

BAB III

METODOLOGI PENELITIAN

3.1. Gambaran Umum Objek Penelitian

Objek penelitian yang akan digunakan merupakan komentar pada postingan media sosial *Instagram* dari akun *E-Commerce*. *E-Commerce* yang akan digunakan sebagai objek adalah *Shopee*, *Tokopedia*, *Bukalapak*, *Lazada*, dan *Blibli*. Pemilihan 5 jenis *E-Commerce* didasarkan data jumlah pengunjung *E-Commerce* dari website *iPrice*. *iPrice Group* merupakan website yang beroperasi di beberapa negara di Asia Tenggara yang secara teratur membahas berbagai wawasan mengenai topik yang berkaitan dengan *E-Commerce*, *Startup*, dan berbagai hal lainnya [26]. Pada komentar akan dilihat komentar negatif dan positif dari setiap postingan tersebut. Akun yang digunakan sebagai objek adalah sebagai berikut :

3.1.1. Shopee

Shopee merupakan platform *E-Commerce* terkemuka di Asia Tenggara dan Taiwan. Pada tahun 2015, *Shopee* diluncurkan di Singapura, Indonesia, Malaysia, Thailand, Vietnam, Filipina, dan Taiwan yang memberikan pelanggan pengalaman belanja *online* yang mudah, aman, dan cepat. Akun media sosial *Instagram* yang digunakan sebagai objek penelitian merupakan @shopee_id [27].

3.1.2. Tokopedia

Tokopedia merupakan perusahaan teknologi Indonesia yang memiliki misi untuk mencapai pemerataan ekonomi secara digital. *Marketplace Tokopedia* telah menyediakan berbagai pilihan produk dari Indonesia dan telah bekerja sama dengan lebih dari 10 juta penjual. Akun media sosial Instagram yang digunakan sebagai objek penelitian merupakan @tokopedia [28].

3.1.3. Bukalapak

Bukalapak merupakan perusahaan teknologi Indonesia yang bermisi untuk menciptakan perekonomian yang adil untuk masyarakat Indonesia. Dengan menggunakan *platform online* dan *offline*, *Bukalapak* dapat memberikan kesempatan dan pilihan kepada semua orang untuk mendapatkan hidup yang lebih baik dan layak. Akun media sosial *Instagram* yang digunakan sebagai objek penelitian merupakan @bukalapak [29].

3.1.4. Lazada

Lazada merupakan destinasi belanja dan berjualan *online* di Asia Tenggara yang hadir di Indonesia, Malaysia, Filipina, Singapura, Thailand, dan Vietnam. Melalui *platform marketplace* yang didukung oleh berbagai macam layanan, *Lazada* telah membantu lebih dari 135.000 penjual lokal dan internasional. Akun media sosial *Instagram* yang digunakan sebagai objek penelitian merupakan @lazada_id [30].

3.1.5. Blibli

Blibli merupakan *e-commerce* yang berasal dari Indonesia dengan *model* bisnis yang berfokus pada *B2B*, *B2C*, dan *B2B2C* (*Business to Business to Customer*). *Blibli* menawarkan berbagai pilihan produk yang berkualitas dan disediakan oleh lebih dari 100.000 mitra usaha. Akun media sosial Instagram yang digunakan sebagai objek penelitian merupakan @blibliidotcom [31].

3.2. Variabel Penelitian

3.2.1. Variabel Independen

Variabel Independen merupakan variabel yang menjadi penyebab atau terpengaruhnya variabel terikat [32]. Berikut merupakan variabel independen pada penelitian ini :

1. Tanggapan dari komentar *Instagram* akun *E – Commerce Shopee*
2. Tanggapan dari komentar *Instagram* akun *E – Commerce Tokopedia*
3. Tanggapan dari komentar *Instagram* akun *E – Commerce Bukalapak*
4. Tanggapan dari komentar *Instagram* akun *E – Commerce Lazada*
5. Tanggapan dari komentar *Instagram* akun *E - Commerce Blibli*

3.2.2. Variabel Dependen

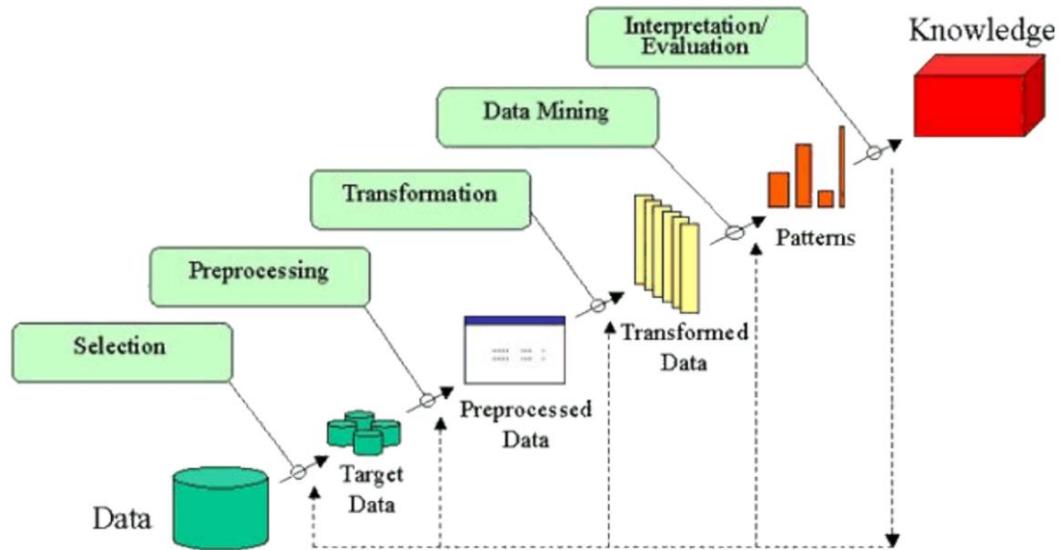
Variabel Dependen merupakan variabel terikat yang dipengaruhi karena adanya variabel independen [32]. Variabel Dependen pada penelitian ini merupakan tanggapan yang bersifat positif dan negatif sehingga terdapat 2 *class*

klasifikasi yaitu *positive* dan *negative*. Variabel tersebut bergantung kepada setiap variabel independen yaitu masing – masing tanggapan dari komentar *Instagram* akun *E – Commerce* tersebut.

3.3. Metode Pengumpulan Data

Metode pengumpulan *data* pada penelitian ini menggunakan *PyCharm* untuk melakukan pengumpulan *data* dan akan mengambil komentar dari akun media sosial *Instagram E – Commerce*. Berdasarkan jurnal penelitian serupa dapat diketahui bahwa penggunaan *data* sebanyak 1.000 dapat menghasilkan hasil yang valid. Namun, semakin banyak *data* yang digunakan maka *data* tersebut akan semakin valid sehingga *data* yang akan diperoleh merupakan 10.000 komentar yang dibagi menjadi 2.000 untuk 5 akun *Instagram E – Commerce* [33]. Pembagian *data* akan dilakukan dengan menggunakan teknik *cluster random sampling* yang membagi *E - Commerce* menjadi 5 *cluster* dengan jumlah data sebesar 2.000 per *E-Commerce* berdasarkan 5 jumlah urutan tertinggi pengunjung *E-Commerce* dari *data website iPrice* dan akan mengambil *data* mulai dari Maret 2020 hingga 2021 [25].

3.4. Metode Penyelesaian Masalah



Gambar 3. 1. Alur KDD

(Sumber : Asroni et al., 2018)

3.4.1. Data Selection

Pada tahapan ini akan dilakukan seleksi terhadap *attribute* yang akan digunakan dalam penelitian. Berdasarkan tujuan penelitian yaitu untuk melakukan *review* terhadap 5 *E-Commerce* maka *attribute* yang akan digunakan dalam penelitian ini adalah komentar yang didapatkan dari akun *Instagram* 5 *E-Commerce* yaitu *Shopee, Tokopedia, Bukalapak, Lazada, dan Blibli*.

3.4.2. Pre-Processing

Pada tahapan ini *data* yang telah di *mining* akan di *Pre-Processing* untuk dapat digunakan di tahapan selanjutnya dan akan menggunakan aplikasi

Rapid Miner. Berikut merupakan tahapan – tahapan dari *Pre-Processing* yang akan dilakukan pada penelitian ini :

1. *Case Folding*

Tahapan untuk mengubah setiap kata menjadi bentuk yang sama yaitu menjadi kata dengan huruf kapital atau tidak [8]. *Case Folding* digunakan untuk menghindari adanya *data* duplikasi dengan huruf yang berbeda.

2. *Data Cleansing*

Tahapan untuk menghilangkan delimiter koma, titik, dan tanda baca lainnya [8]. *Data Cleansing* digunakan untuk menghilangkan karakter atau simbol yang tidak memiliki fungsi dan makna dalam analisis sentimen.

3. Normalisasi Bahasa / *Word Replacer*

Tahapan untuk menormalisasi bahasa pada bahasa yang tidak baku [8]. Normalisasi Bahasa digunakan untuk mengembalikan kata yang disingkat menjadi kata sebelum disingkat agar saat pembobotan kata tersebut dapat dihitung sebagai satu kesatuan.

4. *Stopword Removal*

Tahapan untuk menghilangkan daftar kata umum yang artinya tidak penting dan tidak digunakan [8]. *Stopword Removal* digunakan untuk menghindari penggunaan kata yang tidak memiliki makna dan fungsi dalam analisis sentimen.

5. *Stemming*

Tahapan untuk mencari kata dasar dari hasil *stopword removal* [8].

Stemming digunakan untuk mengembalikan kata yang ada dalam penelitian menjadi kata dasar agar pada pembobotan kata tersebut tidak dihitung lebih dari sekali.

6. Tokenisasi

Tahapan untuk memotong dokumen menjadi beberapa pecahan kecil yang dapat berupa paragraf, kalimat, dan kata [8]. Tokenisasi digunakan untuk mempermudah dalam melakukan pembobotan kata dengan memecahkan setiap kalimat menjadi pecahan kata.

7. *Generate Bigram*

Tahapan untuk menggabungkan 2 kata menjadi 1 token agar dapat meningkatkan konteks dalam kalimat tersebut dan meningkatkan akurasi dari model [13]. *Generate Bigram* digunakan untuk menyatukan 2 kata yang sebelumnya sudah dipecah menjadi 1 agar makna dalam kata tersebut dapat berubah seperti kata tidak yang negatif dan kata jelek yang negatif digabungkan menjadi tidak jelek yang bermakna positif.

3.4.3. Transformation

Pada tahapan ini *data* akan diubah agar sesuai dengan kebutuhan penelitian. Pada penelitian ini, *data* yang sudah dibersihkan akan diubah supaya *data* dapat diimplementasikan ke algoritma yang akan digunakan.

3.4.4. Data Mining

Pada tahapan ini akan dipilih *model* yaitu *Support Vector Machine* dan *Naïve Bayes* dengan menggunakan pendekatan metode *K-Fold Cross*

Validation. Penggunaan algoritma akan menghasilkan hasil akurasi untuk melihat seberapa tepat algoritma dalam melakukan klasifikasi.

3.4.5. Interpretation

Pada tahapan ini akan didapatkan hasil prediksi dan klasifikasi dari tahapan sebelumnya. Hasil dari klasifikasi dan prediksi akan divisualisasikan menggunakan aplikasi *Tableau*