

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Data yang digunakan dalam penelitian analisis sentimen ini berasal dari ulasan pengguna di platform *Play store* tentang aplikasi Spotify, yang akan menjadi fokus penelitian. Data aplikasi Spotify berisi *feedback* serta pendapat dari masyarakat. Tahap selanjutnya data analisis sentimen akan dianalisis dan topik utamanya mengenai peningkatan layanan melalui ulasan pengguna. Penelitian ini akan berfokus pada hasil perbandingan kedua algoritma yaitu *Naïve Bayes* dan *Support Vector Machine* untuk menganalisis sentimen *user* spotify. Selanjutnya hasil analisis tersebut akan dipakai untuk meningkatkan kualitas layanan spotify.

Penelitian ini menguji hasil model untuk menemukan algoritma terbaik dalam menangani kasus sentimen aplikasi Spotify. Pengujian model menggunakan 2 algoritma yaitu *Support Vector Machine* dan *Naïve Bayes*. Alasan memilih algoritma *Support Vector Machine* dan *Naïve Bayes* karena, pada penelitian sebelumnya penelitian sentimen terhadap analisis *rating* dan *review* menghasilkan perbandingan antara *Support Vector Machine* dan *Naïve Bayes* dengan tingkat akurasi lebih baik dimiliki oleh *Support Vector Machine*. Selain itu, *Support Vector Machine* memiliki tingkat akurasi yang sangat tinggi dibandingkan dengan algoritma lainnya. Namun, pada penelitian lainnya menyebutkan bahwa *Naïve Bayes* mendapatkan hasil lebih baik jika dibandingkan dengan algoritma seperti k-NN [32][33].

3.2 Metode Penelitian

Penelitian ini akan menganalisis sentimen publik berdasarkan data aplikasi Spotify dari Play Store. Dalam kasus *machine learning* metode yang sering digunakan untuk menganalisis sentimen data adalah *Support Vector Machine*

(SVM) dan *Naïve Bayes*. Berikut tabel 3.1 dibawah merupakan tabel perbandingan *Support Vector Machine* dan *Naïve Bayes*[34][35].

Tabel 3 1 Perbandingan *Support Vector Machine, Naïve Bayes, dan KNN* [34][35]

Perbandingan	<i>Support Vector Machine (SVM)</i>	<i>Naïve Bayes (NB)</i>	<i>K- Nearest Neighbors (KNN)</i>
Kinerja pada <i>dataset</i>	Cenderung memberikan kinerja yang baik pada <i>dataset</i> kecil hingga menengah	Cenderung memberikan kinerja yang baik pada <i>dataset</i> kecil hingga besar	Cenderung memberikan kinerja yang baik pada <i>dataset</i> kecil hingga menengah.
<i>Training data</i>	Memerlukan jumlah data pelatihan yang cukup besar untuk memberikan hasil yang baik, terutama pada model yang kompleks	Lebih sederhana dengan jumlah data pelatihan yang lebih kecil dan dapat memberikan hasil yang baik meskipun dengan <i>dataset</i> yang terbatas.	Memerlukan jumlah data pelatihan yang lebih besar, terutama data yang kompleks atau memiliki dimensi fitur yang tinggi
Kecepatan pelatihan dan prediksi	Memakan waktu lebih lama, terutama untuk <i>dataset</i> besar.	Memakan waktu yang lebih cepat karena sederhana, terutama pada <i>dataset</i> yang cukup besar	Memakan waktu yang lambat, apabila <i>dataset</i> yang digunakan besar khususnya dimensi fitur tinggi
Atribut	Lebih cocok untuk data numerik dan kontinu	Dapat menangani baik atribut numerik maupun kategorikal tanpa perlu tranformasi tambahan	Dapat menangani baik atribut numerik maupun kategorikal tanpa perlu transformasi tambahan

Pemilihan *Support Vector Machine* dan *Naïve Bayes* untuk analisis sentimen pada ulasan aplikasi Spotify didasarkan pada keunggulan masing-masing metode yang saling melengkapi. SVM terkenal karena kemampuannya dalam menangani data yang kompleks dan besar. Algoritma ini dapat membedakan antara kelas-kelas yang berbeda dengan sangat baik, meskipun data memiliki dimensi yang tinggi dan banyak fitur, yang sering ditemukan dalam ulasan pengguna yang beragam dan rumit. Di sisi lain, *Naïve Bayes* adalah algoritma yang cepat dan sederhana, serta memiliki performa yang sangat baik dalam pemrosesan data teks. Kecepatan dan efisiensi *Naïve Bayes* sangat penting dalam konteks analisis ulasan pengguna, karena sering

melibatkan sejumlah besar teks yang membutuhkan pemrosesan cepat untuk mendapatkan hasil analisis yang *real-time*.

3.3 Teknik Pengumpulan Data

Penelitian ini menggunakan data primer. Data diperoleh dengan scraping dari Play Store menggunakan bahasa pemrograman Python melalui Google Colab. Data yang diambil merupakan opini-opini dari Google Play store terkait dengan aplikasi Spotify. Opini yang diambil adalah opini yang hanya berbahasa Indonesia. *Output* data berupa file .csv.

Data diambil dalam periode waktu 18 Agustus 2023 hingga 30 Januari 2024. Penentuan periode waktu ditentukan berdasarkan ulasan *most relevant*. *Most relevant* berarti data yang mendominasi atau sering muncul terhadap komentar setiap pengguna

3.4 Teknik Pengambilan Sampel

Populasi yang akan diambil untuk penelitian ini adalah ulasan dari Google Play store pada aplikasi Spotify. Teknik *sampling* yang digunakan adalah Teknik *purposive sampling*. *Purposive sampling* adalah sebuah metode pengambilan sampel dimana peneliti telah secara jelas menetapkan fokus dan tujuan penelitian. Sampel diambil dengan menggunakan *library Google-play-scraper* melalui Google Colab dengan menggunakan bahasa pemrograman Python.

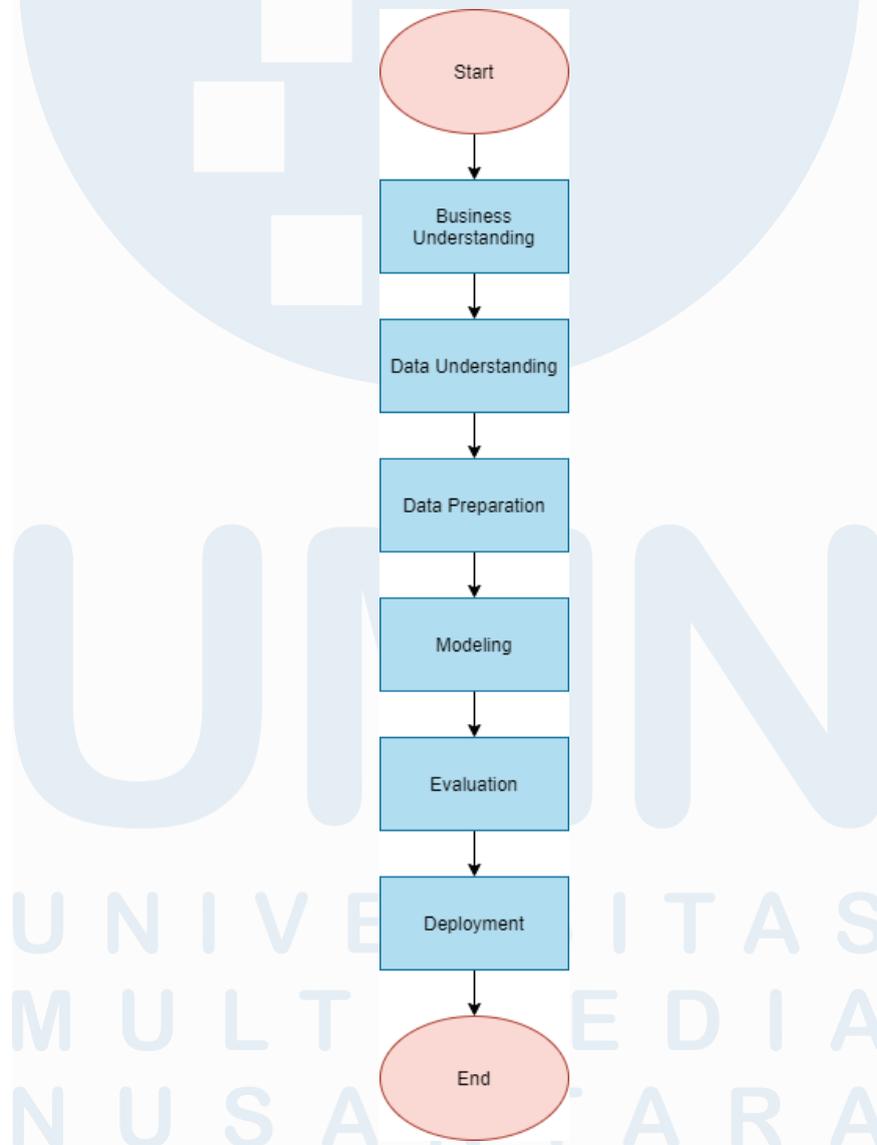
3.5 Teknik Analisis Data

Dalam penelitian ini, teknik analisis data yang diterapkan adalah *Cross Industry Standard Process for Data Mining (CRISP-DM)* [36][25]

Tabel 3. 2 Perbandingan *Framework Data Mining* [36][25]

CRISP-DM	KDD
<i>Business Understanding</i>	<i>Pre KDD</i>
<i>Data Understanding</i>	<i>Selection</i>
	<i>Preprocessing</i>
<i>Data Preparation</i>	<i>Transformation</i>
<i>Modelling</i>	<i>Data Mining</i>
<i>Evaluation</i>	<i>Interpretation/Evaluation</i>
<i>Deployment</i>	<i>Post KDD</i>

Tabel 3.2 merupakan tabel perbandingan Framework Data Mining. Penelitian ini memilih menggunakan CRISP-DM karena fokusnya yang kuat pada pemahaman yang mendalam terhadap kebutuhan dan tujuan bisnis serta implementasi model yang efektif. Kelebihan CRISP-DM terletak pada kemampuannya dalam menganalisis dan memodelkan data yang sudah terstruktur dengan baik, sementara KDD lebih sesuai untuk eksplorasi data yang belum terstruktur. CRISP-DM dipilih karena mampu mendukung seluruh proses, mulai dari pemahaman bisnis hingga pengolahan data dan penerapan model. Berikut pada gambar 3.1 dibawah merupakan tahapan CRISP-DM yang terdiri dari enam tahap:



Gambar 3 1 Alur Penelitian

3.6.1. Business Understanding

Dalam tahap *Business Understanding*, penelitian ini bertujuan untuk meningkatkan layanan aplikasi Spotify berdasarkan pemahaman terhadap sentimen pengguna dari ulasan. Hasil dari analisis sentimen ini diharapkan dapat memberikan pemahaman yang lebih mendalam tentang preferensi dan harapan pengguna, sehingga dapat menjadi dasar untuk pengambilan keputusan yang lebih baik dalam mengembangkan dan meningkatkan layanan Spotify di masa depan. Pendekatan ini menggunakan dua algoritma klasifikasi, yaitu *Support Vector Machine* dan *Naïve Bayes* untuk melakukan analisis sentimen. *Tools* yang digunakan yaitu Google Colab dengan bahasa pemrograman Python.

3.6.2 Data Understanding

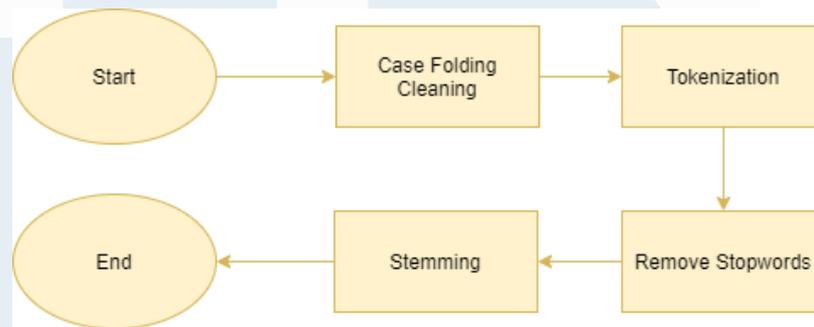
Proses *Data Understanding* dimulai dengan mengumpulkan informasi yang relevan dari ulasan pengguna aplikasi Spotify. Tahap pertama melibatkan pengambilan ulasan menggunakan Google Play Store dengan menggunakan fitur *library* `google_play_scraper` untuk memperoleh informasi. Selanjutnya, data dikumpulkan menggunakan bahasa pemrograman Python. Pengambilan data dilakukan sejak 18 Agustus 2023 hingga 30 Januari 2024, dengan fokus pada ulasan "most relevant". Hasil pengumpulan data disimpan ke dalam format CSV yang terdiri dari beberapa kolom, termasuk:

- a. *username* (Nama Pengguna)
- b. *rating* (*rating* 1-5)
- c. *date* (Tanggal dan waktu)
- d. *review* (Komentar/*Feedback*)

3.6.3 Data Preparation

3.6.3.1 Data preprocessing

Tahap data *pre-processing* merupakan tahap mengubah data dengan cara membersihkan isi teks sehingga menghasilkan data sesuai dengan yang diinginkan. Berikut merupakan tahap-tahap melakukan *preprocessing*, antara lain:



Gambar 3.2 Tahapan *Data preprocessing*

Gambar 3.2 merupakan tahapan *preprocessing* dibagi ke dalam enam langkah. Langkah pertama disebut *case folding*, yang merupakan proses untuk mengubah teks huruf besar menjadi huruf kecil. Selanjutnya, terdapat langkah *cleaning*, yang melibatkan pembersihan elemen-elemen yang dianggap tidak relevan, seperti *hashtag*, simbol dan emoji. Langkah ketiga adalah *Tokenization*, di mana teks dipisahkan menjadi bagian yang lebih kecil yang disebut 'token'. Langkah selanjutnya adalah *stopword*, di manakata umum seperti preposisi, kata ganti, dan kata sambung dihapus. Terakhir merupakan langkah *stemming*, yang melibatkan perubahan kata imbuhan menjadi kata dasar.

3.6.3.2 Data Labeling

Setelah mengumpulkan data, langkah berikutnya adalah pelabelan data. Data-data diberikan label positif dan negatif. Tujuan pemberian pelabelan data ini dilakukan untuk menguji akurasi agar dapat mengetahui performa dari masing-masing model pada saat

dilakukan pengujian model. Data *Labeling* pada penelitian ini menggunakan *library transformers* Python. *Library transformers* merupakan model untuk memberikan label atau kategori kelas sentimen positif dan negatif.

3.6.3.2 Split Data

Pada tahap ini, Pemisahan data dilakukan dengan memisahkan data membangun model klasifikasi dan validasi (*training*) dan menguji model yang sudah dibangun (*testing*). Umumnya *dataset* untuk *training* dan *testing* dengan rasio pembagian 90:10. 90% data *training* dan 10% data *testing*.

3.6.4 Modeling

Pada tahap *Modeling* menggunakan dua algoritma klasifikasi yaitu *Support Vector Machine* dan *Naïve Bayes*. Dua algoritma akan dibandingkan untuk mengevaluasi dan memilih model prediksi dengan kinerja terbaik. Proses ini akan dilakukan dengan memanfaatkan Google Colab dan bahasa pemrograman Python.

3.6.5 Evaluation

Pada tahap evaluasi ini, dilakukan penilaian hasil pengujian kedua algoritma, yaitu *Support Vector Machine* dan *Naïve Bayes*, dengan menggunakan parameter seperti *accuracy*, *precision*, *recall*, dan *F1-Score*. Setelah menganalisis performa model, dipilih model yang memberikan hasil terbaik untuk mengambil kesimpulan akhir dan memberikan rekomendasi terkait peningkatan kualitas layanan aplikasi Spotify.

3.6.6 Deployment

Tahap *deployment* merupakan tahapan terakhir dari metode *CRISP-DM*. Pada tahap ini akan memasukkan data dan dijalankan ke model analisis menggunakan algoritma *Naïve Bayes* dan menghasilkan sebuah prediksi.