

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Gambaran Umum Objek Penelitian

Gambaran Umum pada objek penelitian ini akan menampilkan data dari sistem pengoperasian berbeda tapi terstruktur dan siap untuk proses analisis. Biasanya data-data ini akan sangat berguna untuk *predictive modeling*, *machine learning* dan pengaplikasian analisis lainnya. Gambaran umum objek penelitian yaitu melakukan pengamatan terhadap pengguna *e-commerce* yang ada di Indonesia. Pengamatannya berupa review pengguna baik pada aplikasi playstore maupun review dalam aplikasi Shopee, Bukalapak, Lazada, dan Tokopedia.

**Tabel 3. 1** E-commerce terbaik di indonesia[25]

Shopee	Lazada	Bukalapak	Tokopedia
<ul style="list-style-type: none"><li>• Harga produk terjangkau</li></ul>	<ul style="list-style-type: none"><li>• Reputasi ulasan yang baik</li></ul>	<ul style="list-style-type: none"><li>• Harga murah</li></ul>	<ul style="list-style-type: none"><li>• Barang lengkap</li></ul>
<ul style="list-style-type: none"><li>• Reputasi ulasan yang baik</li></ul>	<ul style="list-style-type: none"><li>• Harga terjangkau</li></ul>	<ul style="list-style-type: none"><li>• Barang lengkap</li></ul>	<ul style="list-style-type: none"><li>• Didominasi UMKM</li></ul>
<ul style="list-style-type: none"><li>• Sering ada event dan gratis ongkir</li></ul>	<ul style="list-style-type: none"><li>• Barang lengkap</li></ul>	<ul style="list-style-type: none"><li>• Didominasi UMKM</li></ul>	<ul style="list-style-type: none"><li>• Banyak promosi yang dilakukan</li></ul>

Berdasarkan tabel 3.1 mendapat gambaran alasan Masyarakat memilih e-commerce tersebut . Masyarakat memilih e-commerce shopee karena harga produk terjangkau, memiliki reputasi ulasan yang baik pada penjualan produk, selain itu shopee sering mengadakan event seperti 11.11 sehingga barang discount dan memiliki ongkir gratis. Sedangkan alasan Masyarakat memilih Lazada karena Lazada memiliki reputasi ulasan yang baik, harga barang barang yang terjangkau, dan memiliki variasi barang yang terbilang lengkap. Sedangkan alasan Masyarakat memilih bukalapak karena harga murah, memiliki variasi barang yang lengkap, dan

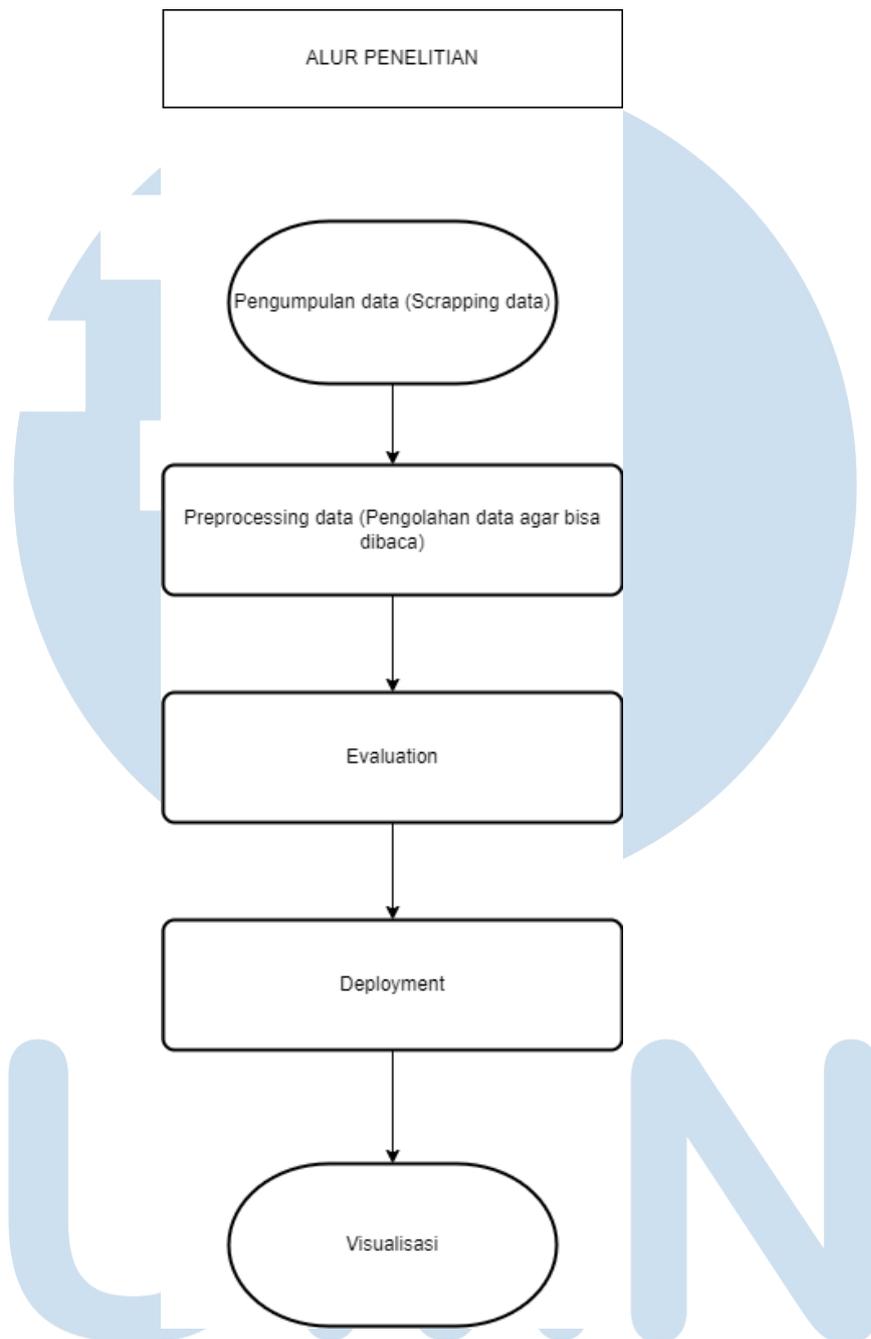
didominasi oleh pelaku usaha UMKM. Sedangkan alasan Masyarakat memilih Tokopedia karena Tokopedia memiliki variasi barang yang lengkap, didominasi oleh pelaku usaha UMKM, dan Banyak promosi yang dilakukan[25].

### **3.2 Metode Penelitian**

Metode penelitian menggunakan algoritma clustering yaitu algoritma klasifikasi. Algoritma klasifikasi adalah suatu algoritma yang mengidentifikasi berdasarkan kelompok yang ada. Algoritma yang harus digunakan adalah *Support Vector Machine* , *Naïve Bayes Classifier* dan *K-Nearest Neighbors (K-NN)*. Metode penelitian menggunakan *Support Vector Machine*, *Naïve Bayes Classifier* dan *K-Nearest Neighbors (K-NN)* karena *Support Vector Machine*, *Naïve Bayes Classifier* dan *K-Nearest Neighbors (K-NN)* lebih mudah untuk dibaca dan ditafsirkan tanpa perlu memahami statistik lebih mendalam.

#### **3.2.1 Alur Penelitian**

Alur Penelitian menggunakan metode CRISP-DM (Cross-Industry Standard Process for Data Mining). CRISP-DM diadopsi secara luas dan memiliki dokumentasi yang luas serta dukungan dari komunitas data mining. Hal ini umumnya dipandang sebagai metodologi yang praktis dan efektif untuk proyek penambangan data.



**Gambar 3.1** Alur penelitian

### 3.3 Teknik Pengumpulan Data

Teknik pengumpulan data untuk penelitian dilakukan dengan cara melakukan *scrapping data review* pada google playstore tentang aplikasi *e-commerce* yang ada di indonesia, dan melakukan review berdasarkan komentar komentar yang

diberikan oleh pengguna masing masing *e-commerce* terhadap barang barang, pelayanan, kepuasan pelanggan yang ada pada *e-commerce* tersebut.

### 3.3.1 Scraping data

*Scraping data* adalah metode pengumpulan data dengan cara mengambil data dari *website* yang ditentukan. Pada penelitian ini menggunakan *google playstore* untuk mengambil data dari komentar-komentar pengguna aplikasi *e-commerce*.

### 3.4 Variabel Penelitian

Penelitian ini bertujuan untuk menganalisis sentimen pelanggan dalam manajemen e-commerce dengan menggunakan algoritma *Support Vector Machine*, *Naive Bayes Classifier* dan *K-Nearest Neighbors (K-NN)* pada konteks *Big Data*. Variabel penelitian yang dilakukan menggunakan Variabel Independen, Variabel Dependen, dan Variabel Kontrol. Alasan memilih variabel independen karena Melibatkan jumlah dan kompleksitas data yang dihasilkan oleh interaksi pelanggan dan ulasan produk pada platform *e-commerce*. Alasan memilih variabel dependen karena mencerminkan opini, perasaan, atau tanggapan pelanggan terhadap produk atau layanan di platform e-commerce. Alasan memilih variabel kontrol karena menyelidiki tingkat akurasi, ketepatan waktu, dan kebersihan data yang nantinya digunakan dalam analisis sentiment.

### 3.5 Teknik Analisis Data

Tabel 3. 2 Perbandingan metode data mining untuk analisis data[26]

KDD	CRISP-DM
Data Cleaning	Business Understanding
Data Integration	Data Understanding
Data Transformation	Data Preparation
Data Mining	Modeling
Pattern Evolution	Evaluation
Knowledge Presentation	Deployment

Berdasarkan tabel 3.2 KDD dan CRISP-DM sama-sama memiliki tujuan untuk melakukan data mining dengan efektif. Tujuan dari KDD adalah untuk

menghasilkan pengetahuan baru yang dapat membantu perusahaan dalam pengambilan keputusan yang lebih baik. Tujuan dari CRISP-DM adalah untuk membantu perusahaan dalam mengelola dan menganalisis data dengan efektif, serta menghasilkan wawasan baru yang dapat mendukung pengambilan keputusan yang lebih baik. Perbedaan dari keduanya adalah CRISP-DM menyediakan standar proses baku untuk data mining yang dapat diterapkan ke dalam strategi pemecahan masalah umum pada bisnis atau pada unit penelitian[26]. Oleh karena itu merupakan suatu alasan saya lebih memilih Teknik analisis CRISP-DM.

Teknik analisis data yang dilakukan adalah CRISP-DM. Berikut ini adalah langkah-langkah yang dilakukan untuk analisis dan mengolah data sehingga hasil dari penelitian ini dapat dimanfaatkan oleh orang banyak. Berikut ini adalah langkah-langkah dari CRISP-DM:

### **3.5.1 Business Understanding**

*Business Understanding* adalah tahap pertama dalam Teknik analisis CRISP-DM. Pada tahap ini membutuhkan pengetahuan dan riset yang luas tentang dunia bisnis, mulai dari cara mendapatkan data yang bertujuan untuk membangun bisnis dari data yang didapatkan tersebut. Pada tahap ini melakukan penelitian dari beberapa orang pengguna *e-commerce* untuk mengetahui bagaimana tingkat kepercayaan pengguna dengan melakukan pendaftaran dengan data pribadi untuk penggunaan *e-commerce*.

### **3.5.2 Data Understanding**

*Data Understanding* adalah tahap untuk memeriksa data dan melakukan identifikasi masalah yang ada pada data tersebut. Pada *data Understanding* ini data yang telah dikumpulkan dari data *scrapping* dianalisa dan disesuaikan dengan algoritma yang ada agar proses CISDM dapat berjalan dengan lancar.

### **3.5.3 Data preparation**

*Data preparation* adalah tahap untuk memperbaiki masalah yang ada dalam data. Pada tahap ini memastikan data yang tepat untuk algoritma yang digunakan. *Data* didapatkan dengan menggunakan metode *scrapping* pada *google play store* di *e-commerce*. Data yang dikumpulkan akan berbentuk *.csv*.

### **3.5.4 Modeling**

*Modelling* adalah tahap untuk membuat model prediktif atau deskriptif. Pada tahap ini menggunakan metode *Machine learning* yang bertujuan untuk merapikan data agar data lebih terstruktur. Tujuan data dirapikan agar bisa dikembangkan untuk melihat akurasi hasil yang didapatkan dari data tersebut. Data dirapikan dengan cara di preprocessing agar dapat dibaca dan dianalisis.

### **3.5.5 Evaluation**

*Evaluation* adalah tahap hasil dari proses yang dilakukan pada tahap modeling sebelumnya. Hasil akhir dari modeling tersebut akan di lakukan evaluasi lebih mendalam apakah memenuhi tujuan penelitian atau tidak. Pada tahap evaluation ini akan menghasilkan hasil analisis berupa sentimen analisis positif, negatif, dan netral. Selain itu juga menghasil *wordcloud* yang akan menampilkan kata yang sering banyak muncul pada ulasan.

### **3.5.6 Deployment**

Deployment adalah tahap akhir dengan hasil yang telah di evaluasi dan menghasilkan skor keputusan akhir yang akurat dan sudah memenuhi tujuan untuk dilakukan suatu penelitian. Pada tahap Deployment ini selain akurasi juga akan menghasilkan visualisasi data

## **3.6 Algoritma yang digunakan untuk pengujian**

Pengujian sentimen Analisis menggunakan beberapa algoritma yaitu Support Vector Machine, Naïve Bayes Classifier dan K-Nearest Neighbors (K-NN) dengan tujuan menemukan akurasi terbaik[27]. Penelitian ini menggunakan tiga algoritma berbeda yaitu, Support Vector Machine, K-NN, dan Naïve Bayes untuk mengekstrak pemikiran atau perasaan pengguna melalui review mereka dan mengelompokkan datanya ke dalam kategori yang berbeda[28]. Berdasarkan pendekatan terhadap klasifikasi yang telah dilakukan seperti Naïve Bayes, Support Vector Machine, dan K-Nearest Neighbor(K-NN) telah diterapkan untuk menemukan hasil paling optimal[28]. Klasifikasi adalah suatu langkah dimana kita mencari suatu pola atau aturan tertentu yang dapat mengelompokkan data ke dalam kategori-kategori tertentu. Hal ini bertujuan untuk membangun suatu model yang dapat digunakan untuk memprediksi kategori atau kelas dari suatu objek

berdasarkan atribut-atributnya, bahkan jika kelasnya belum diketahui sebelumnya[28]

### **3.6.1 Support Vector Machine**

*Support Vector Machine (SVM)* merupakan metode untuk memproyeksikan hasil prediksi di masa depan, baik dalam hal menganalisis pola data untuk memperkirakan nilai numerik (regresi) maupun untuk mengelompokkan data ke dalam kategori atau kelas tertentu (klasifikasi) [29]. Metode SVM dimanfaatkan untuk mencari hyperplane terbaik yang dapat memisahkan data dengan nilai target yang berbeda [29]. Support Vector Machine (SVM) merupakan kumpulan teknik pembelajaran terawasi yang memeriksa data serta mengidentifikasi corak, dipakai untuk mengklasifikasikan serta menganalisis regresi[19]. Analisis sentimen dengan menggunakan Support Vector Machine (SVM) mampu mengklasifikasikan opini dalam kalimat untuk membedakan sentimen positif atau negatif [19].

### **3.6.2 Naïve Bayes**

Algoritma *Naïve Bayes* adalah merupakan salah satu algoritma klasifikasi yang banyak digunakan pada *data mining* ataupun *text mining*[30]. *Naïve, Bayes*, adalah metode, machine learning untuk probabilitas[31]. Dalam kata lain, Naïve, Bayes, merupakan, metode untuk, klasifikasi, text, dengan, kecepatan pemrosesan yang tinggi jika dalam data besar[31]. Menghasilkan data yang specific dari masing-masing literatur dalam efisiensi penggunaan metode *naïve bayes* terhadap prediksi kelulusan mahasiswa tepat waktu[32].

### **3.6.3 K-Nearest Neighbor (K-NN)**

K-Nearest Neighbor (KNN) adalah salah satu algoritma klasifikasi supervised learning yang digunakan untuk mengkasifikasikan objek berdasarkan atribut kelas dan data training[33]. Algoritma k-NN adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data training yang jaraknya paling dekat dengan objek tersebut[34]. Salah satu metode atau algoritma yang berfungsi untuk mengklasifikasikan setiap opini menjadi suatu sentimen positif, sentimen negatif, maupun hanyalah sentimen netral ialah K-Nearest Neighbor (K-NN)[35].