BAB II

KAJIAN PUSTAKA

2.1 Teori Dasar

Teori dasar dalam penelitian ini membahas sejumlah konsep yang memiliki keterkaitan langsung dengan topik yang diangkat. Pembahasan teori ini menjadi aspek penting guna memperkuat validitas dan landasan konseptual penelitian. Teori-teori yang diuraikan mencakup definisi serta istilah-istilah umum yang relevan dengan proses dan metode yang digunakan dalam penelitian. Adapun teori yang menjadi dasar dalam penelitian ini meliputi: (1) *Machine Learning*, (2) Pengolahan Bahasa Alami, (3) *Text Mining*, (4) *Sentiment Analysis*, (5) *Text Processing*, (6) Klasifikasi Data, (7) *SVM*, (8) Kepuasan Pelanggan, (9). Aplikasi Grab, (10) Ulasan Pengguna, dan (11) *Python*.

2.1.1 *Machine Learning*

Machine learning atau pembelajaran mesin merupakan segala hal tentang belajar, berpikir, dan bertindak berdasarkan data. Ini dilakukan dengan cara membangun program komputer yang memproses data, mengekstrak informasi berguna, membuat prediksi tentang permasalahan yang tidak diketahui, dan menyarankan tindakan yang harus diambil atau keputusan yang harus diambil. Hal yang mengubah analisis data menjadi pembelajaran mesin adalah bahwa proses ini otomatis dan program komputer dipelajari dari data. Ini berarti bahwa program komputer generik digunakan, yang disesuaikan dengan keadaan khusus aplikasi dengan menyesuaikan pengaturan program berdasarkan data pelatihan yang

diamati, yang disebut data pelatihan. Oleh karena itu, bahwa pembelajaran mesin adalah cara pemrograman berdasarkan contoh (Faiza et al., 2022).

Penelitian terbaru mengungkapkan bahwa *machine learning* terbagi menjadi tiga kategori: *Supervised Learning*, *Unsuversiped Learning*, dan *Reinforcement Learning*. Teknik yang digunakan oleh *Supervised Learning* ialah metode klasifikasi dengan menggunakan data yang terlebih dahulu diberikan label sebagai data *train* untuk mengklasifikasikan kelas yang tidak dikenal. Sementara itu, teknik *Unsupervised Learning* seringkali disebut sebagai *cluster*. Hal ini dikarenakan tidak ada kebutuhan untuk melakukan pemberian label dalam kumpulan data di dalam kelas yang telah ditentukan. Sedangkan *Reinforcement Learning* adalah sebuah kategori *Machine Learning* yang berada diantara *Supervised* dan *Unsupervised Learning*, teknik ini bekerja dalam lingkungan yang dinamis dengan menggunakan konsep untuk mencapai tujuan tanpa adanya pemberitahuan dari komputer secara khusus apabila tujuan telah terselasaikan (Baharuddin & Tjahyanto, 2022)

2.1.1.1 Jenis-Jenis Machine Learning

Machine Learning (ML) atau pembelajaran mesin merupakan cabang dari kecerdasan buatan (*Artificial Intelligence*) yang memungkinkan sistem komputer untuk belajar dari data dan melakukan prediksi atau pengambilan keputusan tanpa perlu diprogram secara eksplisit. Komputer mempelajari pola atau relasi dari data historis (data pelatihan) dan menggunakan pengetahuan tersebut untuk memproses data baru. Salah satu karakteristik utama dari machine learning adalah sifatnya yang

adaptif, yaitu algoritma dapat memperbaiki kinerjanya seiring dengan bertambahnya data yang diproses (Diana et al., 2023).

Secara umum, *machine learning* terbagi ke dalam beberapa kategori berdasarkan cara mesin belajar dari data. Tiga kategori utama yang banyak digunakan dalam literatur dan aplikasi nyata adalah *Supervised Learning, Unsupervised Learning,* dan *Reinforcement Learning,* Di samping ketiga kategori utama ini, dikenal juga pendekatan lain seperti *Semi-Supervised Learning dan Self-Supervised Learning*, yang merupakan hibrida dari dua pendekatan utama atau pendekatan baru dalam penelitian terbaru (Diana et al., 2023).

1. *Supervised Learning* (Pembelajaran Terawasi)

   Supervised Learning adalah metode pembelajaran mesin di mana model dilatih menggunakan data yang sudah diberi label. Artinya, setiap input memiliki output yang diketahui. Model akan belajar memetakan input ke output tersebut. Pendakatan ini dapat diterapkan untuk melakukan klasifikasi sentiment ulasan pengguna. Misalnya, ulasan-ulasan pengguna yang sudah dilabeli sebagai “positif” atau “negatif” digunakan untuk melatih model klasifikasi seperti Support Vector Machine (SVM). Model ini kemudian mampu mengkategorikan ulasan baru berdasarkan pola yang telah dipelajari dari data historis. Pendekatan ini sangat efektif ketika tersedia data dengan label yang akurat dan representatif.

2. *Unsupervised Learning* (Pembelajaran Tak Terawasi)

   Dalam *Unsupervised Learning*, data tidak memiliki label, dan model berusaha menemukan struktur atau pola tersembunyi di dalam data.

Pendekatan ini dapat digunakan untuk segmentasi pengguna berdasarkan perilaku penggunaan. Misalnya, dengan menggunakan algoritma K-Means Clustering, data aktivitas pengguna seperti frekuensi penggunaan GrabCar, GrabFood, atau pembayaran GrabPay dapat dikelompokkan menjadi beberapa segmen pengguna, seperti pengguna aktif, pengguna GrabFood loyal, atau pengguna yang hanya sesekali menggunakan layanan. Hasil segmentasi ini dapat digunakan untuk meningkatkan personalisasi promosi atau penyusunan strategi pemasaran yang lebih tepat sasaran.

3. *Reinforcement Learning* (Pembelajaran Penguatan)

   Reinforcement Learning melibatkan agen (*agent*) yang belajar dari lingkungan melalui trial-and-error untuk mencapai tujuan. Agen menerima reward (penghargaan) atau penalty (hukuman) berdasarkan tindakannya. Contoh penerapan metode ini adalah optimasi rute pengiriman J&T Express atau Shopee Express. Sistem dapat belajar secara dinamis dari kondisi lalu lintas, waktu tempuh, dan efisiensi biaya, untuk memilih rute pengiriman yang optimal. Pendekatan ini sangat berguna untuk skenario yang bersifat dinamis dan kompleks, kondisi yang mengharuskan keputusan harus diambil secara real-time.

4. *Semi-Supervised Learning*

   Metode ini adalah kombinasi dari supervised dan unsupervised learning, di mana sebagian data diberi label dan sebagian tidak. Biasanya digunakan ketika pelabelan data mahal atau sulit dilakukan. Metode *semi-supervised learning* dapat digunakan untuk klasifikasi sentimen ulasan pengguna ketika

hanya sebagian data yang telah diberi label. Dengan menggunakan algoritma seperti *self-training* atau *label propagation*, model dapat belajar dari data berlabel dan menerapkan hasil pembelajaran tersebut untuk mengklasifikasikan data tak berlabel, sehingga meningkatkan performa klasifikasi tanpa memerlukan proses pelabelan menyeluruh.

5. *Self-Supervised Learning*

   Metode ini masih berkembang dan sering digunakan dalam bidang NLP dan computer vision. Model belajar dari data tanpa label dengan membuat tugas prediksi internal dari data itu sendiri. Metode ini dapat diterapkan pada layanan Grab, metode ini dapat digunakan untuk mengembangkan chatbot layanan pelanggan yang cerdas. Misalnya, model seperti BERT dilatih dengan memprediksi kata-kata yang hilang dari kalimat (masked language modeling), sehingga mampu memahami konteks bahasa alami dalam ulasan atau pertanyaan pengguna. Dengan pendekatan ini, sistem dapat memberikan respon yang lebih akurat, relevan, dan sesuai konteks, meskipun tidak bergantung pada data yang telah dilabeli secara manual.

2.1.2 Pengolahan Bahasa Alami

Pengolahan bahasa alami atau yang dikenal dengan istilah *Natural Language Preprocessing* (*NLP*) merupakan cabang dari ilmu komputer dan linguistik, serta bagian penting dari *text mining* dan merupakan sub bidang dar kecerdasan buatan. Dalam fokusnya NLP mempelajari tentang bagiamana membuat komputer dapat memahami dan mengerti makna dar bahasa manusia. Hal ini dapat dicapa dengan cara mengubah bahasa manusia ke dalam dokumen dan menjadikan dokumen

tersebut lebih formal agar lebih mudah dimanipulasi oleh program komputer dan memberikan respon yang sesuai. Tujuan NLP ialah mengatasi batasan dalam mengolah teks berbasis sintaks, (kerap disebut sebagai 'perhitungan kata') dan melanjutkan ke pemahaman yang lebih mendalam dengan memproses bahasa alami denagan memperhatikan aspek semantik, tata bahasa, dan konteks (Kelviandy & Ihsan, 2022).

Implementasi NLP membutuhkan beberapa komponen utama untuk melakukan analisis makna kalimat (Amin et al., 2023), yaitu:

a. *Parser*

Kata-kata yang telah di *input* pengguna akan dilakukan identifikasi oleh sistem, yang selanjutnya akan mengorganisasikan sesuai dengan tata bahasa. Pengelompokan kata yang sering dilakukan adalah dengan memisahkan kata-kata berdasarkan kategori, mulai dari subjek, kata kerja, kata keterangan, kata benda, kata sifat, kata hubung, atau frasa kata benda-keterangan.

b. Sistem Representasi Pengetahuan

Sistem Representasi Pengetahuan adalah sistem yang menginspeksi keluaran dari *parser* untuk mengidentifikasi maknanya, yang kemudian akan menghasilkan keluaran sesuai dengan masukan yang diberikan oleh *parser*.

c. *Output Translator*

Output translator adalah sebuah terjemahan yang menggambarkan pengetahuan sistem dan dapat berupa respons terhadap bahasa alami atau keluaran khusus yang sesuai dengan program komputer.

Implementasi NLP saat ini telah mengalami pertumbuhan pesat dan merambah ke berbagai sektor dan industri yang berbeda. Sebagai contoh, NLP telah digunakan dalam berbagai aplikasi seperti sistem penjawab pesan otomatis atau *chatbot*, *machine translation*, *spam filtering*, serta identifikasi bahasa. Dalam mengolah teks dan bahasa, NLP mengikuti beberapa tingkatan pengolahan yang melibatkan analisis fonetik untuk pemahaman suara, analisis morfologi untuk memahami struktur kata, analisis sintaksis untuk memahami struktur kalimat, analisis semantik untuk memahami makna kata dan kalimat, serta pemanfaatan pengetahuan leksikal dan aspek pragmatik dalam memahami bahasa dalam konteks yang lebih luas.

Pertumbuhan dan pengembangan dalam bidang NLP terus membuka pintu untuk inovasi baru dan peningkatan kemampuan komunikasi komputer dengan bahasa manusia. Aplikasi NLP yang semakin berkembang telah membantu meningkatkan efisiensi dan kualitas komunikasi manusia dengan komputer, serta memberikan solusi yang lebih canggih dalam berbagai industri seperti teknologi informasi, perbankan, e-commerce, hingga layanan pelanggan. Dengan terus berkembangnya teknologi dan pemahaman terhadap bahasa manusia, NLP akan terus menjadi bidang penelitian dan pengembangan yang menarik di masa depan.

2.1.3 *Text Mining*

Text mininig adalah proses penambangan teks untuk mengekstrak informasi yang berguna dari sumber data berukuran besar yang tidak terstruktur melalui identifikasi dan eksplorasi pola tertentu (Pranata et al., 2022). Dataset yang digunakan dalam *text mining* adalah dokumen-dokumen teks. *Text mining* merupakan pengembangan dari penelitian tentang *data mining* sehingga terjadi kesamaan arstitektur dalam melakukan pengerjaan *data mining* dan *text mining*. Contoh kesamaannya adalah *text mining* dan *data mining* masing-masing mempunyai tahapan *preprocessing*, *algortima*, pola, hasil, serta *tools* atau alat bantu lain yang hampir sama. Dalam pengerjaannya *text mining* mengimplementasikan banyak pola untuk mendapatkan hasil yang lebih akurat sehingga untuk *text mining* tahapan *preprocessing* menjadi sangat penting karena di tahapan inilah dilakukan identifikasi dan ekstraksi berbagai fitur yang mewakili sebuah dokumen. Tugas dari *text mining* adalah pengkategorisasian teks (*text categorization*) dan pemngelompokan teks (*text clustering*) (Pranata et al., 2022).

Tujuan utama dari *text mining* ialah memperoleh informasi yang sesuai dan relevan dari data yang diolah. Akan tetapi implementasinya bukan tanpa halangan, salah satu permasalahan yang dihadapi dalam implementasi *text mining* adalah dataset yang akan digunakan berdimensi tinggi, memiliki pola *noise* yang berbeda-beda, data yang tidak terstruktur atau semistruktur, tidak kompleks adan tidak lengkap. Seringkali data yang akan diolah tidak tersatandarisasi, berbagai jenis bahasa bercampur dalam satu data, serta translasi yang tidak akurat dan merubah

nuansa teks (Hardi et al., 2021). Pada *text mining* terdapat fitur-fitur pendukung yang digunakan antara lain:

1. ***Character*:** Komponen individual yang ada dalam bagian *text mining* seperti huruf, angka, karakter spesial dan spasi. Karakter ini merupakan level yang paling tinggi dalam pembentukan *semantik feature* seperti kata, *terms* dan konsep. Akan tetapi analisa *character-based* jaranf sekali digunakan pada saat analisis sentimen.
2. ***Words*:** Kata-kata yang dipilih secara langsung dari dataset yang menjadi dasar atau tingkat dasar semantik.
3. ***Terms*:** Kata tunggal atau kata jamak yang terpilih secara langsung dari kopus. Representasi *term-based* dari dokumen tersusun dari subset *term* dan dokumen.
4. ***Concept*:** Merupakan *feature* yang dibuat dari sebuah dokumen secara manual, *rule-based*, ataupun menggunakan metodologi lain.

2.1.4 *Sentiment Analysis*

Sentiment analysis atau analisis sentimen adalah sebuah studi komputas yang berhubungan dengan sikap, emosi, pendapat, penilaian, serta opini dari sekumpulan teks yang tujuan dari analisis sentimen dilakukan adalah untuk melakukan ekstraksi, identifkasi atau menemukan karakteristik sentimen dalam unit teks menggunakan metode NLP (*Natural Language Processing*), statistik, atau *machine learning* (Manullang & Prianto, 2023).

Analisis sentimen merupakan proses klasifikasi dokumen tekstual yang dipisah ke dalam beberapa kelas atau label. Kelas sentimen dapat berupa positif, negatif, atau netral. Topik ini menjadi cukup banyak dikembangkan dalam beberapa waktu kebelakang dikarenakan pengaruh dan manfaat dari penelitian ini sangat besar. Akan tetapi meskipun bergitu, dalam implementasinya berbagai hamatan harus dihadapi seperti tingkat ambiguitas yang tinggi dari dataset, tidak adanya intonasi dalam sebuah teks, serta perkembangan bahasa.

Manfaat lain dari analisis sentimen adalah dapat melakukan analisa tentang sebuah produk dan dilakukan secara cepat. Sehingga dapat digunakan sebagai alat bantu untuk melihat respon pengguna terhadap produk tersebut, hasil dari analisis sentimen dapat menjadi bahan pertimbangan optimalisasi yang sesuai dengan langkah strategis. Pada tugas akhir ini penulis melakukan penelitian tentang analisis sentimen terhadap layanan ojek *online* dengan menggunakan *Support Vector Machine* (Alfandi Safira & Hasan, 2023).

2.1.5 *Text Preprocessing*

Tahapan pra-proses dalam analisis sentimen merupakan tahapan awal dan paling penting yang harus dilakukan agar dataset dapat menjadi lebih terstruktur dan bisa digunakan ditahapan berikutnya (Arsi & Waluyo, 2021). Tahapan atau langkah ini akan merubah bentuk dan struktur awal data menjadi bentuk yang lebih baik. Jadi tujuan sebenarnya dari dilakukan *text preprocessng* adalah untuk membersihkan data yang akan digunakan saat klasifikasi sentimen (Pravina et al., 2019). Secara umum tahapan pra-pemrosesan terdiri sebagai berikut.

1. *Data Cleaning*

Cleaning atau pembersihan data adalah proses pembersihan dokumen dari karakter-karakater yang tidak diperlukan untuk mengurangi derau atau *noise* di dalam dataset. Karakter-karakter yang dihapus dalam dokuemn biasanya berupa emotikon, simbol-simbol, dan angka.

2. *Case Folding*

Case Folding adalah proses untuk merubah keseluruhan teks di dalam dokumen menjadi huruf kecil.

3. Normalisasi Kata Tidak Baku

Salah satu permasalahan dalam *text mining* adalah penggunaan kata tidak baku oleh masyarakat, misalnya penggunaan kata *slang*, bahasa daerah, dan kata-kata yang ditulis dengan disingkat. Hal ini harus dinormalisasi dengan menggunakan kamus khusus yang berisikan kata-kata yang tidak baku dengan kata-kata bakunya.

4. Penghapusan *Stopword*

Stopwords adalah kata-kata yang terdapat dalam dokumen akan tetepi memiliki bobot sentimen yang rendah sehingga diperlukan sebuah kamus khusus yang berisikan kata-kata umum untuk dihapus. Pada penelitian ini penulis akan menggunakan kamus *stopword* dari penelitian (Pebiana et al., 2022) dengan jumlah kata sebanyak 205 kata.

2.1.6 Klasifikasi Data

Klasifikasi merupakan suatu proses yang bertujuan untuk mengidentifikasi atau membedakan konsep atau kelas data dengan maksud untuk memprediksi kelas dari sebuah objek yang kelasnya belum diketahui. Dalam tahap klasifikasi, sekelompok data yang disebut sebagai *training set* diberikan, yang terdiri dari berbagai atribut, baik dalam bentuk kontinu maupun kategori. Salah satu atribut dalam data ini digunakan untuk menentukan kelas dari setiap rekaman (Arisusanto et al., 2023).

Tujuan utama dari melakukan klasifikasi adalah untuk menghasilkan sebuah model dari set data pelatihan yang dapat memisahkan entri-entri tersebut ke dalam kategori atau kelas yang sesuai. Model yang telah dibuat dapat digunakan untuk melakukan klasifikas data atau entri baru yang kelasnya belum diketahui. Tujuan kedua dari melakukan klasifikasi adalah membantu pengambilan suatu keputusan dengan memprediksi suatu kasus didadasarkan hasil klasifikasi yang diperoleh.

Tahapan proses klasifikasi data terdiri dari pembelajaran/pembangunan model dan proses klasifikasi. Pembelajaran atau pembangunan model adalah proses dimana tiap data di dalam data analisis berdasarkan nilai-nilai atributnya dengan menggunakan suatu algoritma klasifikasi untuk mendapatkan model. Selanjutnya ketika model sudah selesai dibangun maka dilakukan tahapan selanjutnya yaitu proses klasifikasi, pada tahapan ini data yang belum diketahui kelas atau labelnya dimasukkan ke dalam model untuk dilakukan klasifikasi. Setelah dilakukan klasifikasi maka hasilnya akan dilakukan evaluasi untuk mengetahui tingkat akurasi dari model yang dihasilkan. Apabila tingkat akurasi yang diperoleh sesuai dengan

nilai yang ditentukan, maka model tersebut dapat digunakan untuk mengklasifikasi *record-record* data baru yang belum pernah dilatih atau diuji sebelumnya (Ramadhan & Khoirunnisa, 2021).

2.1.6.1 Jenis-Jenis Klasifikasi Data

Proses klasifikasi terdiri dari dua tahap utama, yaitu pembangunan model (*training*) dan pengujian model (*testing*). Pada tahap pembangunan model, data pelatihan (*training set*) digunakan untuk melatih algoritma klasifikasi agar mengenali pola berdasarkan atribut tertentu. Setelah model terbentuk, tahap selanjutnya adalah menguji kemampuan model dalam mengklasifikasikan data baru yang belum diketahui labelnya. Evaluasi akurasi model biasanya dilakukan menggunakan metrik seperti akurasi*, precision, recall,* dan *f1-score* (Pranata et al., 2022).

Pada analisis sentimen, terdapat beberapa henis pendekatan klasifikasi yang umum digunakan, diantaranya (Arisusanto et al., 2023):

1. Klasifikasi Biner (*Binary Classification*)

   Merupakan jenis klasifikasi yang hanya memiliki dua kelas target, misalnya positif dan negatif. Ini adalah pendekatan yang paling umum dalam analisis sentimen sederhana, ulasan pengguna dikategorikan apakah mendukung (positif) atau tidak puas (negatif) terhadap suatu layanan. Misalnya, ulasan pengguna Grab yang menyatakan “pengemudinya sangat ramah” akan diklasifikasikan sebagai sentimen positif.

2. Klasifikasi Multikelas (*Multiclass Classification*)

Berbeda dengan klasifikasi biner, klasifikasi multikelas memiliki lebih dari dua kelas target. Dalam analisis sentimen, pendekatan ini digunakan untuk mengkategorikan data ke dalam tiga kelas atau lebih, seperti positif, negatif, dan netral. Hal ini memberikan fleksibilitas lebih dalam memahami opini pengguna secara lebih detail dan nyaring.

3. Klasifikasi Multilabel (*Multilabel Classification*)

   Jenis klasifikasi ini memungkinkan satu data memiliki lebih dari satu label secara bersamaan. Sentimen dalam sebuah teks dapat mengandung lebih dari satu jenis sentimen, tergantung pada konteks kalimatnya. Meskipun kurang umum dalam analisis sentimen dasar, multilabel sering digunakan dalam pengklasifikasian opini yang kompleks seperti pada ulasan produk multiaspek.

4. Klasifikasi Hierarkis (*Hierarchical Classification*)

   Merupakan pendekatan yang menggunakan kelas-kelas data memiliki struktur bertingkat atau hierarki. Jenis klasifikasi ini dapat digunakan dalam skenario di mana sentimen terbagi secara bertahap, misalnya dari "sangat negatif", "negatif", "netral", hingga "sangat positif". Model ini lebih kompleks dan memerlukan struktur data yang mendukung hierarki label.

2.1.7 Algoritma *Support Vector Machine*

Algoritma *Support Vector Machine* (SVM) mempunyai asal-usul yang menarik dalam sejarah pembangunannya. Pada tahun 1992, ketika acara tahunan Workshop Teori Pembelajaran Komputasi berlangsung, sekelompok peneliti yang

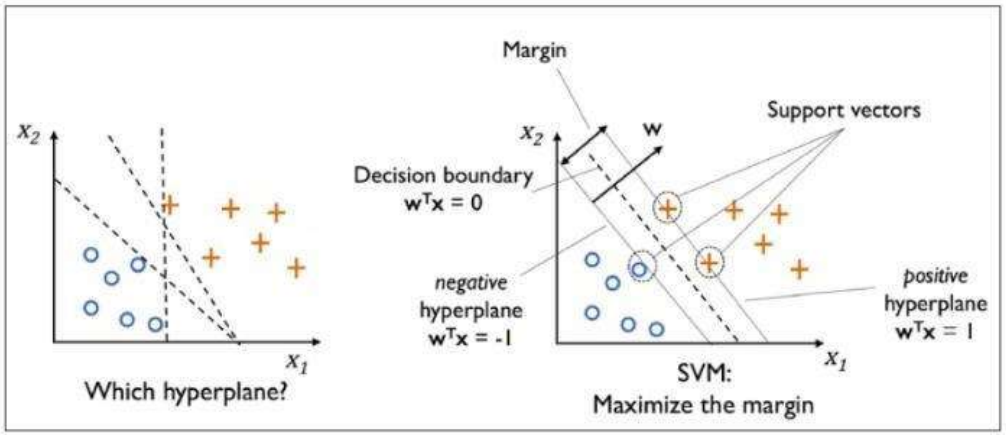
terdiri dari Boser, Guyon, dan Vapnik memperkenalkan SVM kepada dunia. SVM adalah salah satu metode pembelajaran terawasi yang cukup revolusioner dalam dunia pemrosesan data dan pembelajaran mesin (Wahyudi & Kusumawardana, 2021).

SVM dikenal karena pendekatannya yang unik dalam melakukan klasifikasi data. Ia mengoperasikan data di dalam sebuah ruang fitur yang sering kali memiliki dimensi yang tinggi. Ruang fitur ini menciptakan sebuah konteks abstrak di mana SVM mencoba untuk menemukan pemisahan linier yang optimal antara berbagai kategori atau kelas data yang diberikan.

Pendekatan ini memiliki beberapa keunggulan, salah satunya adalah kemampuannya untuk menangani data yang tidak linier dengan menerapkannya ke dalam ruang berdimensi tinggi. Dengan kata lain, SVM dapat mengatasi masalah klasifikasi yang tidak dapat dipisahkan dengan garis lurus atau polinomial sederhana.

Sejak pengenalan awalnya pada tahun 1992, SVM telah menjadi salah satu algoritma yang paling berharga dalam pembelajaran mesin, digunakan dalam berbagai aplikasi seperti klasifikasi gambar, deteksi spam, analisis biomedis, dan banyak lagi. Kesuksesannya adalah hasil dari kombinasi antara matematika yang kuat dan konsep inti yang dapat diaplikasikan pada berbagai masalah dunia nyata. Dengan terus berkembangnya teknologi dan pengetahuan dalam bidang pembelajaran mesin, SVM tetap menjadi salah satu algoritma yang relevan dan diterapkan luas.

Pada Gambar 2.1 yang disajikan di bawah, kita dapat melihat penjelasan mengenai konsep dasar dari metode klasifikasi SVM. Dalam konsep SVM ini, tujuannya adalah untuk mencari fungsi pemisah terbaik, yang juga disebut sebagai hyperline, di antara berbagai fungsi yang jumlahnya tidak terbatas. Dengan pendekatan ini, SVM memastikan bahwa kemampuan untuk menggeneralisasi pola data ke data yang akan datang menjadi sangat tinggi.



Gambar 2. 1 *Hyperline* Terbbaik dan Margin Maksimum
Sumber: (Medium)

Pada awalnya prinsip kerja dari SVM yaitu mengklasifikasi secara linear (*liniear classifier*). Kemudian SVM dikembangkan sehingga dapat bekerja pada klasifikasi *non-linear*. Formulasi optimalisasi SVM untuk masalah klasifikasi dibedakan menjadi dua kelas yaitu klasifikasi *linear* dan klasifikasi *non-linear* (Styawati et al., 2021).

2.1.8 Kepuasan Pelanggan

Kepuasan pelanggan merupakan tingkat perasaan senang atau kecewa yang timbul setelah membandingkan antara harapan pelanggan dengan kinerja nyata dari suatu produk atau layanan. Dalam konteks layanan digital seperti aplikasi Grab,

kepuasan pelanggan mencerminkan persepsi mereka terhadap pengalaman menggunakan layanan transportasi online, pemesanan makanan, pengiriman barang, dan berbagai fitur lain yang disediakan dalam satu ekosistem aplikasi (Sasongko, 2021).

Menurut teori dalam ilmu pemasaran, kepuasan pelanggan merupakan hasil evaluasi subjektif yang dapat dipengaruhi oleh berbagai faktor seperti kecepatan layanan, keramahan driver, keakuratan estimasi waktu, kemudahan navigasi aplikasi, kualitas layanan pelanggan, hingga keamanan dan kenyamanan. Ketika pengguna merasa layanan Grab memenuhi atau melampaui ekspektasi mereka, maka hal tersebut dikategorikan sebagai kepuasan pelanggan. Sebaliknya, ketidaksesuaian antara harapan dan kenyataan akan menghasilkan ketidakpuasan (Wahyudi & Kusumawardana, 2021).

Pada penelitian berbasis *data mining* dan *machine learning*, kepuasan pelanggan dapat diukur secara tidak langsung melalui analisis ulasan atau komentar pengguna. Komentar-komentar ini biasanya menggambarkan opini atau sentimen yang berkaitan dengan pengalaman mereka terhadap layanan tertentu. Dengan mengumpulkan dan menganalisis ulasan tersebut, peneliti dapat mengidentifikasi kecenderungan sentimen (positif atau negatif) yang mengindikasikan tingkat kepuasan secara umum.

Salah satu metode yang efektif untuk mengklasifikasikan sentimen pengguna berdasarkan ulasan adalah *Support Vector Machine* (SVM). SVM bekerja dengan memetakan data teks ke dalam ruang berdimensi tinggi dan membangun garis pemisah (*hyperplane*) yang dapat memisahkan kelas sentimen seperti "puas" dan

"tidak puas" secara optimal. Dengan menerapkan SVM pada data ulasan pengguna Grab, peneliti dapat mengidentifikasi pola-pola tertentu yang menunjukkan faktor-faktor dominan dalam membentuk kepuasan pelanggan (Wahyudi & Kusumawardana, 2021).

Melalui pendekatan ini, penelitian tidak hanya berfokus pada opini eksplisit, tetapi juga mampu menangkap *insight tersembunyi* dalam data teks yang bersifat tidak terstruktur. Analisis ini menjadi penting bagi perusahaan seperti Grab untuk melakukan evaluasi dan perbaikan layanan secara berkelanjutan berbasis data yang aktual.

2.1.9 Aplikasi Grab

Grab adalah sebuah aplikasi penyedia layanan pemesanan transportasi online yang berbasis di Singapura (Wahyudi & Kusumawardana, 2021). Perusahaan ini telah beroperasi sejak tahun 2012, menyediakan layanan transportasi yang diakui sebagai opsi yang nyaman, cepat, dan aman. Saat ini, Grab telah berkembang menjadi sebuah platform mobile yang menawarkan beragam layanan, termasuk transportasi, pengantaran makanan, layanan kurir, dompet digital, serta layanan keuangan lainnya. Aplikasi Grab mendukung sistem pembayaran baik dengan uang tunai maupun nontunai, seperti OVO dan kartu kredit. Aplikasi Grab dapat diunduh secara gratis melalui toko aplikasi PlayStore untuk perangkat Android dan AppStore untuk perangkat iOS.

Gambar 2. 2 Logo Grab
Sumber: (Data Penelitian, 2025)

Pada tahun 2014, Grab memasuki pasar Indonesia dan memulai operasinya dengan nama GrabTaxi. Seiring waktu, perusahaan ini memperluas layanannya dengan menyediakan berbagai jenis layanan transportasi lainnya, termasuk GrabBike, GrabCar, dan GrabExpress. Grab menjalankan kebijakan yang mengutamakan kenyamanan dan fleksibilitas bagi para mitra pengemudi serta memberikan berbagai kemudahan bagi pelanggan. Kebijakan tersebut mencakup jam kerja yang fleksibel bagi pengemudi, sistem tarif yang kompetitif, serta fitur pemesanan sebelumnya untuk meningkatkan kenyamanan dan kepuasan pelanggan.

2.1.10 Ulasan Pengguna

Ulasan pengguna (*user review*) merupakan bentuk umpan balik atau tanggapan yang diberikan oleh pengguna terhadap layanan atau produk digital yang mereka gunakan. Dalam konteks aplikasi berbasis layanan seperti Grab, ulasan pengguna biasanya berisi opini, pengalaman, kepuasan, atau keluhan yang diungkapkan melalui platform aplikasi, media sosial, atau toko aplikasi seperti Google *Play Store*. Ulasan tersebut dapat bersifat positif, negatif, atau netral,

tergantung pada pengalaman masing-masing pengguna saat menggunakan layanan seperti transportasi online, pesan-antar makanan (GrabFood), pengiriman barang (GrabExpress), dan lainnya (Wahyudi & Kusumawardana, 2021).

Ulasan pengguna memegang peranan penting dalam mengevaluasi kualitas layanan, kinerja *driver,* kemudahan penggunaan aplikasi, hingga respon sistem terhadap kendala teknis. Dalam studi analisis sentimen, ulasan ini menjadi sumber data utama yang dapat dianalisis untuk mengetahui persepsi publik secara kolektif terhadap suatu fitur atau layanan. Menurut literatur data mining, kumpulan ulasan tersebut dapat dikategorikan dan dianalisis menggunakan pendekatan *text mining* guna memperoleh pola tertentu yang bermanfaat bagi penyedia layanan untuk pengambilan keputusan yang berbasis data (Idris et al., 2023).

Lebih lanjut, ulasan yang bersifat terbuka dan publik memungkinkan pengguna lain untuk mempertimbangkan kualitas layanan sebelum memutuskan menggunakan aplikasi tersebut. Oleh karena itu, bagi perusahaan seperti Grab, ulasan pengguna menjadi indikator penting dalam mempertahankan reputasi, meningkatkan pelayanan, serta mengidentifikasi area yang perlu diperbaiki.

2.1.11 Bahasa Pemrograman Python

Python merupakan salah satu bahasa pemrograman yang memiliki ciri khas kuat dan *fleksibel* dalam dunia komputasi. Bahasa ini dikenal sebagai bahasa interpretatif serbaguna yang menekankan pada kemudahan pembacaan dan pemahaman sintaksis. Salah satu prinsip utama yang mendasari pengembangan

Python adalah penyusunan kode yang bersih dan mudah dipahami oleh manusia (Lestari et al., 2023).

Gambar 2. 3 Python
Sumber: (Data Penelitian., 2025)

Keunggulan *Python* tidak hanya terletak pada tingkat keterbacaannya yang tinggi, tetapi juga pada kemampuannya yang tangguh serta sintaks yang sederhana dan intuitif. *Python* juga didukung oleh koleksi pustaka standar yang sangat luas dan lengkap, yang memungkinkan pengembangan perangkat lunak dilakukan secara lebih efisien. Meskipun memiliki tampilan sintaks yang ringkas, *Python* tetap menawarkan kemampuan pemrograman tingkat lanjut. Ketersediaan berbagai pustaka ini turut membantu pengembang dalam membangun aplikasi modern tanpa mengorbankan kemudahan dalam membaca dan memelihara kode. Oleh karena itu,

Python menjadi salah satu alat yang andal dalam menciptakan solusi perangkat lunak yang efektif dan mudah dikembangkan (Refo et al., 2022).

2.2 *Software* Pendukung

Perangkat lunak atau *software* pendukung adalah aplikasi-aplikasi atau *tools* yang digunakan untuk dapat melakukan dan menyelesaikan penelitian terhadap tugas akhir ini. Pada proses penelitian ini akan melibatkan *tools* seperti google colabs yang digunakan untuk melakukan pengkodean program *python* mulai dari penambangan ulasan pengguna hingga implementasi SVM. Berikut ini adalah penjelasan mengenai *software* pendukung yang digunakan dalam menyelesaikan penelitian dengan judul "*Analisis Sentimen Terhadap Layanan Ojek Online Maxim Dengan Menggunakan Support Vector Machine*"

2.2.1 *Google Colaboratory*

Google Colaboratory, yang lebih dikenal sebagai Google Colab, merupakan sebuah platform yang dikembangkan oleh perusahaan teknologi global, Google, untuk mendukung aktivitas komputasi data, khususnya dalam bidang *machine learning* dan *deep learning*. Platform ini memberikan kemudahan bagi peneliti dan praktisi dalam menjalankan berbagai eksperimen berbasis kecerdasan buatan. Meski menawarkan beragam fitur unggulan, Google Colab memiliki keterbatasan dalam hal durasi dan kapasitas komputasi yang dapat diakses secara gratis (Lewis et al., 2021).

Salah satu kelebihan utama dari Google Colab adalah tersedianya akses GPU secara cuma-cuma yang dapat digunakan hingga kurang lebih 12 jam per sesi.

Fasilitas ini menjadi solusi ideal bagi pengguna yang membutuhkan sumber daya komputasi tinggi tanpa perlu berinvestasi pada perangkat keras khusus. Selain itu, *Google Colab* dibangun di atas infrastruktur Jupyter Notebook, sehingga antarmuka dan cara penggunaannya sangat familiar bagi pengguna Jupyter. Perbedaan mencolok terletak pada sistem penyimpanannya, di mana Google *Colab* terintegrasi langsung dengan Google *Drive,* serta berjalan secara penuh di lingkungan *cloud,* sehingga mendukung kolaborasi daring secara *real-time* (Guntara, 2023).

Gambar 2. 4 *Google Colaboratory*
Sumber: (Data Penelitian, 2025)

Google Colab juga menyediakan runtime *Python* versi 2 dan 3 yang telah dilengkapi dengan berbagai pustaka penting seperti *TensorFlow, Matplotlib,* dan Keras. Hal ini memungkinkan pengguna untuk segera memulai proyek-proyek *machine learning* tanpa harus melakukan proses instalasi atau konfigurasi yang kompleks.

2.3 Penelitian Terdahulu

Penelitian dengan topik *data mining* dan analisis sentimen maupun implementasi algoritma *Support Vector Machone* (SVM) telah banyak dilakukan sebelumnya, untuk memberikan gambaran dan perbandingan mengenai penelitian ini, berikut penelitian-penelitian terdahulu.

1. Penelitian yang telah dilakukan oleh (Prastyo et al., 2021) dengan judul "*A Combination Expansion Rangking and GA SVM for Improving Indonesian Sentiment Classification Performance*" dan terindeks SCOPUS. Penelitian ini menerapkan pendekatan gabungan antara *Quality Evaluation Research* (QER) dan *Genetic Algorithm–Support Vector Machine* (GA–SVM) untuk melakukan seleksi fitur dalam analisis sentimen. Data diperoleh dalam rentang waktu 8 Juli hingga 29 Juli 2020 melalui pemanfaatan API GetOldTweets3, yang menghasilkan total 14.097 data. Dari keseluruhan data tersebut, sebanyak 4.000 catatan (terdiri dari 2.000 sentimen positif dan 2.000 sentimen negatif) digunakan sebagai data latih setelah melalui tahapan *preprocessing*, yang mencakup penghapusan simbol khusus Twitter, *case folding*, normalisasi bahasa tidak baku, *stemming*, dan *tokenization*. Proses ekstraksi fitur dilakukan menggunakan pendekatan TF-IDF. Berdasarkan hasil uji-t statistik, ditemukan perbedaan signifikan antara metode yang diusulkan dan algoritma konvensional, ditunjukkan oleh nilai *p-value* yang kurang dari 0,05. Pendekatan ini menunjukkan performa klasifikasi yang sangat baik, dengan nilai *precision* sebesar 96,78%, *recall* 96,76%, *f1-score*

96,75%, dan AUC sebesar 98,66%, serta mampu mengurangi waktu komputasi secara signifikan.

2. Penelitian yang dilakukan oleh (Lestari et al., 2023) berjudul *"Penerapan Algoritma Support Vector Machine pada Analisis Sentimen Terhadap Identitas Kependudukan Digital"* dan telah terindeks pada SINTA 2. Studi ini menyoroti dinamika opini publik, baik yang mendukung maupun menolak, terkait implementasi Identitas Kependudukan Digital (E-KTP). Data diperoleh melalui teknik *crawling* dari komentar pengguna di platform Facebook. Analisis sentimen dilakukan menggunakan algoritma *Support Vector Machine* (SVM) dan menghasilkan tingkat akurasi sebesar 77%. Dari total 902 data yang dianalisis, sebanyak 78,27% mengandung sentimen negatif, 12,97% netral, dan hanya 8,76% yang bersentimen positif. Temuan ini mencerminkan bahwa sebagian besar masyarakat menunjukkan ketidakpuasan terhadap kehadiran sistem identitas kependudukan digital.
3. Penelitian oleh (Idris et al., 2023) yang berjudul *"Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma SVM"* dan terindeks SINTA 3, berfokus pada pengklasifikasian sentimen pengguna terhadap aplikasi Shopee. Studi ini memanfaatkan sebanyak 3.000 ulasan pengguna yang dikumpulkan melalui metode *scraping* dari platform Playstore. Ulasan dibedakan berdasarkan skor bintang, di mana bintang 1 hingga 3 dikategorikan sebagai sentimen negatif, dan bintang 4 hingga 5 sebagai sentimen positif. Melalui pemerapan algoritma *Support Vector Machine (SVM*) menggunakan berbagai parameter, penelitian ini berhasil mencapai

performa optimal dengan nilai akurasi dan *f1-score* masing-masing sebesar 98%. Hasil tersebut menunjukkan efektivitas SVM dalam mengklasifikasikan opini pengguna aplikasi *e-commerce* berbasis ulasan daring.

4. Penelitian yang dilakukan oleh (Pranata et al., 2022) berjudul *"Klasifikasi Sentimen Terhadap Maxim Menggunakan Algoritma SVM pada Media Sosial Twitter"* dan terindeks SINTA 5, dengan tujuan untuk mengidentifikasi opini publik terhadap layanan ojek online Maxim melalui analisis sentimen berbasis data dari Twitter. Studi ini mengumpulkan sebanyak 1.200 data tweet menggunakan Twitter API yang kemudian diberi label secara manual oleh ahli bahasa Indonesia menjadi tiga kategori sentimen: positif, netral, dan negatif. Proses pengolahan data dilakukan melalui beberapa tahapan *text preprocessing*, meliputi case folding, cleaning, tokenizing, filtering, dan stemming dengan algoritma Nazief & Adriani. Selanjutnya dilakukan pembobotan menggunakan metode TF-IDF dan klasifikasi sentimen menggunakan algoritma Support Vector Machine (SVM). Penelitian ini menguji beberapa kernel SVM seperti RBF, Polynomial, dan Sigmoid dengan variasi rasio data latih dan uji (90:10, 80:20, dan 70:30). Hasil penelitian menunjukkan bahwa kernel RBF dan Polynomial memberikan performa terbaik dengan akurasi 85% pada skenario pembagian data latih dan uji sebesar 90:10. Selain itu, analisis menunjukkan bahwa sebagian besar tweet pengguna berisi opini positif terhadap layanan Maxim. Temuan ini menegaskan efektivitas algoritma SVM dalam mengklasifikasikan sentimen media sosial secara akurat serta menunjukkan bahwa pendekatan ini dapat

dimanfaatkan oleh penyedia layanan untuk mengevaluasi persepsi publik terhadap layanannya.

5. Penelitian oleh (Permata Aulia et al., 2021) berjudul *"Perbandingan Kernel Support Vector Machine (SVM) dalam Penerapan Analisis Sentimen Vaksinasi COVID-19"* dan terindeks SINTA 3 bertujuan mengevaluasi performa berbagai kernel pada algoritma SVM dalam mengklasifikasikan sentimen masyarakat terhadap vaksin COVID-19. Sebanyak 1.977 tweet berbahasa Indonesia dikumpulkan dari Twitter menggunakan kata kunci "vaksin COVID-19", lalu diproses melalui tahapan *text preprocessing* dan pembobotan TF-IDF. Model SVM dikembangkan menggunakan empat jenis kernel: linear, sigmoid, polynomial, dan RBF, dengan tiga kategori sentimen: positif, negatif, dan netral. Hasil evaluasi menunjukkan bahwa kernel linear dan sigmoid memberikan performa terbaik dengan akurasi 87%, precision 90%, recall 77%, dan f-measure 82%. Penelitian ini menegaskan bahwa pemilihan kernel yang tepat pada SVM berperan penting dalam meningkatkan akurasi klasifikasi teks, terutama dalam konteks opini publik terhadap isu sosial.
6. Penelitian oleh (Arsi & Waluyo, 2021) berjudul "*Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine*" yang terindeks SINTA 2. Penelitian ini bertujuan untuk menganalisis sentimen publik terkait isu pemindahan ibu kota Indonesia melalui media sosial Twitter menggunakan algoritma *Support Vector Machine* (SVM). Sebanyak 1.236 tweet dikumpulkan menggunakan metode

crawling dengan kata kunci tertentu, lalu diklasifikasikan menjadi dua kelas sentimen: positif dan negatif. Data dianalisis melalui tahapan preprocessing seperti case folding, cleaning, tokenizing, stopword removal, stemming, dan pembobotan TF-IDF. Model SVM dikembangkan dan divalidasi menggunakan metode 5-fold cross validation. Hasil evaluasi menunjukkan kinerja sangat baik dengan nilai akurasi 96,68%, precision 95,82%, recall 94,04%, dan AUC 0,979. Penelitian ini menegaskan efektivitas SVM dalam menangani data opini yang kompleks di media sosial serta memberikan wawasan terhadap persepsi masyarakat dalam isu kebijakan publik.

7. Penelitian yang dilakukan oleh (Atmajaya et al., 2023) berjudul *"Metode SVM dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter"*dan terindeks SINTA 5. Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma klasifikasi *Support Vector Machine* (SVM) *dan Naive Bayes* dalam menganalisis sentimen pengguna Twitter terhadap ChatGPT. Dataset yang digunakan terdiri dari 1.000 tweet yang diperoleh dari Kaggle dengan kata kunci "ChatGPT", dan dianalisis menggunakan dua model pelabelan sentimen populer, yaitu Vader dan RoBERTa. Setelah melalui proses *preprocessing* dan ekstraksi fitur menggunakan TF-IDF, model dievaluasi berdasarkan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa kombinasi SVM dengan Vader memberikan performa terbaik dengan akurasi 59%, presisi 59%, recall 59%, dan F1-score 55%. Kombinasi SVM dengan RoBERTa menghasilkan akurasi 55%, sedangkan *Naive Bayes* menunjukkan performa yang lebih rendah, khususnya dengan

RoBERTa yang hanya mencapai akurasi 43% dan F1-score 26%. Penelitian ini menyimpulkan bahwa metode SVM lebih unggul dibanding *Naive Bayes* dalam klasifikasi sentimen, dan pemilihan metode pelabelan sentimen (Vader atau RoBERTa) sangat mempengaruhi hasil klasifikasi.

8. Penelitian oleh (Putra et al., 2023) dengan judul "*Perbandingan Feature Extraction TF-IDF dan BOW untuk Analisis Sentimen Berbasis SVM*" dab terindeks SINTA 5. Penelitian ini membandingkan dua teknik ekstraksi fitur, yaitu TF-IDF dan Bag of Words (BOW), dalam analisis sentimen menggunakan algoritma *Support Vector Machine* (SVM). Data penelitian diperoleh dari Twitter dengan kata kunci "#jne", mencakup 300 tweet yang telah melalui proses *crawling*, pelabelan manual, *preprocessing*, dan validasi. Sentimen dibagi menjadi dua kategori: positif dan negatif. Hasil evaluasi menunjukkan bahwa kombinasi TF-IDF + SVM menghasilkan performa lebih baik dengan akurasi 86%, precision 85%, recall 85%, dan f1-score 85%. Sementara itu, BOW + SVM mencatat akurasi 85%, precision 80%, recall 89%, dan f1-score 84%. Dengan demikian, TF-IDF unggul dalam keseluruhan metrik evaluasi, kecuali pada nilai recall yang lebih tinggi pada metode BOW. Penelitian ini menegaskan pentingnya pemilihan teknik ekstraksi fitur yang tepat untuk meningkatkan akurasi klasifikasi sentimen pada data teks.
9. Penelitian yang dilakukan oleh (Muhammadin & Sobari, 2021) berjudul *"Analisis Sentimen pada Ulasan Aplikasi Kredivo dengan Algoritma SVM dan NBC"*dan terindeks SINTA 5. Penelitian ini bertujuan untuk

membandingkan performa dua algoritma klasifikasi, yaitu Support Vector Machine (SVM) dan Naive Bayes Classifier (NBC), dalam mengklasifikasikan sentimen ulasan pengguna terhadap aplikasi Kredivo. Sebanyak 10.000 data ulasan berbahasa Indonesia dikumpulkan melalui *web scraping* dari Google Play Store, kemudian diproses melalui tahap *preprocessing* seperti case folding, tokenizing, stopword removal, normalization, dan stemming. Fitur teks diekstraksi menggunakan metode TF-IDF, dan data dibagi dengan proporsi 80% data latih dan 20% data uji. Hasil pengujian menunjukkan bahwa SVM menghasilkan akurasi sebesar 83,3%, dengan presisi 81% dan recall 88% untuk sentimen positif, serta presisi 87% dan recall 79% untuk sentimen negatif. Sementara itu, NBC menghasilkan akurasi 80,8%, dengan presisi dan recall yang sedikit lebih rendah. Berdasarkan evaluasi menggunakan confusion matrix dan metrik f1-score, algoritma SVM terbukti memberikan performa klasifikasi yang lebih baik dibandingkan Naive Bayes. Penelitian ini menegaskan efektivitas algoritma SVM dalam menganalisis sentimen ulasan pengguna aplikasi berbasis teks secara lebih akurat.

10. Penelitian yang dilakukan oleh (Pratiwi et al., 2021) berjudul *"Analisis Sentimen pada Review Skincare Female Daily Menggunakan Metode Support Vector Machine (SVM)"* dan terindeks SINTA 4. Penelitian ini bertujuan untuk mengklasifikasikan sentimen ulasan produk kecantikan lokal di platform Female Daily. Sebanyak 1.249 ulasan dikumpulkan melalui teknik *scraping*, dengan 649 ulasan berkategori "Ya" (positif) dan 600 "Tidak"

(negatif). Proses *preprocessing* mencakup tahapan *case folding*, *filtering*, dan *tokenizing*, serta pembobotan fitur menggunakan metode TF-IDF. Data kemudian dibagi dengan proporsi 80% data latih dan 20% data uji, dan diklasifikasikan menggunakan algoritma Support Vector Machine. Hasil evaluasi menunjukkan performa model yang cukup baik, dengan akurasi sebesar 87%, recall 90%, precision 84,90%, dan f1-score 87,37%. Penelitian ini menunjukkan bahwa algoritma SVM efektif dalam menangani data teks ulasan pengguna, khususnya dalam konteks produk skincare yang bersifat subjektif dan emosional.

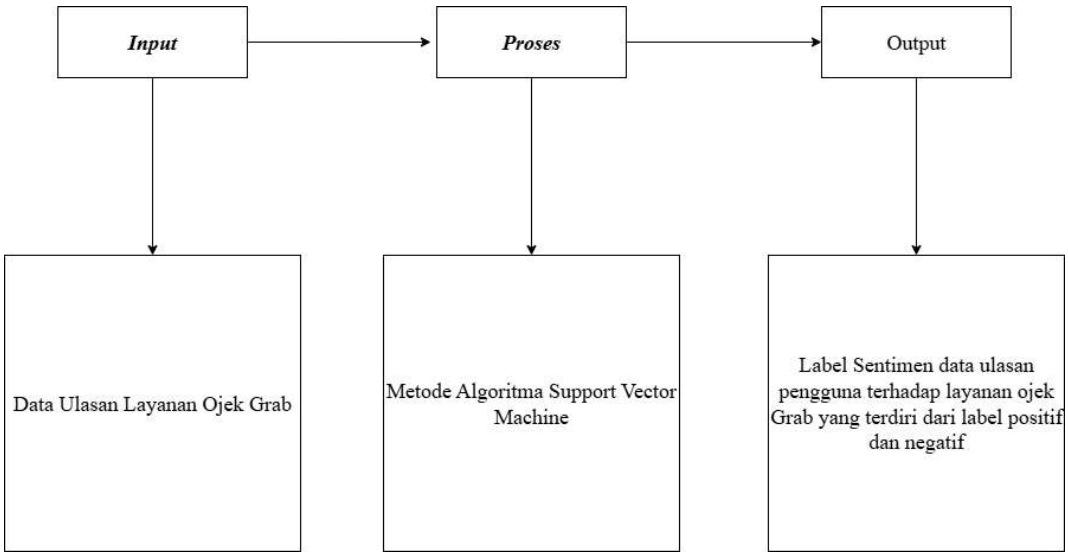
2.4 Kerangka Pemikiran

Kerangka pemikiran dalam penelitian ini berfokus pada analisis sentimen kepuasan pelanggan terhadap layanan transportasi online Grab, yang dianalisis berdasarkan ulasan atau komentar pengguna. Data ulasan tersebut berisi tanggapan, pengalaman, maupun persepsi pengguna terhadap berbagai aspek layanan Grab seperti kenyamanan, kecepatan, keramahan driver, dan keandalan sistem. Input utama dalam penelitian ini adalah data teks ulasan pengguna yang dikumpulkan dari platform digital seperti Play Store atau media sosial.

Untuk mengolah data tersebut, penelitian ini menerapkan metode *Support Vector Machine* (SVM), sebuah algoritma *machine learning* yang terbukti efektif dalam klasifikasi teks. Proses analisis dilakukan melalui beberapa tahapan, mulai dari *preprocessing* data (seperti *case folding*, tokenisasi, *stopword removal*, dan *stemming*), pembobotan fitur menggunakan TF-IDF, hingga tahap klasifikasi sentimen menjadi dua kategori utama, yaitu positif dan negatif.

Penerapan SVM, sistem akan mempelajari pola dalam data pelatihan dan membentuk model klasifikasi yang dapat digunakan untuk memprediksi sentimen dari data baru. Proses ini bertujuan untuk mengevaluasi tingkat kepuasan pelanggan secara otomatis berdasarkan opini yang disampaikan dalam ulasan. Hasil analisis diharapkan dapat memberikan gambaran objektif mengenai persepsi pengguna terhadap layanan Grab, serta menjadi bahan pertimbangan dalam meningkatkan kualitas dan kepuasan layanan.

Kerangka pemikiran ini menjadi dasar metodologis dalam melakukan analisis sentimen berbasis data teks, serta memberikan landasan penting dalam pemanfaatan teknologi pembelajaran mesin untuk memahami kebutuhan dan harapan pelanggan. Adapun diagram kerangka pemikiran penelitian ini disajikan pada gambar berikut.



Gambar 2. 25 Kerangka Pemikiran
Sumber: (Data Penelitian, 2025)

Berdasarkan gambar di atas, penelitian dimulai dengan tahap input, yaitu pengumpulan data ulasan pengguna yang diperoleh dari Google *Play Store*. Data ini berupa teks yang mengandung opini atau tanggapan pengguna terhadap layanan,

baik yang bersifat positif maupun negatif. Tahapan selanjutnya merupakan proses klasifikasi sentimen yang dilakukan dengan menerapkan metode algoritma SVM. Proses ini mencakup berbagai tahap pra-pemrosesan data, seperti tokenisasi, penghapusan *stopword*, stemming, serta transformasi data ke dalam bentuk vektor numerik menggunakan metode representasi teks seperti TF-IDF. Setelah itu, data yang telah diproses akan digunakan dalam pelatihan model untuk mengklasifikasikan sentimen menjadi dua kategori, yakni positif dan negatif. Pada tahap akhir, sistem menghasilkan output berupa label sentimen dari setiap data ulasan yang dianalisis. Hasil klasifikasi ini diharapkan dapat memberikan gambaran objektif mengenai persepsi dan kepuasan pelanggan terhadap layanan Grab, serta menjadi dasar pertimbangan bagi pengambilan keputusan strategis oleh penyedia layanan.