-validation K-Fold</i> dengan pembagian data 90:10. 121 positif dan 613 negatif untuk <i>Application Service</i> , serta 282 positif dan 563 negatif untuk <i>Customer Service</i> .
10	Penerapan Metode <i>Extreme Programming</i> Pada Rancang Bangun Sistem Analisis Sentimen Portal Berita. (Premana Putra <i>et al.</i> , 2023)	SVM	Hasil pengujian menunjukkan bahwa sistem berhasil menampilkan <i>score sentimen</i> dan klasifikasi emosional kata. Meskipun sudah bermanfaat, pengembangan fitur dan evaluasi berkala diperlukan agar sistem tetap efektif dan mengikuti perkembangan teknologi.

B. Landasan Teori

1. Analisis Sentimen

Analisis sentimen adalah proses untuk menentukan dan mengklasifikasikan sikap atau opini yang terkandung dalam sebuah teks, apakah bersifat positif, negatif, atau netral. Menurut Pang & Lee (2008), analisis sentimen digunakan untuk memproses data teks dalam bentuk ulasan atau opini, dengan tujuan untuk memahami perasaan dan pendapat publik terhadap suatu entitas. Menurut Liu (2015), analisis sentimen dapat dibagi menjadi dua pendekatan utama: pendekatan

berbasis kamus dan pendekatan berbasis pembelajaran mesin. Pendekatan berbasis kamus menggunakan daftar kata positif dan negatif untuk menentukan sentimen, sedangkan pendekatan berbasis pembelajaran mesin melibatkan pelatihan model menggunakan data yang telah dilabeli.

Dalam analisis sentimen, data sentimen umumnya dikategorikan menjadi tiga kelas utama: positif, negatif, dan netral.

a) Sentimen Positif

Merupakan kategori yang menunjukkan bahwa opini atau perasaan pengguna terhadap suatu topik adalah baik atau mendukung. Data sentimen positif sering kali digunakan untuk mengukur kepuasan pelanggan dan keberhasilan suatu produk atau layanan (Pang & Lee, 2008).

b) Sentimen Negatif

Kategori ini menunjukkan bahwa opini atau perasaan pengguna terhadap suatu topik adalah buruk atau menolak. Data sentimen negatif penting untuk mengidentifikasi masalah dan area yang perlu diperbaiki dalam produk atau layanan (Pang & Lee, 2008).

c) Sentimen Netral

Kategori ini mencakup opini atau perasaan yang tidak menunjukkan sikap positif atau negatif yang jelas. Data sentimen netral sering kali digunakan untuk analisis yang lebih mendalam,

di mana pengguna mungkin memiliki pandangan yang beragam atau tidak pasti (Liu, 2015).

Analisis sentimen tidak hanya membantu dalam memahami opini publik tetapi juga dalam pengambilan keputusan berbasis data. Dengan mengklasifikasikan data sentimen ke dalam kategori-kategori ini, organisasi dapat lebih mudah mengidentifikasi tren, pola, dan area yang memerlukan perhatian lebih lanjut.

2. Metode *Machine Learning*

Metode *Machine Learning*, khususnya dalam konteks analisis sentimen, melibatkan penggunaan algoritma untuk mengklasifikasikan teks berdasarkan data pelatihan. Penelitian ini menggunakan algoritma *machine learning* untuk menganalisis sentimen, yaitu *Naive Bayes* yang efisien dan akurat untuk klasifikasi teks dengan data besar berdasarkan teorema *Bayes* (Rish, 2001). *Support Vector Machine* (SVM) yang mencari *hyperplane* optimal untuk memisahkan kelas pada teks berdimensi tinggi (Cortes & Vapnik, 1995), *K-Nearest Neighbor* (KNN) yang mengklasifikasikan data berdasarkan kedekatan jarak antar titik (Cover & Hart, 1967), serta *Random Forest*, metode *ensemble* berbasis *decision trees* yang akurat dan tahan terhadap *overfitting* (Breiman, 2001). Zhang dan Wallace (2015) menjelaskan bahwa pemilihan algoritma yang tepat dapat mempengaruhi akurasi dan efektivitas model dalam mengklasifikasikan sentimen. Zhang dan Wallace (2015) menjelaskan bahwa pemilihan algoritma yang tepat dapat

mempengaruhi akurasi dan efektivitas model dalam mengklasifikasikan sentimen.

Setiap algoritma memiliki kelebihan dan kekurangan, dan pemilihan algoritma yang tepat harus didasarkan pada karakteristik data dan tujuan analisis. Misalnya, Naive Bayes sering digunakan untuk klasifikasi teks karena kesederhanaannya dan kecepatan dalam pelatihan, sementara SVM dapat memberikan hasil yang lebih baik pada dataset yang lebih kompleks.

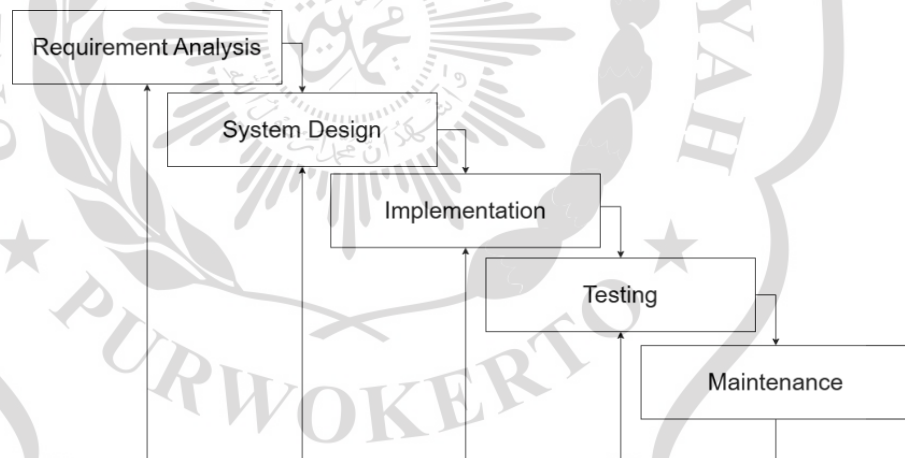
3. *Natural Language Processing (NLP)*

Natural Language Processing (NLP) adalah cabang dari kecerdasan buatan yang berfokus pada interaksi antara komputer dan manusia menggunakan bahasa alami. Menurut Klabunde (2002), teknik NLP seperti *tokenization*, *stemming*, dan *stopwords removal* sangat penting dalam mempersiapkan data teks untuk analisis lebih lanjut. Teknik-teknik ini membantu dalam membersihkan dan memformat data agar lebih mudah dianalisis.

Tokenization memecah teks menjadi kata-kata atau frasa, sementara *stemming* mengubah kata-kata ke bentuk dasarnya. *Stopwords removal* menghapus kata-kata umum yang tidak memberikan informasi penting dalam analisis. Dengan menerapkan teknik-teknik ini, data teks dapat diproses dengan lebih efisien, meningkatkan akurasi analisis sentimen.

4. *Waterfall*

Metode *waterfall* adalah model pengembangan perangkat lunak yang terstruktur dan linier, yang membagi proses pengembangan menjadi beberapa tahapan yang saling bergantung, yaitu *requirements*, *design*, *implementation*, *testing*, dan *maintenance*. Setiap tahap harus diselesaikan sebelum melanjutkan ke tahap berikutnya, sehingga memberikan kontrol yang jelas dan dokumentasi yang terstruktur dalam pengembangan aplikasi (Pressman, 2009). Metode *waterfall* dipilih dalam penelitian ini untuk memastikan setiap tahap pengembangan aplikasi dilakukan secara sistematis dan terdokumentasi. Ilustrasi *waterfall* dapat dilihat pada gambar 2.1.



Gambar 2.1 Metode Waterfall

5. *Flask Framework*

Flask adalah sebuah *web framework* berbasis *Python* yang ringan dan fleksibel. *Flask* memungkinkan pengembangan aplikasi *web* dengan cepat dan mudah, serta mendukung integrasi dengan berbagai pustaka *Python*, termasuk pustaka untuk analisis data dan visualisasi.

Menurut Grinberg (2018), *Flask* sangat cocok untuk pengembangan aplikasi berbasis *web* yang membutuhkan fungsionalitas sederhana namun kuat, seperti aplikasi analisis sentimen dengan visualisasi interaktif.

6. Visualisasi Data

Visualisasi data merupakan teknik untuk menyajikan data dalam bentuk grafis yang memudahkan pemahaman dan interpretasi. Dalam konteks analisis sentimen, visualisasi seperti *pie chart*, *bar chart*, dan *wordcloud* digunakan untuk menunjukkan distribusi sentimen dan frekuensi kata yang sering muncul dalam data. Menurut Few (2012), visualisasi yang baik dapat membantu pengguna untuk memahami pola dan tren dalam data, serta mengambil keputusan berdasarkan hasil analisis.

7. Black Box Testing

Black Box Testing adalah metode pengujian perangkat lunak yang berfokus pada pengujian fungsionalitas aplikasi tanpa memeriksa kode internalnya. Pengujian ini dilakukan berdasarkan spesifikasi kebutuhan, memastikan bahwa setiap fungsi aplikasi bekerja sesuai yang diharapkan dari perspektif pengguna. Dalam penelitian ini, *Black Box Testing* diterapkan untuk menguji semua fitur utama aplikasi, seperti *preprocessing* teks, analisis sentimen, dan visualisasi data (Myers *et al.*, 2011). Dengan metode ini, setiap keluaran yang dihasilkan oleh aplikasi

dapat dibandingkan dengan spesifikasi untuk memastikan aplikasi berjalan sesuai dengan fungsinya.

8. *SQLite* sebagai Sistem Basis Data

SQLite adalah sistem manajemen basis data relasional yang ringan dan berbasis *file*, yang sering digunakan untuk aplikasi *web* skala kecil hingga menengah. Dalam penelitian ini, *SQLite* digunakan untuk menyimpan data hasil analisis sentimen dan visualisasi. Menurut Steinberg (2016), *SQLite* memberikan kinerja yang cepat dengan memori yang efisien dan dapat diintegrasikan dengan berbagai aplikasi berbasis *Python*, seperti *Flask*, untuk penyimpanan data yang efektif.

