

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Gambaran Umum Objek Penelitian

Penelitian ini membahas pemanfaatan figur publik, baik dari kalangan artis Indonesia maupun Korea, sebagai *brand ambassador* untuk produk lokal Indonesia di *platform* media sosial X. Tujuan utama dari studi ini adalah untuk mengidentifikasi adanya perbedaan sentimen publik terhadap penggunaan artis Korea sebagai duta merek produk lokal. Melalui pendekatan analisis sentimen, studi ini mengeksplorasi persepsi serta reaksi pengguna media sosial terhadap keterlibatan artis Korea dan artis Indonesia dalam kampanye pemasaran di *platform* X. Dalam proses analisis, digunakan metode klasifikasi berbasis algoritma Naïve Bayes, *Support Vector Machine* (SVM), dan *Random Forest* untuk mendeteksi dan mengkategorikan sentimen pengguna ke dalam tiga kategori: positif, negatif, dan netral. Sumber data berasal dari komentar-komentar yang ditulis oleh pengguna di *platform* X, yang kemudian diolah menjadi data tekstual untuk dianalisis. *Tweet* ini digunakan untuk menilai respons publik terhadap pemanfaatan artis korea sebagai *brand ambassador* produk lokal. Melalui pemanfaatan teknik klasifikasi seperti *Naïve Bayes*, SVM, dan *Random Forest* penelitian ini mengelompokkan opini pengguna berdasarkan sentimen yang ditunjukkan, guna memahami sejauh mana tanggapan mereka terhadap produk lokal yang memakai artis Korea sebagai duta merk mereka. Hasil dari kajian ini diharapkan dapat memberikan masukan strategis bagi pelaku usaha dan pemasar produk makanan lokal, terutama dalam merancang strategi promosi yang efektif di media sosial. Temuan ini juga dapat menjadi acuan penting dalam pengambilan keputusan terkait pemilihan *brand ambassador* yang paling sesuai dengan preferensi konsumen di *platform* X.

#### 3.2 Metode Penelitian

##### 3.2.1 Alur Penelitian

Metode CRISP-DM diterapkan sebagai alur penelitian untuk

mengklasifikasi data dalam studi ini. CRISP-DM ialah suatu metode yang mempunyai 6 tahapan dalam melakukan prosesnya. Berikut dibawah ini merupakan tahapan dalam penelitian ini :

1. *Business Understanding*

Memahami tujuan dari analisis sentimen pemanfaatan artis Korea sebagai duta merk produk lokal Indonesia serta mengidentifikasi bagaimana persepsi pengguna X pada produk lokal yang melibatkan artis Korea sebagai duta merk dan melihat apakah ada perbedaan yang signifikan dalam sentimen antara keduanya.

2. *Data Understanding*

Mengumpulkan data yang berasal dari *tweet* X terhadap penggunaan artis Korea sebagai *brand ambassador* produk lokal Indonesia dengan Teknik *scraping* data.

3. *Data Preparation*

Melakukan *preprocessing data* yang ditujukan untuk melakukan *cleansing* dan *filtering* pada data yang sudah dikumpulkan dari kolom komentar *tweet* X.

4. *Modelling*

Melakukan pengujian data untuk dilakukan analisis sentimen. Pendekatan yang digunakan dalam proses pembuatan model adalah dengan menggunakan *machine learning* yaitu algoritma *Naïve Bayes*, *SVM (Support Vector Machine)*, dan *Random Forest* yang digunakan untuk memodelkan sentimen terhadap produk lokal Indonesia yang memanfaatkan artis Korea sebagai *brand ambassador* mereka pada kolom komentar *tweet* X, kemudian hasil yang diberikan merupakan klasifikasi data ke dalam kategori positif, negatif atau netral.

5. *Evaluation*

Mengevaluasi hasil dari analisis sentimen pemanfaatan artis Korea sebagai *brand ambassador* produk lokal Indonesia dan memeriksa tingkat ke akurasian terhadap 3 model yang telah digunakan,

kemudian data yang telah diuji akan diubah menjadi vektor fitur TF-IDF yang akan dilatih secara bersamaan dengan model *Naïve Bayes*, SVM, dan *Random Forest*. Setelah data uji berhasil diubah, model akan melakukan prediksi terhadap data tersebut dan akan menghasilkan label prediksi untuk setiap sampel yang terdapat dalam data yang di uji. Hasil prediksi yang telah didapatkan kemudian akan dibandingkan dengan label sebenarnya yang berasal dari data uji agar dapat digunakan untuk menghitung metrik evaluasi yang akan disajikan dalam bentuk presisi, *recall*, serta *f1-score* serta *confussion matrix* untuk menunjukkan pendistribusian hasil prediksi yang benar serta salah dalam setiap kelasnya.

#### 6. *Deployment*

Merupakan tahapan dimana hasil dari visualisasi model yang sudah diterapkan untuk penelitian ini.

### 3.2.2 Metode Pengembangan Sistem / Metode Data Mining

Dalam studi ini dijalankan perbandingan antara dua metode yaitu algoritma *Naïve Bayes*, SVM serta *Random Forest* guna mengukur perbandingan akurasi tertinggi yang dihasilkan oleh kedua metode algoritma yang diterapkan dalam melakukan analisis sentimen mengenai produk lokal yang menggunakan artis Korea sebagai *brand ambassador* produk mereka.

### 3.3 Teknik Pengumpulan Data

Dalam penelitian ini metode pengumpulan data yang diterapkan adalah Teknik *Scrapping* data. Data yang dikumpulkan berasal dari *tweet* pada *platform X* yang melibatkan artis Indonesia dan artis Korea sebagai *brand ambassador* produk lokal Indonesia dari tahun 2021-2025.

### 3.4 Teknik Analisis Data

Metode analisis data yang digunakan dalam penelitian ini mengacu pada pendekatan CRISP-DM. Penelitian ini memilih pendekatan CRISP-DM dikarenakan kerangka kerjanya yang memberikan pemahaman menyeluruh

terhadap tujuan dan kebutuhan bisnis, serta kemampuannya untuk mendukung penerapan model secara optimal. Salah satu keunggulan utama dari CRISP-DM ialah kapasitasnya guna menangani dan menganalisis data yang telah terstruktur dengan baik. Oleh karena itu, CRISP-DM dipilih dalam penelitian ini karena dianggap mampu dalam mencakup seluruh tahapan proses data, mulai dari identifikasi kebutuhan bisnis hingga pemrosesan data dan implementasi terhadap model yang sesuai[28].

#### **3.4.1 Business Understanding**

Pada tahap *Business Understanding*, penelitian ini bertujuan untuk mengevaluasi adanya perbedaan persepsi masyarakat terhadap keterlibatan artis Korea sebagai duta merek produk lokal Indonesia, serta menyajikan visualisasi dari hasil analisis sentimen pengguna di *platform X*. Diharapkan, hasil analisis ini mampu memberikan pemahaman yang lebih komprehensif terkait pandangan dan kecenderungan pengguna, sehingga dapat dijadikan dasar dalam pengambilan keputusan yang lebih strategis terkait pemanfaatan artis Korea dalam kampanye pemasaran produk lokal. Dalam penelitian ini digunakan tiga metode klasifikasi, yakni *Naïve Bayes*, *Support Vector Machine*, dan *Random Forest*, untuk melakukan analisis sentimen. Proses implementasi dilakukan menggunakan platform Google Colab dan Jupyter Notebook dengan bahasa pemrograman Python.

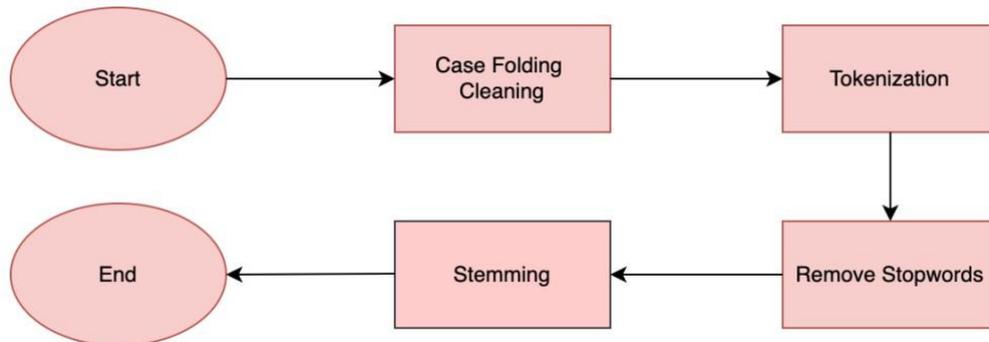
#### **3.4.2 Data Understanding**

Tahap *Data Understanding* diawali dengan mengumpulkan informasi yang didapatkan dari para pengguna *platform X*. Tahapan awal dalam proses ini melibatkan pengambilan komentar-komentar dari beberapa threads yang relevan pada *platform X*. kemudian, dikumpulkanlah data yang menggunakan Bahasa pemrograman Python. Data yang diambil adalah data pada kolom komentar *tweet X* dari tahun 2021 sampai tahun 2025 yang selanjutnya hasil dari pengumpulan data ini akan disimpan dalam bentuk CSV.

#### **3.4.3 Data Preparation**

### 3.4.3.1 Data Preprocessing

*Data preprocessing* adalah tahapan dimana data akan diubah dengan cara membersihkan isi dari teks sehingga nantinya akan mendapatkan data yang sesuai. Berikut dibawah ini merupakan tahapan-tahapan dalam melakukan *data preprocessing* :



Gambar 3. Tahapan data preprocessing

Gambar 3.1 merupakan urutan dalam tahapan *data preprocessing* yang terdiri dari enam langkah utama. Tahapan pertama adalah *case folding*, yaitu proses mengonversi seluruh huruf dalam teks menjadi huruf kecil. Tahap kedua adalah pembersihan, yang bertujuan menghilangkan elemen tidak relevan seperti tanda pagar, simbol, dan emoji. Proses ketiga yaitu *tokenization*, dimana kalimat dipecah menjadi berbagai unit kecil yang disebut sebagai token. Selanjutnya adalah tahap *stopword removal*, yaitu tahapan dimana kata-kata umum seperti konjungsi, kata ganti, dan preposisi yang tidak memiliki makna penting dalam analisis akan dihapus. Tahapan terakhir adalah *stemming*, yakni proses mengubah kata-kata yang mempunyai imbuhan menjadi bentuk dasar atas *root word*-nya.

### 3.4.3.2 Data Labeling

Pada tahap ini, setiap data akan diberikan kategori berupa sentimen positif, negatif ataupun netral. Tujuan dari proses ini adalah untuk mengukur tingkat akurasi sejinnga performa dari masing-masing algoritma dapat diketahui saat proses evaluasi model dilakukan. Dalam

penelitian ini, proses pelabelan memanfaatkan pustaka transformers dari Python, yang memungkinkan klasifikasi data ke dalam label sentiment positif, negatif dan netral secara otomatis.

#### **3.4.3.4 Split Data**

Tahapan ini melibatkan pemisahan data ke dalam dua bagian, yaitu data pelatihan (*training*) yang digunakan untuk membangun model klasifikasi, dan data pengujian (*testing*) yang berfungsi untuk mengevaluasi kinerja model yang telah dilatih. Pada penelitian ini, data yang digunakan dibagi menggunakan rasio 80 banding 20, dimana 80% dialokasikan untuk pelatihan dan sisanya 20% digunakan untuk pengujian.

#### **3.4.3.5 Modeling**

Pada tahap pemodelan, digunakan tiga metode klasifikasi yaitu *Naïve Bayes*, *SVM* serta *Random Forest*. Ketiga algoritma ini akan dibandingkan guna menilai serta menentukan model prediktif mana yang memiliki performa paling optimal. Proses ini dijalankan dengan menggunakan platform Jupyter dan Bahasa pemrograman Python sebagai alat bantu analisis.

#### **3.4.3.6 Evaluation**

Dalam tahap evaluasi, dilakukan analisis terhadap hasil pengujian ketiga algoritma yakni *Naïve Bayes*, *SVM* serta *Random Forest*, dengan menerapkan metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-Score*. Berdasarkan hasil dari kinerja model tersebut, akan dipilih algoritma terbaik yang selanjutnya akan digunakan untuk menyimpulkan hasil akhir serta memberikan kesimpulan dari pemanfaatan artis Korea sebagai duta merk produk lokal.

#### **3.4.3.7 Deployment**

Tahap ini adalah tahapan penutup dalam metode CRISP-DM. Pada bagian ini, data dimasukkan ke dalam model analisis yang telah dibangun menggunakan algoritma *Naïve Bayes*, *Support Vector Machine*, dan *Random Forest* untuk menghasilkan sebuah akurasi yang di dapatkan setiap algoritma yang digunakan.