

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Tujuan dari penelitian ini adalah untuk menganalisis sentimen pengunjung terhadap Gunung Bromo. Industri pariwisata di Gunung Bromo menjadi fokus penelitian ini. Di Jawa Timur, Gunung Bromo merupakan salah satu tempat wisata terkenal yang tidak selalu sepi. Daya tarik utama Gunung Bromo adalah kondisinya yang aktif, pendakiannya yang relatif mudah, dan keunikan kawah Bromo yang dikelilingi pegunungan dan terletak di tengah kaldera Gunung Tengger yang sangat besar. Dengan menggunakan teknik *web scraping* dan bahasa komputer *Python* dan *Jupyter*, data ulasan pengunjung destinasi wisata Gunung Bromo dikumpulkan dari kolom komentar situs *TripAdvisor*.

3.2 Metode Penelitian

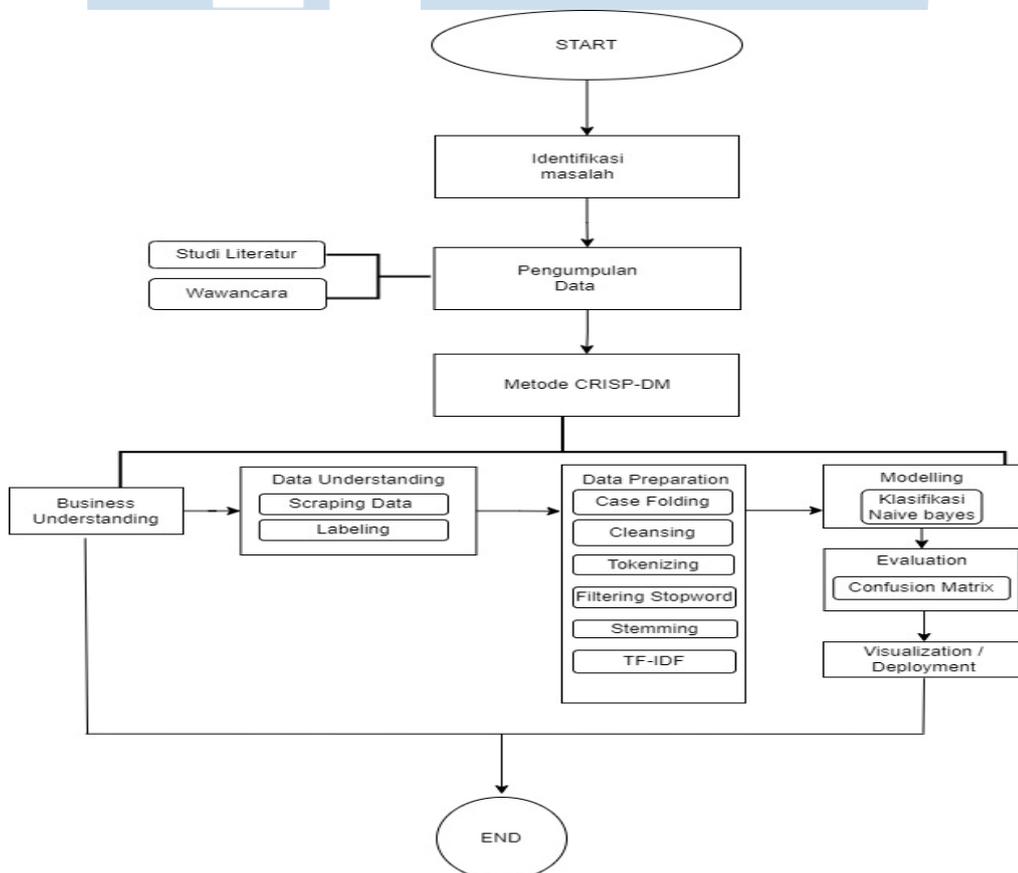
CRISP-DM adalah metodologi penelitian yang digunakan dalam penelitian ini. Sebagai sebuah teknik, *CRISP-DM* memberikan gambaran siklus hidup data mining jika dilihat sebagai model proses. Ini juga mencakup fase-fase proyek, pekerjaan yang dilakukan di setiap fase, dan penjelasan tentang hubungan antara aktivitas-aktivitas ini. Uraian tersebut mengisyaratkan bahwa pendekatan ini menawarkan prosedur standar yang dapat digunakan dalam suatu unit bisnis atau penelitian untuk menyelesaikan masalah secara umum atau non-eksklusif dengan memanfaatkan teknik data mining yang tepat. Enam fase *CRISP-DM* adalah: penerapan, penilaian, pemodelan, persiapan data, pemahaman bisnis, dan pemahaman data, fase tersebut akan digunakan untuk analisis dan hasil penelitian.

Naive Bayes adalah model pembelajaran mesin yang digunakan dalam penelitian ini, yang memanfaatkan teknologi notebook *Jupyter* dan bahasa pemrograman *Python*. *K-Fold Cross Validation* adalah langkah awal yang diselesaikan. *K-Fold Cross Validation* adalah proses yang dirancang untuk membedakan antara data pelatihan dan pengujian. Model dilatih dan dinilai K kali,

setiap kali dengan menggunakan set pengujian yang berbeda dan set pelatihan K-1 yang tersisa. Hasil evaluasi K adalah rata-rata kinerja akhir.

3.3. Alur Penelitian

Dalam menjalankan sebuah penelitian, penggunaan kerangka kerja atau metodologi yang terstruktur sangatlah penting untuk memandu setiap langkahnya. Salah satu metode yang umum digunakan adalah CRISP-DM (Cross-Industry Standard Process for Data Mining). Metode ini memberikan panduan yang terorganisir untuk menjalankan proses penelitian dari awal hingga akhir dengan efisien dan efektif.



Gambar 3. 1 Alur Penelitian

3.3.1. Identifikasi Masalah

Mewawancarai pihak-pihak yang terlibat untuk mengidentifikasi masalah adalah langkah pertama dalam prosedur ini. Fase ini berguna dalam mengidentifikasi masalah yang ada. Permasalahan tersebut dapat dijadikan topik pembahasan untuk penelitian. Dari hasil wawancara yang dilakukan,

permasalahan terjadi karena pengembangan objek wisata yang belum dikelola secara baik.

3.3.2. Pengumpulan Data

Pada tahap pengumpulan data terdapat 2 metode yang digunakan yaitu studi literatur dan wawancara, berikut merupakan penjelasan dari metode-metode tersebut:

a) Studi Literatur

Jurnal *online* dicari atau dibaca untuk studi literatur. Jurnal-jurnal ini akan menjadi sumber informasi untuk penelitian ini. Metode ini digunakan untuk menyajikan ide-ide berbeda yang berkaitan dengan permasalahan yang sedang diteliti atau dihadapi guna memberikan konteks dalam membicarakan temuan penelitian.

b) Wawancara

Wawancara dilakukan dengan pihak terkait pengelola objek wisata Gunung Bromo dengan membahas tentang permasalahan yang sedang terjadi di objek wisata Gunung Bromo dan dampaknya dari permasalahan tersebut.

3.3.3. Business Understanding

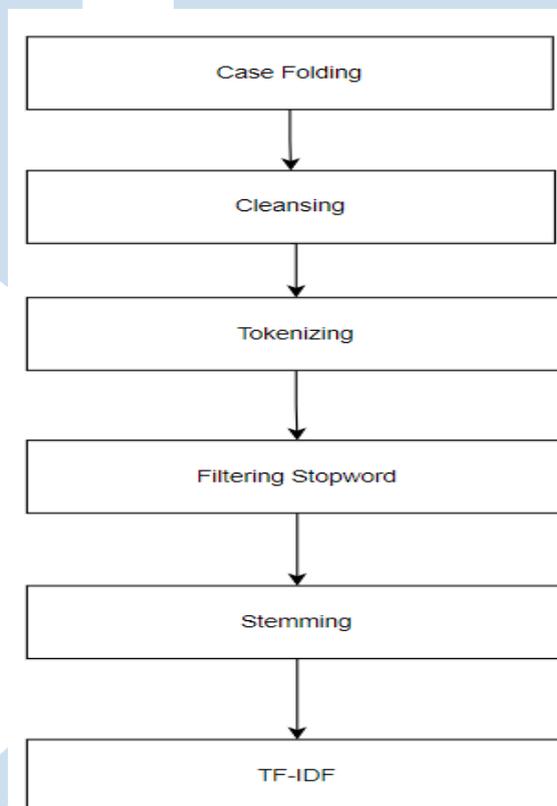
Fase yang dikenal sebagai "Pemahaman Bisnis" bertujuan untuk memahami tujuan dan spesifikasi proyek dari sudut pandang bisnis. Setelah itu, informasi ini diubah menjadi deskripsi tantangan data mining dan rencana proyek awal yang bertujuan untuk mencapai tujuan tersebut. penetapan konteks pembahasan fokus pada isu manajemen destinasi wisata Gunung Bromo yang dapat ditinjau berdasarkan persepsi wisatawan melalui data teks berupa ulasan pengunjung.

3.3.4. Data Understanding

Penelitian ini memanfaatkan informasi dari evaluasi pengguna *TripAdvisor* terhadap destinasi wisata Gunung Bromo. Teknik *scraping* situs *TripAdvisor* yaitu <https://www.tripadvisor.com/> digunakan untuk mengumpulkan data. Pengguna, tanggal, rating, dan ulasan adalah atribut yang dikumpulkan. Sebanyak seribu review destinasi wisata Gunung Bromo berhasil dikumpulkan.

3.3.5. Data Preparation

Tahap *Data Preparation* melibatkan pengolahan teks untuk mengeliminasi fitur-fitur yang tidak digunakan. Tahap ini sangat penting dalam kerangka kerja CRISP-DM karena kualitas data yang baik dan relevan sangat menentukan keberhasilan proses analisis. Dalam konteks analisis sentimen Gunung Bromo, data preparation akan melibatkan pengumpulan data dari berbagai sumber seperti media sosial, platform *review*, atau forum diskusi yang membahas tentang Gunung Bromo. Setelah data dikumpulkan, langkah-langkah berikut ini dapat dilakukan dalam tahap persiapan data. Tahapan *Data Preparation* yang digunakan terdiri dari *case folding*, *cleansing*, *tokenizing*, *filtering stopword*, *stemming*.



Gambar 3. 2 Tahap Data Preparation

Berikut ini merupakan penjelasan dari tahapan Data Preparation yang berurutan mulai dari tahap awal hingga akhir :

1. *Case Folding*

Salah satu jenis metode prapemrosesan untuk teks. Teknik ini bertujuan untuk mengubah seluruh huruf dokumen menjadi huruf kecil. Contoh

mengubah semua huruf dokumen menjadi huruf kecil; hanya huruf "a" sampai "z" yang dikenali. Selain huruf, karakter dihilangkan dan digunakan sebagai pembatas.

2. *Cleaning*

Langkah ini menghapus satu huruf, URL, dan karakter non-huruf (seperti tanda baca dan angka) dari data teks. Hal ini dilakukan karena karakter tunggal, tanda baca, URL, dan karakter non-huruf dianggap sebagai elemen yang tidak berguna dan tidak penting. Item-item ini, jika tidak dicentang, akan diubah menjadi karakteristik pemodelan yang mempengaruhi hasil evaluasi kinerja model.

3. *Tokenizing*

Pada bagian ini, teks diberi token yaitu, dibagi menjadi segmen atau atribut token. Teks dibagi menjadi beberapa bagian sesuai dengan ruang yang digunakan untuk tokenisasi dalam penelitian ini. Untuk menyederhanakan prosedur pada langkah berikut, langkah ini diperlukan.

4. *Filtering Stopwords*

Pemilihan istilah kunci dari hasil pencarian merupakan tahap penyaringan. Daftar kata (yang menyimpan kata-kata kunci) atau metode daftar berhenti dapat digunakan untuk menghilangkan istilah-istilah yang kurang penting. *Stopwords* berpotensi menambah lebih banyak karakteristik yang tidak berguna dan menurunkan performa model akhir jika tidak dihilangkan. Perpustakaan NLTK digunakan untuk menghilangkan kata-kata berhenti. Ada 758 *stopwords* dalam koleksinya.

5. *Stemming*

Kata-kata diubah menjadi bentuk paling dasar pada saat ini. Hal ini perlu dilakukan karena kata-kata yang mempunyai imbuhan yang berasal dari kata dasar yang sama diperlakukan sebagai kata-kata tersendiri. Di sinilah dilakukan proses penggabungan berbagai bentuk kata menjadi satu representasi.

6. *TF-IDF*

Untuk memberikan bobot pada setiap fitur digunakan pembobotan TF-IDF. Setiap atribut diberi nilai bobot karena semuanya mempunyai tingkat

relevansi dan signifikansi yang berbeda-beda. Salah satu teknik konvensional yang sering digunakan untuk pembobotan fitur pada dokumen teks adalah TF-IDF, yang digunakan dalam pembobotan fitur penelitian ini.

7. Pelabelan Data

Penelitian ini menggunakan kelompok emosi positif dan negatif. Pustaka *VADER (Valence Aware Dictionary for Sentiment Reasoning)* digunakan untuk mengidentifikasi persyaratan pelabelan data yang diperlukan untuk membuat model klasifikasi. Berdasarkan kepekaan terhadap pola sentimen positif atau negatif dan kekuatan emosi, perpustakaan ini dapat memastikan sentimen suatu teks [11], [12]. Nilai sentimen teks akan ditentukan oleh *VADERS*. Ditetapkan sebagai kelas emosi negatif jika nilainya kurang dari nol, dan sebagai kelas sentimen positif jika nilainya lebih dari nol.

3.3.6. Modeling

Pada tahapan ini, data tinjauan yang telah diproses sebelumnya akan digunakan untuk pemodelan. Data tinjauan sekarang akan dibagi menjadi dua tahap yaitu data terpisah dan validasi silang. Ini akan dipisahkan menjadi data pelatihan dan pengujian. Pembelajaran yang diawasi adalah algoritma pembelajaran mesin yang digunakan dalam penelitian ini. *Naive Bayes* adalah teknik pembelajaran mesin yang populer untuk klasifikasi.

3.3.7. Evaluation

Matriks kebingungan adalah metode untuk menilai metode klasifikasi. Setelah pengumpulan data dan pra-pemrosesan yang efektif, kumpulan data akhir ditampilkan dengan atributnya yang disusun menurut kelas sentimen. Hasil visualisasi menunjukkan istilah mana yang sering digunakan pada setiap jenis sentimen.

3.4. Teknik Pengumpulan Data

Penelitian ini menggunakan metode penelitian kuantitatif, yaitu akan dilakukan pengumpulan data dari situs *Trip Advisor* dengan menggunakan aplikasi *Octoparse*. Metode kuantitatif diterapkan dengan mengimplementasikan scraping

data dari situs *TripAdvisor* untuk memperoleh data ulasan objek wisata Gunung Bromo.

3.5. Teknik Pengambilan Sampel

Teknik Random Sampling digunakan dalam penelitian ini yaitu dengan menggunakan pendekatan random sampling, setiap orang dalam populasi mempunyai peluang yang sama untuk dipilih menjadi anggota sampel, baik secara sendiri-sendiri maupun secara kolektif. Selain sampel acak dengan nomor acak, peneliti mengumpulkan sampel dari sejumlah pasien pada jarak yang sama yang telah ditentukan.

3.6. Teknik Analisis Data

Penerapan *framework CRISP-DM* pada penelitian ini memanfaatkan tools data mining yang dipilih berdasarkan perbandingan. Di bawah ini merupakan perbandingan antara tools *RapidMiner*, dan *Jupyter Notebook*:

Tabel 3. 1 Perbandingan Rapid Miner dan Jupyter Notebook

Performance	Rapid Miner	Jupyter Notebook
Access	Free Community Edition, Commercial Enterprise Edition	Open Source
Kelebihan	-Mempunyai kapabilitas algoritma terbanyak -Mempunyai user interace yang menarik dan lebih grafis (GUI / graphic user interface).	-Tampilan yang interaktif -Mendukung berbagai visualisasi -Mendukung berbagai bahasa pemograman
Kekurangan	Diharuskan mempunyai license RapidMiner community untuk mengakses aplikasi.	-Sulit untuk bereksperimen -Tidak teorganisir -Sulit untuk di debug

Sumber:[63]

Berdasarkan tabel 3.3 terdapat perbandingan tools data mining, penelitian ini menggunakan tools *Jupyter Notebook* dikarenakan terdapat beberapa keunggulan dibandingkan dengan tools lain nya yaitu sisi kecepatan, *Jupyter Notebook* lebih unggul dibandingkan dengan *Rapid Miner* [64].