

## **BAB II**

### **KAJIAN PUSTAKA**

#### **2.1 Teori Dasar**

Teori dasar dalam penelitian ini akan membahas teori-teori yang berhubungan dengan topik penelitian yang diusulkan, pembahasan teori ini menjadi penting sebagai instrumen untuk meningkatkan validitas dari penelitian yang dilaksanakan. Dalam pembahasannya teori-teori yang dibahas terdiri dari pengertian dan istilah umum teori yang digunakan, adapun teori yang akan dibahas dalam penelitian ini terdiri dari: (1) *Knowledge Discovery in Database* (KDD), (2) *Data Mining*, (3) Kelebihan dan kekurangan SVM dan Maxim, (4) Layanan ojek *online* Maxim di Batam, (5) *Sentiment Analysis*, dan (6) *RapidMiner*.

#### **2.2 *Knowledge Discovery in Database***

Penemuan Pengetahuan dalam Database atau *Knowledge Discovery in Database* (KDD) pertama kali diungkapkan pada tahun 1989. Istilah ini digunakan untuk menekankan bahwa pengetahuan dapat ditemukan melalui penggalian informasi dari database (Alghifari & Juardi, 2021).

*Data mining* seringkali disebut sebagai KDD padahal KDD adalah proses ekstraksi informasi dari data secara keseluruhan, sementara *data mining* merupakan salah satu tahap pemrosesan KDD yang fokus pada identifikasi pola pada data (Pramadani et al., 2018). Proses yang terjadi dalam KDD bersifat berulang, yaitu pengukuran dapat dilakukan berkali-kali dan ditingkatkan, selain itu data baru dapat

diintegrasikan di tengah proses berjalan dan dapat ditransformasi ulang untuk mencari hasil yang berbeda atau lebih baik.

Konsep *Data Mining* saat ini diakui sebagai teknik yang sangat penting dalam bidang manajemen informasi, terutama karena meningkatnya volume informasi yang tersedia. *Data Mining*, yang umumnya dikenal sebagai Penemuan Pengetahuan dalam Database (KDD), adalah ekstraksi sistematis pola, keteraturan, dan korelasi dari basis data yang luas melalui penggunaan data historis (Handoko, 2016).

Tahapan-tahapan yang terdapat dalam KDD terdiri dari 5 tahapan utama, berdasarkan (Alghifari & Juardi, 2021) tahapan-tahapan yang terdapat dalam KDD adalah sebagai berikut:

1. *Data Selection*

Tahapan pemilihan data (*data selection*) dilakukan dengan tujuan memilih subset data yang relevan dan representatif untuk dilakukan analisis lebih lanjut, hal ini diperlukan untuk mengurangi *noise* yang dikandung dalam data. Pemilihan data yang baik akan menghasilkan proses pengolahan dan luaran yang lebih baik.

2. *Pre-processing*

Tahapan pra-proses dalam KDD merupakan langkah-langkah yang dilakukan sebelum melanjutkan ke tahap penambangan data (*data mining*). Tahapan ini sangat penting untuk mempersiapkan data yang akan digunakan dalam analisis. Tahapan pra-proses biasanya terdiri dari beberapa tahapan penting seperti pembersihan data, dan melakukan perbaikan dalam data.

### 3. *Transformation*

Pada tahap ini akan dilakukan proses pencarian fitur-fitur yang bermanfaat untuk mewakili data tergantung pada tujuan yang ingin dicapai, langkah transformasi data yang penting setelah pemilihan data yang relevan. Transformasi ini mengubah data yang telah dipilih agar sesuai dengan persyaratan yang diperlukan dalam proses penambangan data. Proses ini memiliki unsur kreatif dan sangat bergantung pada jenis informasi atau pola yang akan diidentifikasi dalam basis data.

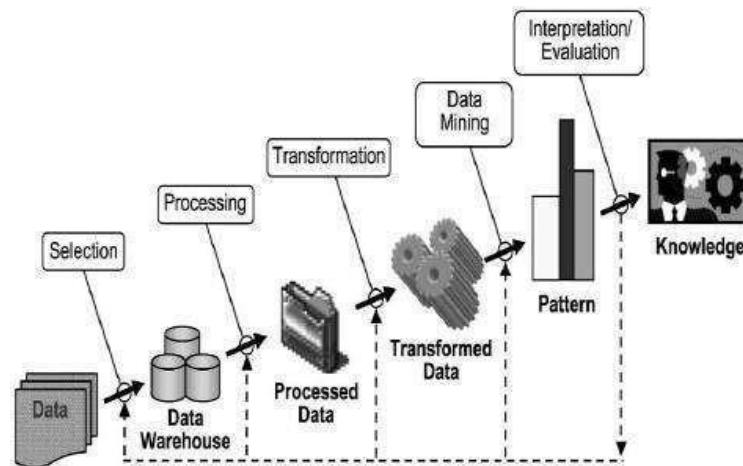
### 4. *Data Mining*

Pada tahapan ini dilakukan berbagai pemilihan mulai dari metode *data mining* yang akan digunakan, seperti klasifikasi, regresi, *clustering*, dll. Serta melakukan pemilihan algoritma untuk melakukan pencarian. Proses *data mining* adalah upaya untuk mencari pola tertentu untuk informasi menarik dalam data terpilih dengan menggunakan berbagai teknik, metode atau algoritma yang sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

### 5. *Interpretation/Evaluation*

Tahapan ini menampilkan hasil penjumlahan pola (*pattern*) yang dihasilkan dari proses *data mining*, tahapan ini juga akan melakukan evaluasi dari pola atau informasi yang didapat sesuai atau bertentangan dengan fakta/hipotesa yang ada sebelumnya.

Adapun proses *Knowledge Discovery in Database* hasil penjabaran di atas dapat dilihat pada gambar di bawah ini:



**Gambar 2. 1** Tahapan Proses *Knowledge Discovery in Database*  
**Sumber:** (Walhaidayat et al, 2021)

### 2.3 *Data Mining*

*Data mining* adalah istilah yang merujuk pada proses pengungkapan pengetahuan yang terdapat dalam basis data. *Data mining* melibatkan penggunaan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstraksi dan mengidentifikasi informasi yang berharga serta pengetahuan yang terkait dari berbagai basis data besar (Yuli Mardi, 2019), sementara dalam (Handoko, 2016) dijelaskan bahwa *Data mining* adalah prosedur sistematis yang menggunakan metodologi statistik, prinsip matematika, pendekatan kecerdasan buatan, dan algoritma *machine learning* untuk mengekstraksi dan memahami informasi penting serta pengetahuan yang relevan dari basis data yang luas. *Data Mining* memiliki akar sejarah dalam berbagai disiplin ilmu akademik, termasuk kecerdasan buatan, *machine learning*, statistik, dan basis data. *Literature data*

*mining* seringkali merujuk kepada beberapa pendekatan, seperti pengelompokan, klasifikasi, penambangan aturan asosiasi, jaringan saraf tiruan, dan algoritma evolusi.

*Data mining* bukan sekadar menggali data, melainkan sebuah pendekatan yang sangat sistematis dan terstruktur. Selama proses *data mining*, data historis yang tersedia dianalisis secara menyeluruh, dilakukan transformasi data jika diperlukan, dan diterapkan berbagai teknik untuk mengungkapkan pola, keteraturan, atau hubungan yang mungkin tersembunyi dalam dataset tersebut.

Menurut Rerung (2018), *Data mining* memiliki beberapa kategori berdasarkan tugas atau pekerjaan yang dapat dilakukan. yaitu:

1. Deskripsi: kategori *data mining* deskriptif adalah salah satu jenis metode *data mining* yang bertujuan untuk mencari pola yang dapat dimengerti oleh manusia dan menjelaskan karakteristik dari data.
2. Estimasi: estimasi dalam *data mining* bertujuan untuk memperkirakan nilai dari suatu variabel berdasarkan data yang ada.
3. Prediksi: prediksi dalam *data mining* bertujuan untuk memprediksi nilai yang belum diketahui berdasarkan data yang tersedia.
4. Klasifikasi: klasifikasi dalam *data mining* bertujuan untuk memprediksi kelas atau kategori data dengan memanfaatkan nilai yang ada pada data. Hasil kualifikasi adalah kategori atau kelas data yang bersifat nominal atau ordinal.
5. Pengklusteran: pengklusteran dalam *data mining* bertujuan untuk mengelompokkan data dengan karakteristik yang sama ke dalam satu

kelompok yang sama, dan data dengan karakteristik yang berbeda ke kelompok lain.

6. Asosiasi: asosiasi dalam *data mining* bertujuan untuk menemukan aturan asosiasi antar suatu kombinasi item. Metode asosiasi sering digunakan dalam analisis keranjang belanja atau *market basket analysis*.

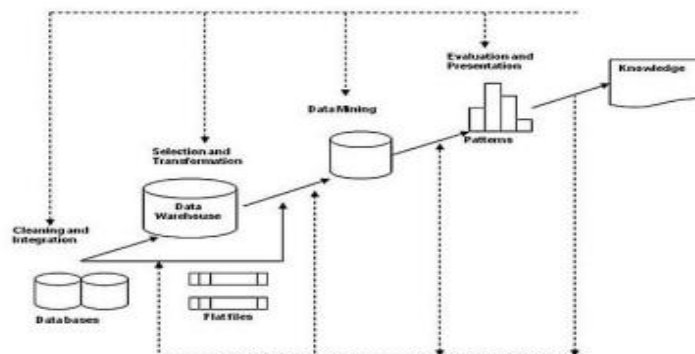
Tahapan-tahapan dalam *data mining* menurut Rerung (2018) dibagi menjadi enam bagian yaitu:

1. Pembersihan data (*data cleaning*), sebelum melakukan proses *data mining*, diperlukan tahapan *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* meliputi penghapusan duplikasi data, pemeriksaan data yang tidak konsisten, dan perbaikan kesalahan pada data seperti kesalahan cetak.
2. Integrasi data (*data integration*), integrasi data adalah proses menggabungkan data dari berbagai *database* ke dalam satu basis data baru. Seringkali, data yang diperlukan untuk keperluan penambangan data tidak hanya berasal dari satu sumber *database*, tetapi dapat datang dari beberapa basis data atau berkas teks yang berbeda. Integrasi data ini berfokus pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik, seperti nama, jenis produk, nomor pelanggan, dan lainnya. Pentingnya integrasi data sangat terlihat dalam ketelitian eksekusinya, karena kesalahan dalam proses integrasi data dapat menghasilkan hasil yang tidak sesuai dan bahkan menyesatkan saat pengambilan keputusan di masa mendatang. Sebagai contoh, jika integrasi data didasarkan pada jenis produk dan menggabungkan produk dari kategori

yang sebenarnya berbeda, maka akan timbul korelasi antar produk yang sebenarnya tidak ada.

3. Seleksi data (*data selection*), tidak semua data yang terdapat dalam sebuah basis data selalu digunakan; oleh karena itu, hanya data yang relevan untuk analisis yang akan diekstraksi dari basis data. Sebagai ilustrasi, dalam sebuah studi yang mengevaluasi faktor kecenderungan pembelian dalam analisis keranjang belanja, tidak diperlukan untuk mengambil nama pelanggan, cukup dengan mengambil identifikasi pelanggan (ID pelanggan) saja.
4. Transformasi data (*data transformation*), data diproses atau disesuaikan ke dalam bentuk yang cocok untuk digunakan dalam *data mining*. Beberapa teknik *data mining* memerlukan format data tertentu sebelum dapat diaplikasikan. Sebagai contoh, beberapa metode yang umum seperti analisis asosiasi dan pengelompokan hanya dapat menerima data yang bersifat kategorikal. Oleh karena itu, data berupa angka *numerik* yang kontinu harus diubah menjadi interval-interval tertentu. Proses ini sering disebut sebagai transformasi data.
5. Proses *mining*, merupakan tahap yang sangat krusial ketika suatu metode digunakan untuk menggali informasi berharga dan tersembunyi dari data.
6. Evaluasi pola (*pattern evaluation*), untuk mengidentifikasi pola-pola menarik yang ditemukan dalam pengetahuan yang didasarkan pada data. Dalam tahap ini, hasil-hasil dari teknik data mining berupa pola-pola yang unik serta model prediksi dievaluasi untuk menentukan apakah hipotesis yang ada telah terverifikasi.

7. Presentasi pengetahuan (*knowledge presentation*), tahapan ini melibatkan penyajian dan visualisasi informasi mengenai metode yang digunakan untuk memperoleh pengetahuan yang akan diberikan kepada pengguna. Tahap terakhir dalam proses data mining adalah bagaimana merumuskan keputusan atau tindakan berdasarkan hasil analisis yang telah diperoleh. Terkadang, ini melibatkan pihak yang tidak familiar dengan konsep *data mining*, oleh itu, penting untuk menyajikan hasil *data mining* dalam bentuk pengetahuan yang bisa dipahami oleh semua pihak terkait. Oleh karena itu, dalam tahap presentasi ini, visualisasi juga dapat digunakan untuk membantu menyampaikan hasil *data mining* dengan lebih jelas.



**Gambar 2. 2** Tahapan *Data Mining*  
**Sumber:** (Data Penelitian, 2024)

## 2.4 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 (Nomleni, 2015).

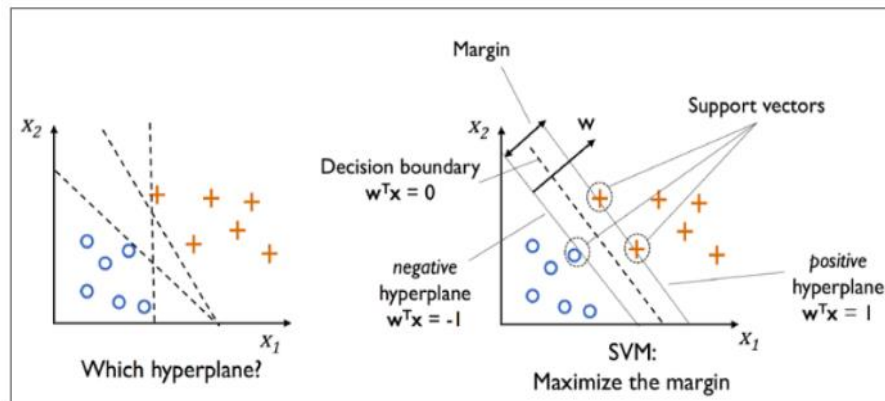
SVM dikenal karena pendekatannya yang unik dalam melakukan klasifikasi data. Ia mengoperasikan data di dalam sebuah ruang fitur yang sering kali memiliki dimensi yang tinggi. Ruang fitur ini menciptakan sebuah konteks abstrak di mana SVM mencoba untuk menemukan pemisahan linier yang optimal antara berbagai 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 2.2 yang disajikan di bawah, dapat dilihat 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. 3** *Hyperplane* Terbaik dan Margin Maksimum  
Sumber: (Medium)

Pada awalnya prinsip kerja dari SVM yaitu mengklasifikasi secara *linear* (*linear 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*.

*Hyperplane* pemisah terbaik antara dua kelas dapat ditemukan dengan mengukur margin *hyperplane* dan menemukan titik maksimumnya. Margin adalah jarak antara *hyperplane* dan data terdekat dari setiap kelasnya.

Data yang tersedia dilambangkan dengan  $x \in R^d$ , sedangkan label terkait dilambangkan dengan  $y_i \in \{-1, +1\}$  untuk  $i = 1, 2, \dots, n$ , dimana  $n$  merupakan jumlah data. Diasumsikan bahwa kedua kelas  $-1$  dan  $+1$  dapat dipisahkan sepenuhnya oleh *hyperplane* berdimensi  $d$ , yang didefinisikan:  $w \cdot x + b = 0$ .

Pola  $x_i$  milik kelas  $-1$  (sampel negatif) dapat dirumuskan sebagai pola memenuhi pertidaksamaan:  $w \cdot x + b = -1$

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara *hyperplane* dan titik terdekatnya, yaitu  $1 / \|d\|$ . Hal ini dapat dirumuskan sebagai masalah pemrograman *quadratic* (QP) (Santoso, 2021), yaitu mencari titik minimum persamaan dengan batas persamaan:

$$\tau(w) = \frac{1}{2} \|w\|^2$$

$$y_i(x_i \cdot w + b) - 1 \geq 0$$

Permasalahan ini dapat dipecahkan dengan menggunakan teknik komputasi *Lagrange Multiplier*.

1. Mencari *Lagrange Multipliers* ( $\alpha_i$ )

$$L(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

**Rumus 2.1** Lagrange Multiplier

Keterangan:

$y_i$  : kelas data latih (+1/-1)

$y_j$  : kelas data latih (+1/-1)

$x_i$ : vektor bobot kalimat komentar

$x_j$ : vektor bobot kalimat komentar

2. Mencari nilai bobot ( $w$ )

$$W = \sum_{i=1}^n (\alpha_i y_i x_i)$$

**Rumus 2.2** Menghitung Nilai Bobot

Keterangan:

$w$ : vektor bobot

$y_i$ : kelas data latih (+1/-1)

$x_i$ : vektor bobot kalimat komentar yang menjadi vektor pendukung

## 3. Mencari Nilai Bias (b)

$$b = \frac{1}{NSV} \sum_{i=1}^{NSV} (W \cdot x_i - y_i)$$

**Rumus 2.3** Menghitung Nilai Bias

Keterangan:

*NSV*: jumlah vektor pendukung*w*: vektor bobot*y<sub>i</sub>*: kelas data latih (+1/-1)*x<sub>i</sub>*: vektor bobot kalimat komentar yang menjadi vektor pendukungProses pengklasifikasian atau pengujian dalam algoritma *Support Vector**Machine* menggunakan persamaan sebagai berikut:

$$f(\vec{t}) = \text{sgn}\left(\sum_{i=1, x_i \in SV}^n a_i y_i < \vec{t} \cdot \vec{x}_i + b\right)$$

**Rumus 2.4** Klasifikasi Kelas dengan SVM

Keterangan:

*t*: vektor bobot data uji*x<sub>i</sub>*: vektor pendukung*b*: nilai bias*y<sub>i</sub>*: kelas atau label dari vektor pendukung (+1/-1)

$\alpha_i$  adalah *multiplikator Lagrange*, yang memiliki nilai nol atau positif ( $\alpha_i \geq 0$ ). Nilai optimal dari persamaan dapat dihitung dengan mengurangi *L* terhadap *w* dan *b*, dan menambahkan *L* terhadap  $\alpha_i$ . Dengan mempertimbangkan sifat bahwa pada titik optimal gradien  $L=0$ , langkah-langkah persamaan dapat diubah menjadi masalah maksimalisasi yang hanya melibatkan  $\alpha_i$ , seperti yang terlihat dalam persamaan *Maximize*:

$$\sum_{i=1} a_i - \frac{1}{2} \sum_{i,j=1} a_i a_j y_i \cdot y_j, \vec{x}_i, \vec{x}_j$$

**Rumus 2.5** Menghitung Nilai Maksimal Margin

## 2.5. Kelebihan dan Kekurangan Algoritma *Support Vector Machine*

Setiap algoritma atau metode pasti memiliki kelebihan dan kekurangannya dalam beberapa hal, termasuk untuk algoritma *Support Vector Machine* (Ichwan et al., 2019), di bawah ini adalah kelebihan dan kekurangan algoritma SVM.

Kelebihan:

1. SVM bekerja relatif lebih baik ketika terdapat margin pemisahan yang jelas antar kelas.
2. SVM lebih efektif pada ruang berdimensi tinggi.
3. SVM efektif jika jumlah dimensi lebih besar daripada jumlah sampel yang tersedia.
4. SVM relatif lebih hemat memori.
5. SVM mampu menghasilkan model klasifikasi yang baik meskipun dilatih dengan himpunan data yang relatif sedikit.

Kekurangan:

1. Algoritma SVM tidak cocok untuk kumpulan data yang besar
2. SVM tidak bekerja dengan baik ketika kumpulan data memiliki lebih banyak gangguan, yaitu kelas target yang tumpang tindih.
3. Jika jumlah fitur untuk setiap titik data melebihi jumlah sampel data pelatihan, SVM akan berkinerja buruk.

4. Karena pengklasifikasi vektor pendukung bekerja dengan meletakkan titik data, di atas dan di bawah *hyperplane* pengklasifikasian, tidak ada penjelasan probabilistik untuk klasifikasi tersebut.

## 2.6 Layanan Ojek *Online* Maxim di Kota Batam

Layanan ojek *online* Maxim adalah perusahaan asal Rusia yang bergerak di bidang inovasi teknologi, didirikan pada tahun 2003 di *Chardinsk*, Pegunungan Ural (Akbar et al., 2022). Perusahaan ini telah beroperasi sejak tahun 2003, menyediakan layanan transportasi yang diakui sebagai opsi yang nyaman dan aman. Saat ini, Maxim telah berkembang menjadi sebuah platform *mobile* yang menawarkan beragam layanan, termasuk transportasi, pengiriman makanan, layanan lainnya, dan masih banyak lagi. Aplikasi Maxim mendukung sistem pembayaran baik dengan uang tunai maupun nontunai, seperti OVO. Aplikasi Maxim dapat diunduh secara gratis melalui toko aplikasi *PlayStore* untuk perangkat *Android* dan *AppStore* untuk perangkat IOS.

Pada tahun 2018, Maxim memasuki pasar Indonesia dan memulai operasinya dengan bekerja sama dengan PT. Teknologi Prima Indonesia. Perusahaan ini tidak hanya menawarkan layanan transportasi *online* yang terutama berfokus pada taksi, tetapi juga berbagai jenis layanan transportasi lainnya, termasuk penyediaan kendaraan seperti taksi konvensional. Maxim menjalankan kebijakan yang menguntungkan baik bagi para pengemudi maupun pelanggan. Ini mencakup jam kerja yang dapat disesuaikan oleh pengemudi, tarif yang terjangkau, serta sistem pemesanan sebelumnya untuk kepuasan pelanggan.

Tahun 2019, Maxim mengembangkan invasinya hingga sampai ke Kota Batam, sejak Februari 2019, aplikasi ini mulai beroperasi di Batam. Hingga pada Agustus 2019 jumlah mitra *driver* Maxim di Kota Batam sudah mencapai 9000 mitra. Tarif yang ditawarkan Maxim di kota ini cukup terjangkau dengan tarif

minimum untuk mobil sebesar Rp 5.000 per Km, sedangkan untuk sepeda motor sebesar Rp. 7000 per 4 Km. Hal ini tentu menjadikan Maxim sebagai salah satu penyedia jasa ojek *online* yang murah dan cukup terjangkau untuk mobilitas masyarakat Kota Batam.

## **2.7 Kelebihan dan Kekurangan Layanan Ojek *Online* Maxim**

Sebagai penyedia jasa layanan ojek *online* Maxim memiliki kelebihan dan kekurangannya tersendiri (Putri et al., 2023).

Kelebihan:

1. Tarif yang ekonomis, biaya perjalanan menjadi jelas setelah rute tercatat, dengan biaya awal mulai dari 3000 untuk perjalanan sepeda motor dan 5000 untuk mobil biasa. Tarif tetap stabil tanpa dipengaruhi oleh lalu lintas dan tidak berubah saat perjalanan berakhir, sehingga penumpang Maxim memiliki kendali penuh dalam mengelola anggaran transportasi mereka.
2. Layanan reservasi tersedia, aplikasi Maxim menawarkan fitur khusus yang jarang ditemukan, yakni kemampuan untuk melakukan pemesanan dan merencanakan rute yang rumit dengan beberapa alamat sekaligus. Selain itu, pengguna dapat berkomunikasi dengan pengemudi tanpa biaya tambahan, yang memberikan jaminan keselamatan tambahan.
3. Bagikan lokasi perjalanan, pengguna Maxim memiliki opsi untuk membagikan lokasi perjalanan dengan orang lain dengan mengirimkan tautan ke peta yang menampilkan rute perjalanan. Fitur ini sangat praktis, terutama ketika perlu menjemput teman di sepanjang rute atau ingin memastikan



bahwa orang yang Anda cintai sedang melakukan perjalanan dengan lancar dan nyaman.

Kekurangan:

1. Kurang populer dikalangan masyarakat, apabila membandingkan Maxim dengan layanan ojek *online* lainnya, dapat disimpulkan bahwa Maxim masih berada di bawah persaingan. Hal ini disebabkan oleh relatifnya keberadaan Maxim yang masih baru dalam pandangan masyarakat, yang menyebabkan sebagian besar orang lebih memilih untuk beralih ke layanan ojek *online* yang sudah lebih mapan dan terkenal.
2. Pelayanan kurang dipercaya, tingkat kepercayaan terhadap layanan yang disediakan oleh Maxim kurang kuat. Hal ini disebabkan oleh adanya sejumlah ulasan negatif dari pelanggan yang mengungkapkan ketidakpuasan mereka terhadap pelayanan Maxim. Banyak pelanggan merasa bahwa pelayanan yang diberikan oleh Maxim kurang memadai dan tidak memuaskan.
3. Tampilan aplikasi membingungkan, sebagian besar masyarakat masih mengalami kebingungan saat berinteraksi dengan antarmuka awal aplikasi Maxim. Mereka merasa kesulitan dalam hal memesan pengemudi Maxim dan berkomunikasi dengan pengemudi, termasuk cara menggunakan fitur chat.

## **2.8 *Sentiment Analysis***

*Sentiment analysis* atau analisis sentimen adalah sebuah studi komputasi yang berhubungan dengan sikap, emosi, pendapat, penilaian, serta opini dari sekumpulan teks yang tujuan dari analisis sentimen dilakukan adalah untuk melakukan ekstraksi, identifikasi atau menemukan karakteristik sentimen dalam unit teks

menggunakan metode NLP (*Natural Language Processing*), statistik, atau *machine learning*.

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 begitu, dalam implementasinya berbagai hambatan 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*.

### **2.8.1 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 di tahapan 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 preprocessing* 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-karakter yang tidak diperlukan untuk mengurangi derau atau *noise* di dalam dataset. Karakter-karakter yang dihapus dalam dokumen biasanya berupa *emoticon*, 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 tetapi 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.

#### 5. Tokenisasi

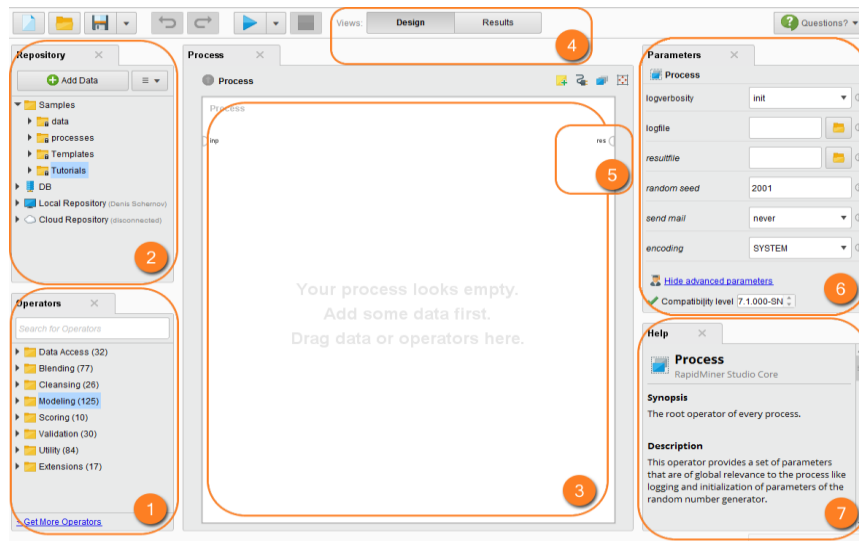
Tokenisasi adalah proses memisahkan setiap kata dari dalam dokumen yang digunakan, tokenisasi diperlukan untuk proses selanjutnya yaitu pembobotan dengan menggunakan TF-IDF dimana bobot setiap kata akan dinilai dalam vektor

### 2.9 *RapidMiner*

*RapidMiner* adalah perangkat lunak yang bersifat *open source* yang menyediakan solusi untuk menganalisis *data mining*, *text mining*, dan analisis prediksi. Dengan menggunakan berbagai teknik deskriptif dan prediksi, *RapidMiner* memberikan wawasan kepada pengguna sehingga mereka dapat membuat keputusan yang optimal. Perangkat ini memiliki lebih dari 500 operator *data mining*, termasuk operator untuk *input*, *output*, *preprocessing* data, dan visualisasi. *RapidMiner* adalah sebuah perangkat lunak independen untuk analisis data dan juga dapat diintegrasikan ke dalam produk lainnya. Perangkat lunak ini ditulis dalam bahasa *Java*, sehingga dapat berfungsi pada berbagai sistem operasi (Musfiroh et al., 2021).

Sebelumnya dikenal sebagai YALE (*Yet Another Learning Environment*), *RapidMiner* mengalami perkembangan awalnya pada tahun 2001 di bawah pengembangan oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer di Unit Kecerdasan Buatan dari *University of Dortmund*. *RapidMiner* didistribusikan dengan lisensi AGPL (*GNU Affero General Public License*) versi 3. Hingga saat ini, ribuan aplikasi yang dibangun menggunakan *RapidMiner* telah digunakan di lebih dari 40 negara. Sebagai perangkat lunak *open source* untuk *data mining*, *RapidMiner* telah mencapai reputasi yang sangat baik di seluruh dunia. Pada *polling* yang dilakukan oleh KDnuggets, sebuah portal data-mining pada tahun 2010-2011, *RapidMiner* bahkan meraih peringkat pertama dalam kategori Perangkat Lunak *Data Mining*.

*RapidMiner* menawarkan sebuah antarmuka grafis (*Graphic User Interface*) yang memungkinkan pengguna untuk merancang suatu alur kerja analisis. Antarmuka ini akan menghasilkan sebuah berkas XML (*Extensible Markup Language*) yang menggambarkan proses analisis yang diinginkan oleh pengguna untuk diterapkan pada data. File ini kemudian dapat dibaca oleh *RapidMiner* untuk menjalankan analisis secara otomatis.



**Gambar 2. 4** *User Interface RapidMiner*  
**Sumber:** (docs.rapidminer.com)

Berikut tabel di bawah ini akan menjelaskan fungsi dari setiap *user interface* *RapidMiner*.

**Tabel 2. 1** *Penjelasan Interface RapidMiner*

Nomor	Nama	Deskripsi
1	Operator	Blok bangunan yang digunakan untuk membuat proses <i>RapidMiner</i> .
2	<i>Repository</i>	Penyimpanan dalam <i>RapidMiner</i> Studio untuk data dan proses <i>RapidMiner</i> .
3	Panel Proses (Proses Utama)	Wilayah kerja untuk membangun proses.
4	Tampilan	Wilayah kerja untuk mengakses fungsionalitas tertentu.
5	Port	Mekanisme masukan dan keluaran untuk operator dan proses. Lihat daftar singkatan port dan maknanya.
6	Parameter	Pengaturan yang memodifikasi perilaku operator
7	Bantuan	Bantuan yang sensitif terhadap konteks untuk operator yang dipilih

**Sumber:** (Data Penelitian, 2024)

## 2.10 Penelitian Terdahulu

Penelitian mengenai *data mining* maupun analisis sentimen atau penggunaan algoritma *Support Vector Machine* (SVM) telah banyak dilakukan sebelumnya, untuk memberikan gambaran dan perbandingan mengenai penelitian ini, beberapa penelitian tersebut akan dipaparkan di bawah ini.

1. (Lestari et al., 2023) Judul penelitian “Penerapan Algoritma *Support Vector Machine* Pada Analisis Sentimen Terhadap Identitas Kependudukan Digital”. Penelitian ini berfokus pada pro dan kontra atas kehadiran Identitas Kependudukan Digital (E-KTP) yang terjadi di masyarakat. Data yang digunakan di *crawling* dari komentar Facebook, dan dilakukan analisis sentimen dengan menggunakan algoritma *Support Vector Machine* dan menghasilkan akurasi 77%. Dari 902 data yang diproses 78.27% mengandung sentimen negatif, 12.97% netral, dan 8.76% positif. Hal ini mengindikasikan bahwa sebagian masyarakat tidak puas akan kehadiran identitas kependudukan digital.
2. (Idris et al., 2023) Judul penelitian “Analisis Sentimen Terhadap Penggunaan Aplikasi *Shopee* Menggunakan Algoritma SVM”. Penelitian ini berfokus pada analisis sentimen terkait komentar positif dan negatif pengguna aplikasi *Shopee*. Data sebanyak 3,000 ulasan diambil melalui *scraping* dari ulasan pengguna di *Playstore*. Ulasan tersebut dibagi menjadi negatif (bintang 1-3) dan positif (bintang 4-5). Melalui proses SVM dengan parameter berbeda, penelitian mencapai akurasi tertinggi sebesar 98% dan f1-score 98%.

3. (Entini, 2023) Judul penelitian “Implementasi *Data Mining* dengan Algoritma *Naive Bayes* untuk Klasifikasi Kelayakan Penerima Bantuan Sembako”. Penelitian ini mengambil studi kasus kelayakan penerima bantuan sembako di kota Batam yang terus meningkat dari tahun 2022, penelitian ini menggunakan algoritma *Naive Bayes* dan dibantu dengan *software RapidMiner* lalu dilakukan pengujian lebih lanjut dengan perhitungan manual sehingga didapat nilai akurasi sebesar 80%.
4. (Simanjuntak et al., 2022) Judul penelitian “*Data Mining* Untuk Klasifikasi Status Pandemi COVID 19”. Penelitian ini dilakukan dengan tujuan untuk memprediksi pandemic COVID 19 dengan menggunakan teknik *data mining* algoritma *Naive Bayes* dan diimplementasikan menggunakan perangkat lunak Tanagra. Hasil dari penelitian ini adalah pengelompokan pandemic COVID 19 dengan probabilitas prediksi benar 90,7% dan prediksi salah sebesar 9,3%.
5. (Pranata et al., 2022) Judul penelitian “Klasifikasi Sentimen Terhadap Maxim Menggunakan Algoritma SVM Pada Media Sosial Twitter”. Penelitian ini difokuskan pada analisis sentimen terhadap *tweet* yang berbicara tentang layanan ojek *online* Maxim. Hasil penelitian menunjukkan bahwa SVM dengan kernel RBF dan kernel *polynomial* mencapai akurasi tertinggi, yaitu 85%. Selain itu, penelitian juga menyimpulkan bahwa kebanyakan komentar terhadap Maxim bersifat positif.
6. (Pamungkas et al., 2021) Judul penelitian “Analisis Sentimen Twitter Menggunakan Metode SVM pada Kasus Benih Lobster 2020”. Penelitian ini berfokus pada klasifikasi sentimen terkait topik Korupsi Benih Lobster



sebagai komoditas ekspor Indonesia. Tujuannya adalah untuk melihat persentase hasil klasifikasi *tweet*, yang dapat menjadi rekomendasi bagi KPK untuk meningkatkan kinerja dan strategi dalam penyelidikan kasus korupsi lainnya. Hasil penelitian menunjukkan bahwa 28% dataset dikategorikan sebagai emosi negatif dan 72% sebagai emosi positif, mengindikasikan dukungan mayoritas masyarakat terhadap KPK dalam menangani korupsi di Indonesia. Selain itu, penelitian mencapai akurasi sebesar 84.21%, sensitivitas sebesar 73.38%, dan nilai spesifisitas sebesar 82.10%.

7. (Faturrahman et al., 2022) Judul penelitian “Klasifikasi Sentimen Ulasan Film Menggunakan *Support Vector Machine*, *Information Gain*, dan *N-Grams*”. Penelitian ini menggabungkan pendekatan *machine learning* dengan *feature selection* untuk mengoptimalkan waktu komputasi dan performa model. Metode klasifikasi SVM terbukti unggul dalam analisis teks, dengan fokus pada implementasi *Information Gain* (IG) dan *N-Grams*. Hasilnya menunjukkan akurasi tertinggi sebesar 86.8% pada klasifikasi SVM dengan unigram tanpa IG dan *stopword removal*, sementara *Naive Bayes* mencapai akurasi tertinggi 80% dengan bigram, IG, dan *stopword removal*. Penelitian ini menggaris bawahi pentingnya pemilihan fitur dan parameter dalam analisis sentimen.
8. (Prastyo et al., 2021) Judul Penelitian “*A Combination Expansion Ranking and GA SVM for Improving Indonesian Sentiment Classification Performance*”. Penelitian ini menggunakan pendekatan yang menggabungkan *Quality Evaluation Research* (QER) dan *Genetic Algorithm*-

*Support Vector Machine* (GA-SVM) untuk pemilihan fitur dalam analisis sentimen. Penelitian ini berhasil menghasilkan performa klasifikasi yang tinggi dengan nilai *precision* sebesar 96.78%, *recall* sebesar 96.76%, *f1-measure* sebesar 96.75, dan nilai AUC sebesar 98.66%, serta berhasil mengurangi waktu komputasi secara signifikan.

9. (Borg & Boldt, 2020) Judul penelitian “*Using VADER Sentiment and SVM for Predicting Customer Response Sentiment*”. Penelitian ini melakukan analisis sentimen terhadap pelayanan sebuah perusahaan Telekomunikasi di Swedia. Data ulasan yang digunakan adalah Email pengguna perusahaan tersebut. VADER bersama leksikon bahasa swedia digunakan untuk memberikan label awal terhadap sentimen, kemudian data tersebut digunakan untuk melakukan pelatihan dari model SVM. Hasil penelitian ini menghasilkan *F1-score* rata-rata 0.8834 dan AUC rata-rata sebesar 0.896.
10. (Arsi & Waluyo, 2021) Judul penelitian “*Analisis Sentimen Wacana Pemindahan Ibu Kota Negara Indonesia Menggunakan Algoritma Support Vector Machine*”. Penelitian ini menggunakan algoritma SVM untuk melakukan analisis sentimen terkait topik pemindahan ibu kota negara Indonesia ke Kalimantan Timur, hasil dari penelitian ini menghasilkan nilai akurasi dari algoritma SVM sebesar 96.68%, *precision* sebesar 95.82%, *recall* = 94.04%, dan AUC sebesar 0.979%.
11. (Refo et al., 2022) Judul Penelitian “*Penerapan SVM untuk Klasifikasi Sentimen pada Review Comment Berbahasa Indonesia di Online Shop*” Penelitian ini menggunakan dataset komentar atau umpan balik dari

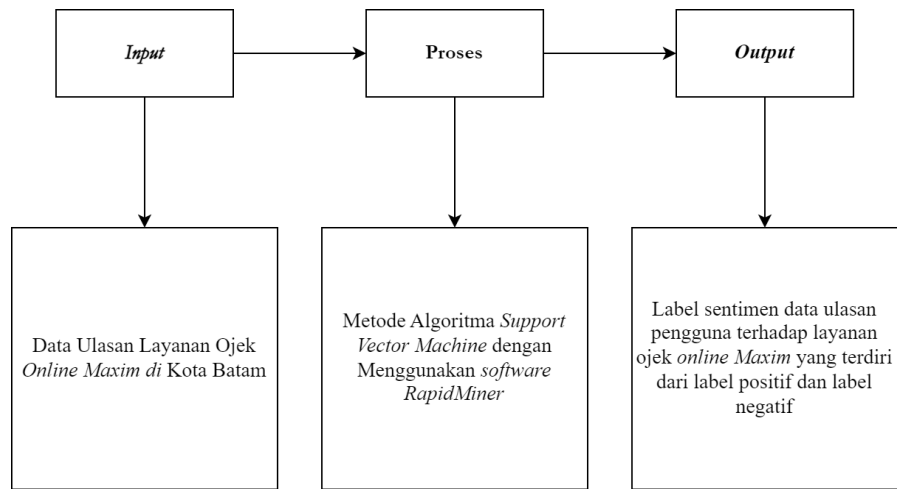
pembelian barang di online shop. Hasil penelitian ini adalah data yang telah melalui tahap pra proses memiliki tingkat akurasi yang lebih baik daripada data yang tidak di praproses. Hasil klasifikasi menggunakan kernel berbeda adalah 88.30% untuk kernel RBF, 87.40% untuk kernel polynomial, dan 87.40% untuk kernel linear.

12. (Darwis et al., 2020) Judul penelitian “Penerapan Algoritma SVM untuk Analisis Sentimen pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia”. Penelitian ini menggunakan metode *Support Vector Machine* (SVM) dengan kernel linear dalam analisis sentimen terkait KPK RI. yang menghasilkan akurasi akhir sebesar 82%. Hasil penelitian menunjukkan bahwa sentimen negatif terhadap kinerja KPK RI mencapai 77%, sedangkan akurasi keseluruhan mencapai 82%.
13. (Tineges et al., 2020) Judul penelitian “Analisis Sentimen Terhadap Layanan *Indihome* Berdasarkan Twitter dengan Metode Klasifikasi SVM”. Penelitian ini fokus pada analisis sentimen menggunakan SVM terhadap layanan *Indihome*, berdasarkan komentar-komentar di Twitter. Tujuan utamanya adalah menilai akurasi metode SVM dan tingkat kepuasan pengguna *Indihome* berdasarkan komentar di Twitter. Hasil sentimen menunjukkan bahwa hanya 18.4% dari komentar dianggap positif, mengindikasikan ketidakpuasan masyarakat terhadap layanan *Indihome*.
14. (Pravina et al., 2019) Judul penelitian “Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritma SVM”. Penelitian ini berfokus pada klasifikasi sentimen terkait opini

masyarakat terhadap maskapai penerbangan. Penelitian ini mencoba meningkatkan akurasi dengan menerapkan *lexicon-based features* untuk mengidentifikasi polaritas kata dalam kamus sentimen yang berisi kata positif dan negatif pada data latih. Hasil akhir penelitian ini menunjukkan akurasi sebesar 40%, precision 40%, recall 100%, dan f1-measure 57.1455, dengan parameter optimal berupa *learning rate* (gamma) 0.03 dan nilai C sebesar 10.

## 2.11 Kerangka Pemikiran

Kerangka penelitian adalah bentuk penelitian ini berfokus pada analisis sentimen terhadap layanan ojek *online* Maxim di Kota Batam berdasarkan data ulasan pengguna. *Input* utama dalam penelitian ini adalah data ulasan pengguna yang mencakup berbagai aspek dalam layanan tersebut. Untuk mengolah data ini, penelitian ini menggunakan metode Algoritma *Support Vector Machine* (SVM) dengan bantuan perangkat lunak *RapidMiner*. Algoritma SVM digunakan untuk mengklasifikasikan sentimen ulasan pengguna menjadi dua label, yaitu label positif dan label negatif. Proses ini akan membantu dalam mengevaluasi persepsi dan pandangan pengguna terhadap layanan ojek *online* Maxim, sehingga perusahaan dapat mendapatkan wawasan yang berharga untuk meningkatkan kualitas layanan di Kota Batam. Dengan demikian, kerangka pemikiran ini memberikan landasan penting dalam melakukan pengolahan dan analisis data ulasan pengguna dengan tujuan akhir mengidentifikasi sentimen yang terkandung dalam ulasan tersebut. Di bawah ini merupakan kerangka pemikiran dalam penelitian ini.



**Gambar 2. 5** Kerangka Pemikiran

Sumber: (Data Penelitian, 2024)