BAB II
TINJAUAN PUSTAKA

2.1 Penelitian Yang Relevan

1. Penelitian yang dilakukan oleh Ade Romadhony, Said Al Faraby, Rita Rismala, Untari Novia Wisesti, dan Anditya Arifianto (2024) dengan judul "Sentiment Analysis on a Large Indonesian Product Review Dataset" membahas penerapan analisis sentimen pada dataset ulasan produk berbahasa Indonesia dalam skala besar, yaitu Female Daily Review (FDReview) yang terdiri dari lebih dari 700.000 ulasan pengguna. Tujuan penelitian ini adalah untuk melakukan klasifikasi sentimen (positif, netral, dan negatif) serta prediksi rating berdasarkan teks ulasan. Penelitian ini menggunakan dua pendekatan utama, yaitu metode *machine learning* konvensional (*Multinomial Naïve Bayes* dan *Support Vector Machine*/SVM) serta metode deep learning (*Long Short-Term Memory/LSTM* dan *Bidirectional LSTM/BiLSTM*).

   Proses penelitian meliputi tahapan pra-pemrosesan data seperti normalisasi kata, penghapusan *stopword*, dan konversi teks informal menjadi bentuk formal. Fitur teks diekstraksi menggunakan metode *Bag of Words* dan *TF- IDF*, kemudian data dibagi menjadi data pelatihan dan pengujian dengan rasio 80:20. Hasil pengujian menunjukkan bahwa metode BiLSTM menghasilkan performa terbaik untuk tugas klasifikasi sentimen dan prediksi rating pada dataset besar, sementara SVM menunjukkan kinerja yang lebih baik pada dataset kecil dan seimbang.

   Kesimpulan dari penelitian ini menunjukkan bahwa ukuran dataset memiliki pengaruh signifikan terhadap kinerja model analisis sentimen. Pendekatan berbasis deep learning seperti BiLSTM terbukti unggul hanya pada data berskala besar, sedangkan metode konvensional seperti SVM dan *Naïve Bayes* masih mampu memberikan hasil kompetitif pada dataset kecil. Penelitian ini memberikan kontribusi penting dengan menyediakan analisis serta sistem dasar (*baseline system*) untuk tugas klasifikasi sentimen dan prediksi rating pada

dataset ulasan produk Indonesia berskala besar, yang sebelumnya masih jarang tersedia untuk penelitian lebih lanjut (Romadhony et al., 2024).

2. Penelitian yang dilakukan oleh Yusuf Ansori dan Khadijah Fahmi Hayati Holle (2022) dengan judul "*Perbandingan Metode Machine Learning dalam Analisis Sentimen Twitter*" membahas penerapan dan perbandingan beberapa algoritma *machine learning* dalam mengklasifikasikan sentimen pada data media sosial *Twitter*. Penelitian ini mengambil studi kasus tanggapan masyarakat terhadap Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 30 Tahun 2021 mengenai penanganan kekerasan seksual di lingkungan kampus. Empat algoritma klasifikasi digunakan dalam penelitian ini, yaitu *Support Vector Machine* (SVM), *K-Nearest Neighbor* (KNN*), Naïve Bayes Classifier* (NBC), dan *Logistic Regression* (LR), untuk mengelompokkan data *tweet* ke dalam dua kategori sentimen, yaitu positif dan negatif.

Tahapan penelitian meliputi proses data crawling menggunakan *Twitter API,* pelabelan data secara manual, serta preprocessing teks yang mencakup *cleansing, case folding, tokenizing, dan stemming*. Data kemudian dibagi menggunakan teknik *k-Fold Cross Validation* agar pembagian data latih dan uji lebih merata. Evaluasi kinerja dilakukan menggunakan metrik akurasi, presisi, *recall*, dan *f- measure*, dengan hasil bahwa algoritma SVM memberikan akurasi tertinggi sebesar 69,15%, diikuti oleh KNN dengan presisi sebesar 69,07% dan *f-measure* sebesar 68,08%.

Kesimpulan dari penelitian ini menunjukkan bahwa metode Support Vector Machine (SVM) memiliki kinerja paling unggul dalam klasifikasi sentimen teks dibandingkan algoritma lainnya. Penelitian ini memberikan kontribusi terhadap pengembangan analisis sentimen berbasis machine learning pada data media sosial berbahasa Indonesia, sekaligus memperlihatkan efektivitas perbandingan beberapa algoritma dalam menentukan model yang paling optimal untuk klasifikasi opini publik (Ansori & Holle, 2022).

3. Penelitian yang dilakukan oleh Lutfi Budi Ilmawan dan Muhammad Aliyazid Mude (2020) berjudul "*Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store*" membahas Studi ini melakukan evaluasi perbandingan kinerja antara *algoritma Support Vector Machine* (SVM) dan *Naïve Bayes* dalam mengklasifikasikan sentimen ulasan aplikasi berbahasa Indonesia di *Google Play Store*. Penelitian tersebut bertujuan untuk menentukan algoritma yang paling optimal dalam memetakan ulasan pengguna ke dalam tiga kelas sentimen, yakni positif, negatif, dan crash. Data yang digunakan merupakan 1818 komentar ulasan pengguna, yang terdiri dari 606 komentar pada setiap kelas sentimen.

Tahapan penelitian meliputi proses *preprocessing* seperti *case folding,* penghapusan *stopword* dan tanda baca, serta penerapan *negation tag* untuk menangani kalimat negatif. Fitur yang digunakan adalah unigram, dan proses pelatihan dilakukan menggunakan *k-fold cross-validation* (K=3) untuk mengukur akurasi model. Implementasi dilakukan menggunakan bahasa pemrograman *Python*, dengan *Naïve Bayes* diimplementasikan manual, sedangkan SVM menggunakan library LibSVM. Hasil pengujian menunjukkan bahwa SVM memiliki tingkat akurasi lebih tinggi sebesar 81,46%, sedangkan *Naïve Bayes* memperoleh akurasi 75,41%.

Kesimpulan dari penelitian ini adalah bahwa metode *Support Vector Machine* lebih unggul dibandingkan *Naïve Bayes* dalam melakukan analisis sentimen terhadap ulasan berbahasa Indonesia, terutama karena kemampuannya menangani data dengan fitur yang kompleks dan distribusi yang tidak seimbang. Penelitian ini memberikan kontribusi penting dalam memperkuat bukti empiris bahwa SVM merupakan salah satu algoritma klasifikasi yang efektif untuk tugas analisis sentimen berbasis teks di bahasa Indonesia, khususnya pada data ulasan aplikasi digital seperti di *Google Play Store* (Ilmawan & Mude, 2025).

4. Penelitian yang dilakukan oleh Jessica Widyadhana Iskandar dan Yessica Nataliani (2021) berjudul "*Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek*" membahas penerapan dan

perbandingan tiga algoritma klasifikasi, yaitu *Naïve Bayes* (NB), *Support Vector Machine* (SVM), dan *k-Nearest Neighbor* (k-NN) untuk melakukan analisis sentimen terhadap komentar pengguna *Youtube* mengenai *Samsung Galaxy Z Flip 3* pada kanal GadgetIn. Penelitian ini bertujuan untuk mengetahui persepsi masyarakat terhadap produk gadget tersebut berdasarkan empat aspek, yaitu desain, harga, spesifikasi, dan citra merek. Data yang digunakan berjumlah 9.597 komentar, yang setelah melalui proses seleksi dan pelabelan manual menghasilkan 1.391 data relevan.

Metode penelitian menggunakan model CRISP-DM (*Cross Industry Standard Process for Data Mining*) yang meliputi tahap *business understanding, data understanding, data preparation, modeling, dan evaluation*. Proses *preprocessing* mencakup *transform case, tokenizing, normalization, dan stemming,* serta penggunaan teknik SMOTE (*Synthetic Minority Over-sampling Technique*) untuk menangani *imbalance* class dalam dataset. Pengujian dilakukan menggunakan perangkat lunak *RapidMiner 9.10* dengan *10-fold cross-validation*, dan evaluasi kinerja model diukur berdasarkan metrik akurasi, presisi, recall, dan f-measure.

Hasil penelitian menunjukkan bahwa algoritma SVM menghasilkan performa terbaik dengan rata-rata akurasi 96,43%, disusul oleh *Naïve Bayes* sebesar 83,54%, dan k-NN sebesar 59,68%. SVM menunjukkan keunggulan tertinggi pada seluruh aspek, dengan akurasi 94,40% untuk aspek desain, 97,44% untuk harga, 96,22% untuk spesifikasi, dan 97,63% untuk citra merek. Kesimpulan penelitian ini menyatakan bahwa SVM merupakan metode yang paling efektif untuk analisis sentimen berbasis aspek pada komentar Youtube berbahasa Indonesia, terutama karena kemampuannya dalam menangani data berukuran besar dan tidak seimbang. Penelitian ini juga merekomendasikan pengembangan lanjutan dengan algoritma lain atau topik berbeda untuk meningkatkan hasil klasifikasi dan memperluas penerapan analisis sentimen di masa depan (Iskandar & Nataliani, 2021).

5. Penelitian berjudul "*Analisis Sentimen Aplikasi Playstore SIREKAP 2024 Pasca Pilpres dengan Perbandingan Metode Support Vector Machine (SVM), Naïve Bayes Classifier, dan Random Forest*" membahas respons publik terhadap aplikasi SIREKAP yang dirilis Komisi Pemilihan Umum (KPU) pada Pemilu 2024. Penelitian ini menggunakan ulasan pengguna pada platform *Google Play Store* sebagai sumber data untuk menilai persepsi masyarakat pasca pemilihan presiden. Komentar-komentar tersebut kemudian diklasifikasikan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Penulis menerapkan tiga algoritma klasifikasi, yaitu SVM, *Naïve Bayes Classifier*, dan *Random Forest*, untuk mengetahui model yang paling akurat dalam mengidentifikasi sentimen publik terkait performa dan reliabilitas aplikasi SIREKAP selama proses rekapitulasi suara.

   Tahapan penelitian meliputi pengumpulan data ulasan melalui web scraping, kemudian dilanjutkan dengan preprocessing berupa *cleaning, tokenizing, stopword removal,* dan *stemming* untuk memperoleh teks yang lebih terstruktur. Data kemudian direpresentasikan dalam bentuk matriks fitur menggunakan metode TF-IDF sebelum dilakukan pelatihan model. Evaluasi performa ketiga algoritma dilakukan menggunakan metrik akurasi, presisi, recall, dan f1-score. Hasil penelitian menunjukkan bahwa SVM memiliki performa paling unggul dibandingkan dua metode lainnya, sedangkan *Naïve Bayes* menunjukkan akurasi yang lebih rendah karena sensitif terhadap distribusi data, dan Random Forest berada di posisi menengah namun kurang stabil pada data teks pendek seperti ulasan aplikasi.

   Penelitian ini menyimpulkan bahwa SVM merupakan metode yang paling efektif untuk analisis sentimen pada ulasan aplikasi Playstore, terutama dalam konteks isu publik yang sensitif seperti penyelenggaraan Pemilu. Studi ini memberikan gambaran yang jelas mengenai kecenderungan opini masyarakat terhadap aplikasi resmi pemerintah pasca Pilpres 2024, sekaligus memperkuat temuan bahwa algoritma berbasis margin maximization cenderung lebih akurat dalam menangani data teks berbahasa Indonesia pada platform digital (TARIGAN et al., 2025).

6. Penelitian yang dilakukan oleh Gishella Septania Al-Husna, Dian Asmarajati, Iman Ahmad Ihsanuddin, dan Rina Mahmudati (2024) berjudul "*Perbandingan Metode Naïve Bayes dan Support Vector Machine untuk Analisis Sentimen pada Ulasan Pengguna Aplikasi LinkedIn*" bertujuan untuk mengidentifikasi kecenderungan sentimen pengguna terhadap aplikasi LinkedIn berdasarkan ulasan di *Google Play Store*. Penelitian ini dilatarbelakangi meningkatnya penggunaan platform digital untuk mencari lowongan pekerjaan, serta pentingnya pemahaman terhadap persepsi pengguna terhadap aplikasi pencarian kerja yang banyak diandalkan masyarakat. Data dikumpulkan melalui teknik web scraping sebanyak 2000 ulasan, kemudian dilakukan proses text preprocessing yang mencakup *data cleaning, case folding, stopword removal, tokenizing,* dan *stemming* untuk menghasilkan data bersih yang siap dianalisis.

   Penelitian ini membandingkan dua algoritma klasifikasi, *yaitu Naïve Bayes* dan *Support Vector Machine* (SVM), yang masing-masing dilatih menggunakan pembobotan TF-IDF dan diuji menggunakan confusion matrix. Hasil pengujian menunjukkan bahwa metode *Naïve Bayes* memperoleh akurasi 88%, presisi 88%, recall 85%, dan f1-score 86%. Sementara itu, metode SVM menunjukkan hasil yang lebih tinggi dengan akurasi 90%, presisi 89%, recall 88%, dan f1-score 88%. Temuan ini menegaskan bahwa SVM memberikan performa lebih baik dibandingkan *Naïve Bayes* dalam mengklasifikasikan sentimen ulasan pengguna aplikasi LinkedIn.

   Kesimpulan penelitian ini menyatakan bahwa SVM merupakan algoritma yang lebih efektif untuk analisis sentimen pada ulasan aplikasi digital, karena mampu menghasilkan prediksi yang lebih akurat dan konsisten. Penelitian ini juga memberikan dasar empiris untuk penggunaan metode machine learning dalam menganalisis persepsi pengguna terhadap aplikasi berbasis layanan digital (Gishella Septania Al-Husna et al., 2024).

7. Penelitian yang dilakukan oleh Nabil Safiq Ramadan dan Dedi Darwis (2025) yang berjudul "*Perbandingan Metode Naïve Bayes Dan Svm Untuk Sentimen Analisis Masyarakat Terhadap Serangan Ransomware Pada Data Kip- K*" meneliti respons masyarakat terhadap kasus serangan ransomware yang menargetkan data KIP-K dengan menggunakan analisis sentimen pada komentar pengguna di media sosial X. Data awal yang dikumpulkan berjumlah 2.648 komentar, yang kemudian melalui rangkaian proses preprocessing seperti *cleaning, case folding, tokenizing, stopword removal,* dan *stemming*, sehingga menghasilkan 1.738 data bersih yang siap dianalisis. Data tersebut diberi label ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral, dengan distribusi yang menunjukkan dominasi sentimen negatif dari masyarakat terhadap insiden tersebut. Penelitian ini menerapkan dua algoritma klasifikasi, yakni *Naïve Bayes Classifier* dan *Support Vector Machine* (SVM), yang keduanya dilatih dan diuji menggunakan proporsi 70% data pelatihan dan 30% data pengujian.

Hasil pengujian menunjukkan bahwa SVM memberikan performa yang lebih baik dibandingkan *Naïve Bayes*, dengan akurasi mencapai 88%, sementara *Naïve Bayes* hanya menghasilkan akurasi 70%. Temuan tersebut mengindikasikan bahwa SVM lebih efektif dalam menangani data teks media sosial yang cenderung bervariasi dan memiliki konteks yang kompleks. Penelitian ini menyimpulkan bahwa sentimen publik terhadap kasus ransomware KIP-K cenderung negatif, mencerminkan kekhawatiran masyarakat terhadap keamanan sistem informasi pemerintah. Selain itu, penelitian ini menegaskan bahwa SVM merupakan metode yang lebih unggul untuk tugas klasifikasi sentimen pada data komentar media sosial berbahasa Indonesia (Ramadan & Darwis, 2024).

8. Penelitian yang dilakukan oleh Eva Rahma Indriyani, Paradise, dan Merlinda Wibowo (2022) yang berjudul "*Perbandingan Metode Naïve Bayes dan Support Vector Machine Untuk Analisis Sentimen Terhadap Vaksin Astrazeneca di Twitter*"mengkaji analisis sentimen masyarakat terhadap vaksin AstraZeneca dengan memanfaatkan data unggahan Twitter. Penelitian ini dilatarbelakangi oleh munculnya perdebatan publik terkait keamanan dan kehalalan vaksin AstraZeneca akibat isu kandungan tripsin babi yang sempat beredar. Data dikumpulkan menggunakan library snscrape pada periode Mei hingga Juni 2021 dan menghasilkan 3.105 tweet setelah pembersihan data dari duplikasi, bahasa asing, serta entri yang tidak relevan. Tahapan preprocessing yang dilakukan mencakup *cleaning, case folding, tokenizing, normalization, stopword removal,* dan *stemming*. Pelabelan

sentimen dilakukan menggunakan pendekatan *lexicon- based* dengan kamus *InSet Lexicon,* yang menghasilkan 1.275 tweet berlabel positif dan 1.830 tweet berlabel negatif. Data kemudian direpresentasikan menggunakan pembobotan TF-IDF untuk proses klasifikasi.

Penelitian ini membandingkan dua algoritma klasifikasi, yaitu *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM), dengan pembagian data 80% untuk pelatihan dan 20% untuk pengujian. Proses evaluasi dilakukan menggunakan *confusion matrix* dengan parameter akurasi, presisi, recall, dan f1- score. Hasil penelitian menunjukkan bahwa SVM memiliki performa terbaik dengan akurasi 87,27%, presisi 90,41%, recall 77,34%, dan f1-score 83,37%. Sementara itu, Naïve Bayes menghasilkan akurasi lebih rendah, yaitu 76,81%, dengan performa evaluasi yang juga berada di bawah SVM. Temuan ini menegaskan bahwa SVM lebih efektif dalam mengklasifikasikan sentimen pada data Twitter yang memiliki ragam bahasa dan konteks tinggi, serta memperkuat penggunaan algoritma tersebut pada analisis opini publik mengenai isu kesehatan seperti vaksin COVID-19 (Indriyani et al., 2022).

9. Penelitian yang dilakukan oleh Acuan Supian, Bagus Tri Revaldo, Nanda Marhadi, Lusiana Efrizoni, dan Rahmaddeni (2024) yang berjudul "*Perbandingan Kinerja Naïve Bayes dan SVM pada Analisis Sentimen Twitter Ibukota Nusantara* ", mengkaji analisis sentimen masyarakat terhadap isu pembangunan Ibukota Nusantara (IKN) dengan menggunakan data dari media sosial Twitter. Penelitian ini dilatarbelakangi oleh tingginya perhatian publik terkait pemindahan ibu kota negara, yang kemudian menimbulkan beragam opini positif maupun negatif di platform media sosial. Data yang digunakan merupakan kumpulan tweet yang diperoleh dari repositori publik *GitHub*, kemudian diproses melalui tahapan preprocessing seperti *cleaning*, *case folding*, *tokenizing*, *stopword removal*, *stemming*, dan normalisasi untuk menghasilkan teks yang siap dianalisis. Selanjutnya, data direpresentasikan menggunakan pembobotan TF- IDF sebelum masuk ke tahap klasifikasi.

   Penelitian ini membandingkan performa dua algoritma klasifikasi, yaitu Naïve Bayes Classifier dan Support Vector Machine (SVM), dengan menggunakan pembagian data 80% sebagai data latih dan 20% sebagai data uji. Evaluasi model dilakukan menggunakan confusion matrix untuk mengukur akurasi, presisi, recall, dan f1-score. Hasil pengujian menunjukkan bahwa SVM memberikan akurasi tertinggi sebesar 94%, sedangkan *Naïve Bayes* menghasilkan akurasi 91%. Temuan tersebut mengindikasikan bahwa SVM lebih efektif dalam mengklasifikasikan sentimen pada data Twitter yang

memiliki variasi bahasa dan konteks yang beragam.

Penelitian ini menyimpulkan bahwa SVM merupakan metode yang lebih unggul untuk analisis sentimen opini publik terhadap IKN, serta menegaskan pentingnya pemanfaatan machine learning dalam memahami respons masyarakat terhadap isu-isu kebijakan nasional (Supian et al., 2024).

10. Penelitian yang dilakukan oleh Mursyid dan Indriyanti (2024) yang dipublikasikan dalam JINACS (Journal of Informatics and Computer Science), dengan judul "*Perbandingan Akurasi Metode Analisis Sentimen Untuk Evaluasi Opini Pengguna Pada Platform Media Sosial (Studi Kasus: Twitter)".* Penelitian tersebut secara spesifik difokuskan pada upaya menganalisis sentimen opini masyarakat yang terekspresikan di media sosial Twitter pasca-Pemilihan Presiden (Pilpres) 2024 di Indonesia. Dalam kerangka metodologisnya, penelitian ini menguji perbandingan kinerja dari tiga algoritma klasifikasi utama—yakni *Naïve Bayes, Random Forest, dan Support Vector Machine* (SVM)—setelah melalui tahapan pre-processing dan pembobotan fitur menggunakan teknik Term Frequency-Inverse Document Frequency (TF-IDF). Total data yang diolah berasal dari Twitter, dengan jumlah sampel mencapai 806 entri.

    Hasil eksperimen yang dicapai dalam penelitian Mursyid dan Indriyanti (2024) menunjukkan bahwa terdapat variasi performa yang jelas di antara ketiga algoritma yang diuji. Secara kuantitatif, algoritma *Support Vector Machine* (SVM) terbukti memberikan performa yang superior dengan meraih tingkat akurasi tertinggi, yakni sebesar 74.79%. Perolehan ini mengungguli hasil yang dicapai oleh Random Forest dan Naïve Bayes. Temuan tersebut mengindikasikan bahwa metode SVM memiliki kemampuan yang paling optimal dalam mengidentifikasi dan mengklasifikasikan sentimen pada konteks data opini publik di media sosial. Dengan demikian, penelitian ini memberikan kontribusi penting dalam memfasilitasi pemahaman yang lebih mendalam mengenai fluktuasi dan polaritas sentimen masyarakat terkait isu politik spesifik, yang dapat menjadi landasan bagi para pemangku kepentingan dalam perumusan kebijakan atau strategi komunikasi (Mursyid & Indriyanti, 2024)

2.2 Landasan Teori

2.2.1 Platform

Dalam perkembangan teknologi informasi, istilah platform sering digunakan untuk menggambarkan suatu lingkungan atau sistem digital yang berfungsi sebagai dasar bagi interaksi antara pengguna, penyedia layanan, dan data. Secara umum, platform menyediakan seperangkat infrastruktur, aturan, serta antarmuka yang memungkinkan pihak-pihak di dalamnya saling berhubungan dan menciptakan nilai baru. Menurut Bartczak (2021), platform teknologi digital dapat diartikan sebagai kumpulan teknologi yang menjadi fondasi bagi terciptanya model bisnis inovatif dan interaksi berbasis data dalam suatu ekosistem digital (Bartczak, 2021).

Lebih lanjut, Jovanovic menjelaskan bahwa platform digital merupakan ekosistem sosio-teknis yang mengorkestrasi hubungan antara data, teknologi, pelaku, dan layanan untuk memungkinkan kolaborasi lintas sektor. Dengan demikian, platform tidak hanya dipahami sebagai perangkat lunak atau perangkat keras, melainkan juga sebagai sistem yang mengintegrasikan berbagai aktor dan sumber daya dalam satu wadah interaktif (Jovanovic et al., 2022).

Dalam konteks penelitian ini, istilah platform merujuk pada sebuah sistem berbasis *Python* yang dirancang untuk memfasilitasi pengguna dalam melakukan analisis sentimen terhadap berbagai komentar atau opini yang diperoleh dari sumber daring. Platform ini tidak hanya berfungsi sebagai wadah input data, tetapi juga mengintegrasikan proses pengolahan teks menggunakan metode *Support Vector Machine* (SVM) untuk mengidentifikasi kecenderungan sentimen positif, negatif, atau netral dari setiap komentar yang dimasukkan. Hasil analisis tersebut kemudian divisualisasikan dalam bentuk persentase atau grafik untuk memberikan gambaran umum mengenai persepsi publik terhadap suatu topik tertentu. Dengan demikian, platform yang dikembangkan bertujuan untuk menjadi sarana interaktif yang tidak hanya menampilkan hasil klasifikasi, tetapi juga mendukung proses pemahaman dan pengambilan keputusan berbasis data opini masyarakat.

Platform digital memiliki beragam bentuk dan fungsi tergantung pada tujuan serta aktor yang terlibat di dalamnya. Secara umum, Gawer dan Evans dalam penelitian yang dikutip oleh Köster, membagi platform digital menjadi tiga kategori utama, yaitu *transaction platforms, innovation platforms, dan integration platforms* (Beverungen et al., 2022). Klasifikasi ini membantu memahami peran dan mekanisme kerja platform dalam berbagai konteks, baik ekonomi, sosial, maupun teknologi.

Platform transaksional (*transaction platform*) merupakan jenis platform yang berfungsi sebagai perantara antara dua pihak atau lebih untuk melakukan pertukaran barang, jasa, atau informasi. Platform ini umumnya ditemukan pada layanan *e-commerce*, *ride-sharing*, atau pasar daring seperti *Tokopedia*, *Gojek*, dan *Grab*, yang memfasilitasi interaksi langsung antara penjual dan pembeli. Menurut Jovanovic et al. (2021), platform jenis ini berperan penting dalam meningkatkan efisiensi pasar dengan mengurangi biaya transaksi dan memperluas jangkauan interaksi digital (Jovanovic et al., 2022).

Lalu, platform inovasi (*innovation platform*) berfungsi sebagai infrastruktur teknologi yang memungkinkan pihak ketiga untuk mengembangkan layanan atau aplikasi tambahan di atas sistem yang telah disediakan. Contoh dari platform jenis ini adalah sistem operasi seperti *Android* dan *iOS*, yang memberikan ruang bagi pengembang untuk menciptakan aplikasi baru dengan memanfaatkan ekosistem yang ada.

Adapun platform integrasi (*integration platform*) merupakan kombinasi dari kedua jenis sebelumnya, di mana sistem tidak hanya menjadi tempat terjadinya transaksi, tetapi juga mendukung pengembangan layanan baru oleh pihak ketiga. Platform jenis ini biasanya digunakan oleh perusahaan besar untuk menggabungkan interaksi pengguna dengan pengembangan teknologi internal, seperti *Google Play* atau *Facebook Platform*.

Menurut Köster, platform integrasi memiliki peran strategis karena memungkinkan pengelolaan data lintas sistem sekaligus mendorong inovasi berkelanjutan dalam satu ekosistem yang saling terhubung (Beverungen et al., 2022).

Berdasarkan klasifikasi tersebut, penelitian ini memposisikan sistem yang dikembangkan dalam kategori *innovation platform* sekaligus memiliki karakteristik *integration platform*. Hal ini dikarenakan platform yang dirancang tidak hanya menyediakan ruang bagi pengguna untuk memasukkan komentar, tetapi juga mengintegrasikan proses analisis sentimen dan visualisasi hasil melalui algoritma

Support Vector Machine (SVM). Dengan demikian, platform ini tidak semata-mata berfungsi sebagai wadah interaksi, melainkan juga sebagai sarana pengolahan dan penyajian data berbasis teknologi kecerdasan buatan.

2.2.2 Analisis Sentimen

Analisis sentimen merupakan salah satu cabang dari *text mining* yang bertujuan untuk mengenali dan mengklasifikasikan opini, emosi, serta sikap pengguna terhadap suatu topik, produk, atau layanan melalui teks yang mereka hasilkan di berbagai platform digital. Komentar atau ulasan publik dalam jumlah besar sangat sulit dan memakan waktu jika harus dievaluasi secara manual, terlebih karena penilaian dalam bentuk *rating* angka kerap kali tidak merepresentasikan maksud sebenarnya dari teks yang dituliskan pengguna (Asri et al., 2022). Oleh karena itu, pendekatan analisis sentimen diperlukan untuk secara otomatis mengekstraksi dan mengklasifikasikan kecenderungan opini ke dalam polaritas tertentu. Menurut penelitian oleh Monika P, Chaitanya Kulkarni, Harish Kumar N, Shruthi S, dan Vani V yang berjudul *"Machine Learning Approaches for Sentiment Analysis: A Survey (2022)"*, analisis sentimen secara umum dilakukan dengan mengolah teks menggunakan algoritma pembelajaran mesin untuk mengidentifikasi kecenderungan positif, negatif, atau netral dari suatu pernyataan. Dalam konteks perkembangan teknologi saat ini, analisis sentimen memiliki peran penting karena dapat membantu organisasi memahami persepsi publik dan mendukung proses pengambilan keputusan berbasis data (Kulkarni et al., 2022).

Penelitian yang dilakukan oleh Yonatan Mamani-Coaquira dan Edwin Villanueva yang berjudul *"A Review on Text Sentiment Analysis With Machine Learning and Deep Learning Techniques (2024)"* menjelaskan bahwa metode pembelajaran mesin seperti *Naive Bayes*, *Support Vector Machine* (SVM), serta pendekatan berbasis *deep learning* banyak digunakan karena memiliki kemampuan dalam mengenali pola kompleks dari teks. Hasil penelitian tersebut juga menegaskan bahwa efektivitas metode SVM dalam mengklasifikasikan data teks masih tergolong tinggi dibandingkan metode klasik lainnya, terutama dalam dataset berukuran sedang hingga besar (Mamani-Coaquira & Villanueva, 2024).

Di Indonesia, penerapan analisis sentimen telah berkembang pesat dalam berbagai

bidang. Berdasarkan penelitian oleh Boy Setiawan yang berjudul "*A Review of Sentiment Analysis Applications in Indonesia Between 2023-2024"*, sebagian besar studi lokal menerapkan analisis sentimen untuk menilai opini masyarakat terhadap layanan publik, isu sosial, hingga produk komersial (Setiawan, 2025). Studi-studi ini menunjukkan bahwa pemanfaatan algoritma pembelajaran mesin mampu menghasilkan pemetaan opini masyarakat yang lebih objektif dan terukur dibandingkan metode survei manual.

Contoh penerapan yang lebih spesifik dapat ditemukan pada penelitian "*Aspect-Based Sentiment Analysis of Reviews for Pandawa Beach Using Naive Bayes and SVM Methods*" (Putri et al., 2025), di mana kedua metode tersebut digunakan untuk menganalisis ulasan wisatawan terhadap destinasi wisata. Hasilnya menunjukkan bahwa kombinasi pendekatan berbasis aspek dengan algoritma SVM mampu meningkatkan akurasi dalam menentukan sentimen setiap ulasan. Hal ini membuktikan bahwa SVM masih relevan digunakan untuk penelitian berbasis teks, termasuk untuk mengidentifikasi opini masyarakat pada konteks tertentu.

2.2.3 Text Mining dan Natural Language Processing (NLP)

Text mining merupakan proses penggalian informasi dari data berbasis teks yang tidak terstruktur dengan tujuan menemukan pola, tren, dan wawasan yang tersembunyi. Menurut penelitian berjudul "*Identifying Research Trends Using Text Mining Techniques: A Systematic Review"* (Vinay, 2024)*, text mining* berfungsi untuk mengekstraksi informasi penting dari kumpulan dokumen teks secara otomatis melalui tahapan pra-pemrosesan seperti tokenisasi, pembersihan data, dan ekstraksi fitur. Proses ini menjadi sangat relevan dalam era digital karena sebagian besar data yang dihasilkan manusia bersifat tekstual, baik dari media sosial, ulasan daring, maupun berita digital. Dengan demikian, text mining menjadi alat penting dalam mengubah data teks yang masif menjadi pengetahuan yang dapat dimanfaatkan untuk analisis lanjutan.

Sementara itu, *Natural Language Processing* (NLP) merupakan cabang dari kecerdasan buatan yang memungkinkan komputer untuk memahami dan menafsirkan bahasa manusia. Berdasarkan penelitian "*A Narrative Literature Review of Natural Language Processing"* (Schoene et al., 2022), NLP bertujuan menjembatani komunikasi antara manusia dan mesin melalui pengolahan bahasa alami. Proses dalam NLP meliputi

beberapa tahapan penting seperti *tokenization*, *stopword removal*, *stemming*, dan *part-of- speech tagging*. Setiap tahapan berfungsi untuk mengubah teks mentah menjadi representasi yang dapat dimengerti oleh algoritma komputasi. Dengan pendekatan ini, sistem mampu mengenali konteks serta makna kata dalam kalimat, yang menjadi fondasi bagi berbagai aplikasi seperti analisis sentimen, *chatbot*, dan sistem rekomendasi.

Selain itu, kombinasi antara text mining dan NLP terbukti mampu meningkatkan efektivitas analisis data teks. Penelitian oleh Yun Liu yang berjudul "*Role of Natural Language Processing in Document Understanding and Sentiment Analysis"* (Liu, 2024), menjelaskan bahwa penggunaan teknik NLP dalam proses *text mining* mampu meningkatkan akurasi klasifikasi sentimen hingga 85–90%, terutama ketika dikombinasikan dengan algoritma pembelajaran mesin seperti *Support Vector Machine* (SVM) atau *Naive Bayes*. Integrasi kedua teknik tersebut tidak hanya mempercepat proses analisis, tetapi juga memungkinkan sistem untuk memahami konteks kalimat secara lebih mendalam. Oleh karena itu, *text mining* dan NLP menjadi dua komponen yang saling melengkapi dalam pengembangan sistem analisis berbasis teks, termasuk platform analisis sentimen yang dirancang untuk memahami opini publik di berbagai bidang.

Selain itu, kombinasi antara text mining dan NLP terbukti mampu meningkatkan efektivitas analisis data teks. Penelitian oleh Yun Liu yang berjudul "*Role of Natural Language Processing in Document Understanding and Sentiment Analysis"* (Liu, 2024), menjelaskan bahwa penggunaan teknik NLP dalam proses *text mining* mampu meningkatkan akurasi klasifikasi sentimen hingga 85–90%, terutama ketika dikombinasikan dengan algoritma pembelajaran mesin seperti *Support Vector Machine* (SVM) atau *Naive Bayes*. Integrasi kedua teknik tersebut tidak hanya mempercepat proses analisis, tetapi juga memungkinkan sistem untuk memahami konteks kalimat secara lebih mendalam. Oleh karena itu, *text mining* dan NLP menjadi dua komponen yang saling melengkapi dalam pengembangan sistem analisis berbasis teks, termasuk platform analisis sentimen yang dirancang untuk memahami opini publik di berbagai bidang.

2.2.1 Term Frequency-Inverse Document Frequency (TF-IDF)

Metode *Term Frequency–Inverse Document Frequency* (TF-IDF) merupakan teknik pembobotan kata yang banyak digunakan dalam bidang *Text Mining* dan *Information Retrieval* untuk merepresentasikan dokumen teks dalam bentuk numerik. TF-IDF bekerja dengan mengukur tingkat kepentingan suatu kata dalam sebuah dokumen relatif terhadap keseluruhan kumpulan dokumen (*corpus*). Konsep dasar metode ini adalah bahwa kata yang sering muncul dalam suatu dokumen namun jarang muncul pada dokumen lain akan memiliki bobot yang lebih tinggi karena dianggap lebih merepresentasikan isi dokumen tersebut(Al-obaydy et al., 2022).

Secara matematis, pembobotan TF-IDF terdiri dari dua komponen utama, yaitu ***Term Frequency* (TF)** dan ***Inverse Document Frequency* (IDF)**. Nilai TF menunjukkan frekuensi kemunculan suatu term ttt dalam dokumen ddd, yang dihitung menggunakan persamaan:

$$\mathrm{TF}(\mathrm{t},\mathrm{d}) = \mathrm{f}_{\mathrm{t,d}} \tag{2.1}$$

Sementara itu, ***Inverse Document Frequency* (IDF)** digunakan untuk mengukur seberapa penting suatu kata dalam keseluruhan dokumen. Kata yang muncul pada banyak dokumen akan memiliki nilai IDF yang lebih kecil karena dianggap kurang informatif. Nilai IDF dihitung menggunakan persamaan:

$$\mathrm{IDF}(\mathrm{t}) = \log\left(\frac{N}{dft}\right) \tag{2.2}$$

di mana N merupakan jumlah total dokumen dalam corpus dan dft adalah jumlah dokumen yang mengandung kata t.

Bobot akhir TF-IDF diperoleh dengan mengalikan kedua komponen tersebut sehingga menghasilkan representasi numerik untuk setiap kata dalam dokumen:

$$\mathrm{TFIDF}(\mathrm{t},\mathrm{d}) = \mathrm{TF}(\mathrm{t},\mathrm{d}) \times \mathrm{IDF}(\mathrm{t}) \tag{2.3}$$

Hasil pembobotan ini kemudian digunakan untuk membentuk vektor fitur yang merepresentasikan dokumen dalam ruang fitur berdimensi tinggi, sehingga dapat digunakan sebagai input dalam proses klasifikasi teks menggunakan algoritma pembelajaran mesin seperti *Support Vector Machine* (SVM)

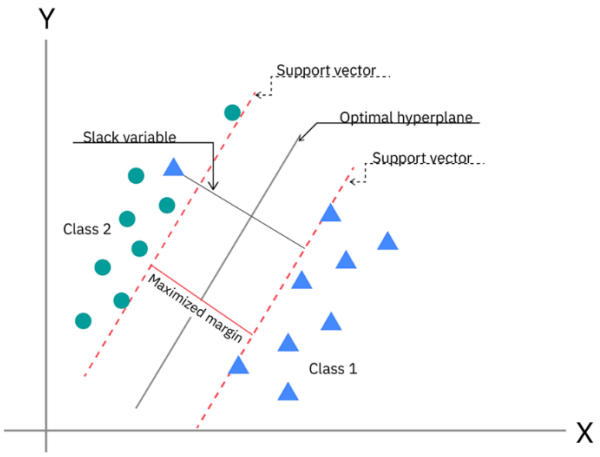
2.2.2 Support Vector Machine (SVM)

Support Vector Machine atau SVM merupakan salah satu algoritma pembelajaran mesin yang paling sering digunakan untuk tugas klasifikasi karena kemampuannya dalam memisahkan data antar kelas secara jelas. Algoritma ini bekerja dengan mencari sebuah *hyperplane* terbaik yang memaksimalkan margin atau jarak antar titik-data terdekat dari kelas berbeda yang biasa disebut vektor dukung sehingga generalisasi model terhadap data baru menjadi lebih baik. Sebagaimana dieksplorasi dalam studi "*Machine Learning Applications based on SVM Classification: A Review*" (Abdullah & Abdulazeez, 2021) yang menjelaskan bahwa SVM telah diterapkan di berbagai domain termasuk pengenalan teks dan pengenalan pola, karena sifat margin maksimumnya yang menciptakan batas keputusan yang kuat untuk klasifikasi.

Tujuan utama dari *Support Vector Machine* (SVM) adalah membangun suatu metode pembelajaran komputasi yang efektif dalam menentukan *hyperplane* pemisah terbaik pada ruang fitur berdimensi tinggi. *Hyperplane* tersebut dirancang untuk memaksimalkan margin antar kelas sehingga menghasilkan batas keputusan yang optimal. Dalam implementasinya, SVM menyediakan beberapa pendekatan berbasis kernel, baik untuk kasus linear maupun non-linear. Pada SVM linear, diasumsikan bahwa data dapat dipisahkan secara langsung menggunakan garis atau bidang pemisah linear. Namun, pada banyak kasus nyata, pola distribusi data tidak dapat dipisahkan secara linear. Oleh karena itu, SVM *non-linear* memanfaatkan fungsi *kernel* untuk mentransformasikan data ke dalam ruang fitur berdimensi lebih tinggi sehingga memungkinkan pemisahan yang lebih efektif.

Fungsi *kernel* dalam SVM berperan sebagai mekanisme transformasi matematis yang mengubah representasi data dari ruang asli berdimensi rendah ke ruang fitur yang lebih kompleks. Dengan pendekatan ini, data yang sebelumnya saling tumpang tindih dalam ruang awal dapat menjadi terpisah secara *linear* pada ruang hasil transformasi. Secara konseptual, *kernel* menjadi komponen inti dalam proses pembelajaran SVM karena menentukan bagaimana pola hubungan antar data direpresentasikan dalam model. Berlandaskan teori pembelajaran statistik, setiap jenis *kernel* dirancang untuk

menangani karakteristik distribusi data tertentu, sehingga pemilihan *kernel* yang tepat sangat berpengaruh terhadap performa klasifikasi yang dihasilkan.. Sebagai contoh, penelitian "*Combination of Lexical Resources and Support Vector Machine for Film Sentiment Analysis*" (Agustina & Putri, 2024) menunjukkan bahwa model SVM dengan pendekatan *kernel* yang tepat dapat mencapai akurasi sekitar 85% dalam klasifikasi teks yang melibatkan aspek leksikal dan sentimen pengguna.



Gambar 2.1 Cara Kerja Metode SVM (Sumber : ibm.com)

Keunggulan *Support Vector Machine* (SVM) dibandingkan dengan metode klasifikasi lainnya seperti *Naïve Bayes* maupun *K-Nearest Neighbors* terletak pada kemampuannya dalam menangani data berdimensi tinggi serta ketahanannya terhadap tingkat kebisingan yang relatif tinggi pada data teks. Stabilitas performa tersebut menjadikan SVM banyak digunakan dalam tugas klasifikasi berbasis teks. Sebagai contoh, dalam penelitian berjudul "*Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store*" (Ilmawan & Mude, 2025), metode SVM memperoleh tingkat akurasi yang lebih tinggi (±81,46%) dibandingkan *Naïve Bayes* (±75,41%) pada data ulasan berbahasa Indonesia. Temuan ini semakin memperkuat relevansi penggunaan SVM dalam penelitian yang berfokus pada klasifikasi sentimen teks.

Secara konseptual, SVM dapat diklasifikasikan menjadi dua jenis, yaitu SVM

linear dan SVM *non-linear*, bergantung pada karakteristik distribusi data. SVM linear digunakan ketika data dapat dipisahkan secara langsung dalam ruang fitur, sedangkan SVM *non-linear* memanfaatkan fungsi *kernel* untuk memetakan data ke ruang berdimensi lebih tinggi agar dapat dipisahkan secara linear. Dalam konteks analisis komentar publik yang bersifat tidak terstruktur serta mengandung variasi bahasa informal dan kata tidak baku, penggunaan SVM dengan pemilihan *kernel* yang sesuai menjadi pendekatan yang relevan untuk mengklasifikasikan sentimen ke dalam kategori positif, negatif, dan netral. Dengan demikian, algoritma SVM dipandang sebagai metode yang tepat untuk menghasilkan model klasifikasi yang memiliki akurasi dan kemampuan generalisasi yang baik pada data teks media sosial.

2.2.2.1 Hyperplane

Hyperplane adalah sebuah bidang datar pemisah yang digunakan sebagai batas keputusan (decision boundary) untuk mengklasifikasikan data. Secara konseptual, ia adalah generalisasi dari sebuah garis atau bidang untuk ruang dengan dimensi berapapun. Di ruang 2-dimensi, sebuah *hyperplane* adalah garis lurus. Di ruang 3-dimensi, ia adalah bidang datar. Untuk ruang dengan 4 atau lebih dimensi, pemisah datar ini disebut *hyperplane*.

Dalam algoritma *Support Vector Machine* (SVM), tujuan utamanya adalah untuk menemukan *hyperplane* optimal yang tidak hanya memisahkan data ke dalam kelas-kelas yang berbeda, tetapi juga memiliki margin atau jarak terbesar ke titik data terdekat dari setiap kelas (Sasidharan, 2026).

$$w \cdot x - b = 0 \tag{2.4}$$

Dimana :

w = vektor yang menentukan orientasi atau kemiringan dari hyperplane

x = mempresentasikan suatu titik data

b = bias

2.2.2.2 Klasifikasi Kelas Positif atau Negatif pada SVM

Support Vector machine mampu menentukan klasifikasi dari hasil dari suatu kelas yang bernilai positif maupun negative, jika bernilai positif maka akan dikategorikan kelas yang memiliki nilai > 1 atau nilainya tidak mengalamai minus dan juga jika suatu klasifikasi mendapatkan angka < 1 atau menghasilkan nilai minus maka akan diklasifikasikan negatif maka akan dikategorikan dalam kelas yang negatif (Amarappa, 2021), Adapun perumusan klasifikasi yang dilakukan oleh *Support Vector Machine* dapat dirumuskan sebagai berikut :

$$f(x) = sign(w.x + b) \tag{2.5}$$

untuk memaksimalkan margin pada *Support vector machine,* penyelesaian permasalahan optimasi sebagai berikut :

$$\underset{w,b}{Min} = \frac{1}{2}||w||^2 \tag{2.6}$$

dengan kendala :

$$y_i = (w.xi. + b) \geq 1 \tag{2.7}$$

2.2.2.3 Perhitungan Optimasi *Soft-margin* pendekatan Regularisasi C

Namun. dikarenakan pada penelitian ini, data teks pada umumnya tidak terpisah secara linear, dan sehingga pada penelitian ini menggunakan pendekatan *soft-margin* dengan pendekatan parameter Regularisasi C yang berfungsi sebagai mengontrol *trade-off* antara margin yang lebar dan kesalahan klasifikasi pada data latih.

$$\min_{w,b} = \frac{1}{2}||w||^2 + C\sum \xi i \tag{2.8}$$

dengan kendala :

$$y_i(w.xi. + b) \geq 1 - \xi i \tag{2.9}$$

dimana :

min w,b = sistem mencari nilai **w (bobot)** dan **b (bias)** yang meminimalkan fungsi.

$||w||^2$ = norm kuadrat dari vektor bobot

C = Parameter Reguralisasi

ξi = Mengukur seberapa besar suatu data melanggar batas margin.

2.2.2.4 Transformasi ke Dual Form

Transformasi ke bentuk dual pada *Support Vector Machine* (SVM) berfungsi untuk menyederhanakan proses optimasi sekaligus memungkinkan penerapan fungsi kernel pada data yang tidak terpisah secara linear. Pada bentuk primal, model secara langsung menghitung vektor bobot dalam ruang fitur asli. Namun, pada data berdimensi tinggi seperti representasi TF-IDF, perhitungan bobot secara eksplisit menjadi kurang efisien dan sulit diterapkan untuk pemetaan non-linear(Sun & Kwon, 2025). Adapun Formulasi Transformasi ke Dual Form pada penelitian ini adalah berikut :

$$w = \sum_{i=1}^{n} \alpha 1 y 1 \tag{2.10}$$

dimana :

α1 = Lagrange Multiplier

yi = Label Kelas

xi = Support Vector

Kemudian, Subtitusi persamaan *Dual Form* tersebut ke dalam fungsi klasifikasi kelas yang kemudian menghasilkan persamaan :

$$f(x) = sign\left(\sum_{i=1}^{n} \alpha 1 y 1(xi + x) + b\right) \quad (2.11)$$

2.2.2.5 Kernel linear

Kernel linear merupakan fungsi *kernel* paling sederhana dalam algoritma *Support Vector Machine* (SVM) yang bekerja dengan membentuk *hyperplane* pemisah secara langsung pada ruang fitur asli tanpa melakukan transformasi ke dimensi yang lebih tinggi. Secara matematis, *kernel* linear menghitung hasil perkalian titik (dot product) antara dua vektor fitur, sehingga fungsi keputusan yang dihasilkan bersifat linear terhadap data input. Pendekatan ini efektif digunakan ketika distribusi data dapat dipisahkan secara linear atau ketika jumlah fitur sangat besar, seperti pada representasi teks menggunakan *TF-IDF*. Pada kasus data teks berdimensi tinggi, meskipun jumlah fitur sangat banyak, pola pemisahan antar kelas sering kali sudah cukup representatif dalam ruang asli, sehingga kernel linear mampu menghasilkan batas keputusan yang optimal tanpa kompleksitas komputasi tambahan.

2.2.2.6 Kernel Non Linear

Kernel *non linear* pada *Support Vector Machine (SVM)* sering digunakan untuk memetakan data ke ruang fitur yang berdemensi lebih tinggi agar data-data yang tidak dapat dipisahkan secara linear dapat dipisahkan dengan menggunakan hyperplane linear di ruang baru. Adapun jenis dari kernel yang digunakan *Support Vector Machine (SVM)* berikut jenis-jenisnya:

1) Radial Basis Function (RBF) / Gaussian Kernel

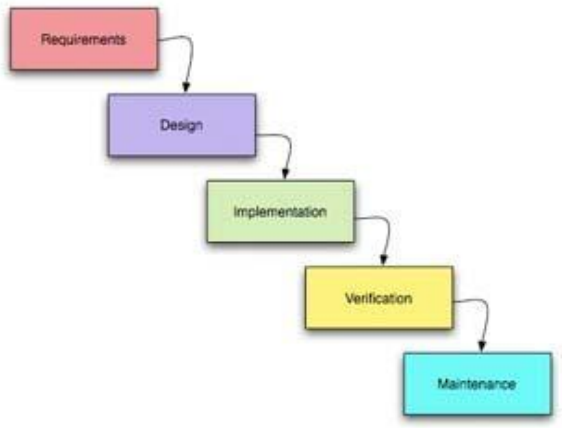
Kernel Radial Basis Function pada *Support Vector Machine (SVM)* dikenal dengan fleksibilitasnya yang dimana menggunakan eksponensial berdasarkan jarak Euclidean antar data.

$$K(x, y) = \exp(-\gamma \parallel x - y \parallel 2) \quad (2.12)$$

2.2.3 Metode Waterfall

Metode *Waterfall* adalah model yang paling umum digunakan untuk tahap pengembangan. Model air terjun (waterfall) sering disebut sebagai model sekuensial linier (squential linear) atau alur hidup klasik (classic cycle). Model air terjun ini menyediakan pendekatan alur hidup perangkat lunak dimulai dengan analisis, desain, pengkodean, pengujian, dan tahap pendukung (Supiyandi et al., 2022).

Metode ini pada setiap fase menghasilkan dokumentasi yang menjadi acuan untuk tahap selanjutnya, sehingga proses pengembangan berjalan secara terstruktur dan terdokumentasi dengan baik. Model ini umumnya digunakan ketika kebutuhan sistem telah terdefinisi secara jelas sejak awal dan perubahan selama proses pengembangan relatif minimal (Senarath, 2021).



Gambar 2. 2 Tahapan Metode Waterfall (Sumber: Ekrut Media)

Tahapan dari Metode *Waterfall* dapat dilihat dari gambar di atas dengan rincian sebagai berikut:

1) **Requirements**

Tahap ini merupakan langkah awal di mana pengembang berkomunikasi dengan pengguna untuk memahami masalah, mengumpulkan kebutuhan sistem,

dan mendefinisikan fitur-fitur perangkat lunak yang akan dibangun.

2) **Design**

Tahap ini berfokus pada perancangan struktur data, arsitektur perangkat lunak, representasi antarmuka, dan detail prosedur algoritma. Tujuannya adalah memberikan gambaran teknis sebelum penulisan kode dimulai.

3) **Implementation**

Tahap ini menggabungkan pembuatan kode (*code generation*) dan pengujian unit. Desain yang telah dibuat diterjemahkan ke dalam bahasa pemrograman yang dapat dimengerti oleh komputer.

4) **Verification**

Tahap di mana perangkat lunak diserahkan kepada pengguna untuk diuji. Pengujian dilakukan untuk memastikan seluruh fungsi berjalan dengan benar dan meminimalisir kesalahan (*error*) sebelum sistem digunakan secara penuh.

5) **Maintenance**

Tahap ini dilakukan setelah perangkat lunak digunakan, meliputi perbaikan kesalahan yang baru ditemukan, adaptasi terhadap lingkungan baru, atau peningkatan fungsionalitas.

2.2.4 Framework Streamlit

Streamlit adalah sebuah framework *open-source* berbasis bahasa pemrograman *Python* yang digunakan untuk membangun aplikasi web interaktif secara cepat, khususnya untuk kebutuhan *data science* dan *machine learning*. *Streamlit* memungkinkan pengembang mengubah skrip *Python* menjadi aplikasi web hanya dengan menambahkan beberapa perintah sederhana tanpa perlu menggunakan *HTML*, *CSS*, atau *JavaScript* secara langsung.

Dari sisi fungsionalitas, *Streamlit* menyediakan berbagai komponen antarmuka (widget) seperti *text input*, *button*, *slider*, *selectbox*, *checkbox*, serta *file uploader* yang dapat diintegrasikan langsung dengan proses komputasi di dalam *script*. Selain itu, *Streamlit* mendukung visualisasi data menggunakan berbagai *library Python* seperti *Matplotlib*, *Plotly*, atau *Altair*, sehingga hasil analisis dan model prediktif dapat ditampilkan dalam bentuk grafik maupun tabel secara interaktif.

Secara arsitektural, *Streamlit* menerapkan *paradigma reactive programming*

berbasis *script rerun*. Setiap interaksi pengguna terhadap komponen input akan memicu eksekusi ulang seluruh skrip dari awal hingga akhir (top-down execution model). Mekanisme ini memastikan bahwa *state* aplikasi dan *output* yang ditampilkan selalu konsisten dengan *input* terbaru (Akkem, 2023). Setiap kali pengguna berinteraksi dengan komponen antarmuka (seperti tombol, input teks, atau upload file), skrip akan dieksekusi ulang dari atas ke bawah dengan mempertahankan *state* tertentu. Hal ini memungkinkan pembaruan tampilan secara dinamis berdasarkan input pengguna.