

BAB III

METODOLOGI PENELITIAN

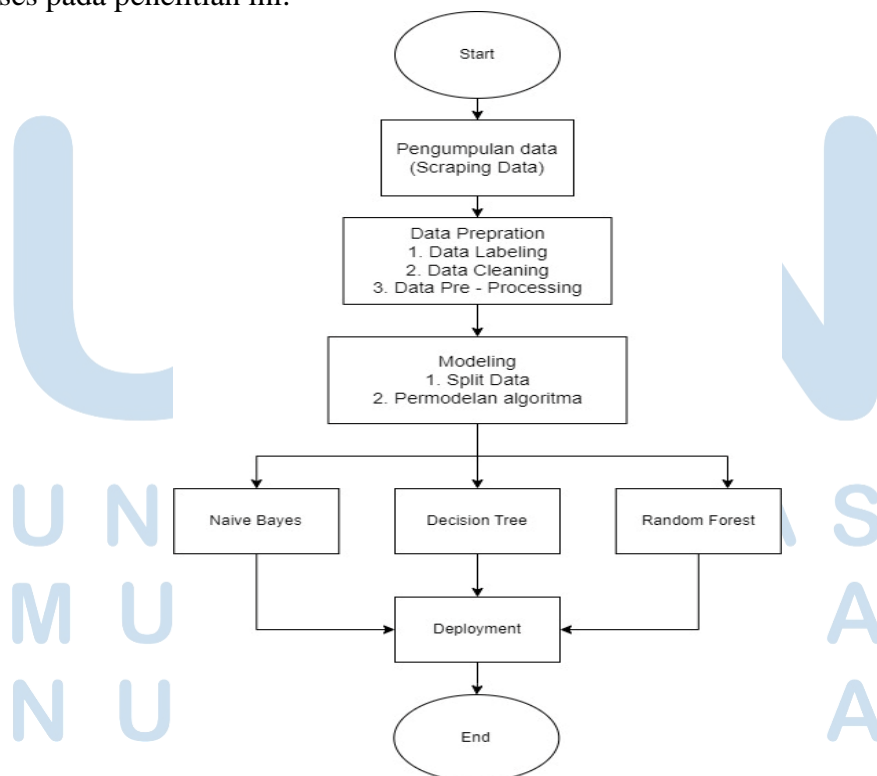
3.1 Gambaran Umum Objek Penelitian

Topik penelitian ini adalah analisis data ulasan atau review dari Google Play Store mengenai aplikasi WeTV untuk menggali bagaimana opini pengguna tentang aplikasi tersebut berkembang sepanjang waktu. Data tersebut akan diolah terlebih dahulu sebelum dibagi menjadi dua kategori sentimen, yaitu positif dan negatif. Selanjutnya, akan dilakukan analisis sentimen menggunakan metode Naïve Bayes, Decision Tree, dan Random Forest. Data yang digunakan dalam penelitian ini mencakup ulasan dari tahun 2021 hingga 2024, memungkinkan untuk menganalisis perubahan opini pengguna dari tahun ke tahun.

3.2 Metode Penelitian

3.2.1 Alur Penelitian

Dalam alur penelitian pada WeTV ini didasarkan dengan diagram dari alur penelitian yang dapat dilihat pada gambar 3.1. Pada alur ini berisikan beberapa proses pada penelitian ini.



Gambar 3. 1 Diagram Alur Penelitian

Pada gambar 3.1 merupakan diagram dari *flowchart* yang dapat memberikan penjelasan bahwa apa saja tahapan yang akan dilakukan dalam penelitian ini yaitu:

1. Pengumpulan Data

Dalam tahapan pertama ini akan dilakukan pengumpulan untuk *data review* dan ulasan dari aplikasi WeTV yang didapatkan dari Google Play Store dengan menggunakan teknik *data scraping*. Dalam pengambilan data ini menggunakan *tools* yaitu Jupyter Notebook yang dapat diakses pada Anaconda Navigator, dalam penelitian ini menggunakan data sebanyak 5001 data.

2. Data Preparation

Dalam Data Preparation ini akan menjadi 3 bagian yaitu terdapat Data Labeling, Data Cleaning dan Data Pre – Processing, Berikut dibawah ini merupakan penjelasan dari ketiga tahapan tersebut.

a. Data Labeling

Setelah mendapatkan data yang akan digunakan dalam penelitian akan masuk tahap dimana akan dilakukan pelabelan untuk setiap sentimen dari ulasan pengguna aplikasi WeTV. Dalam pelabelan ini akan menjadi patokan untuk melatih dari setiap algoritma yang digunakan dan dalam data labeling ini akan diberikan positif untuk setiap sentimen yang memiliki *score* ulasan 4 dan 5 sedangkan untuk negatif dari sentimen yang memiliki *score* ulasan 1, 2 dan 3. Dalam proses pelabelan ini akan dilakukan secara otomatis dengan jupyter notebook menggunakan *syntax* dimana acuan dalam pelabelan dari *score* yang diberikan oleh pengguna dari aplikasi WeTV.

b. Data Cleaning

Kemudian masuk kedalam tahap *data preparation* untuk mempersiapkan seperti *library* dan *sklearn* untuk setiap algoritma yang akan digunakan nantinya, setelah itu data – data ini akan dilakukan *cleaning* seperti huruf *capital* akan menjadi huruf kecil, penghapusan angka, penghapusan *emoticon* dan spasi yang kosong. Setelah data *cleaning* ini akan dilanjutkan ke tahapan *data pre – processing*.

c. Data Pre – Processing

Dalam proses ini data yang sudah di *cleaning* akan masuk dalam tahapan seperti *tokenize*, *stopwords* dan *stemming*. Sehingga data akan lebih rapih dan dapat mudah untuk dilakukan perhitungan setiap masing – masing algoritma yang digunakan dalam penelitian ini seperti perhitungan untuk TF – IDF, *accuracy*, *recall*, *precision* dan *F1 - Score*.

3. Modeling.

a. Split Data

Split Data ini sangat penting untuk menilai efektivitas model machine learning yang digunakan. Dengan teknik ini melibatkan pembagian dataset menjadi dua bagian, yaitu untuk pelatihan dan pengujian, sesuai dengan rasio yang ditetapkan. Pada data splitting untuk data training dan data validation ini akan melakukan perhitungan dari setiap algoritma yang digunakan dengan ukuran 70% untuk data training dan 30% untuk data validation. Selain 70% dan 30% dalam pembagian ini dihitung juga dengan ukuran 80% dan 20% [10]. Sehingga mencari dari ketiga algoritma ini yang lebih baik.

b. Permodelan Algoritma

Dalam tahapan ini merupakan tahapan dimana algoritma yang digunakan akan dilakukan perhitungan seperti *accuracy*, *Precision*, *Recall* dan *F1-Score*. Dalam perhitungan ini menggunakan Data yang sudah dilakukan Splitting dari tahapan sebelumnya.

4. Deployment

Dalam proses ini setelah dilakukan perhitungan dari permodelan ketiga algoritma, pada proses ini merupakan membuat *website* sederhana untuk dilakukan sebuah klasifikasi dari sentimen masyarakat terhadap aplikasi WeTV.

3.2.2 Metode Data Mining.

Pada metode *data mining* akan menggunakan metode CRISP – DM (*Cross-Industry Standard Process for Data Mining*). Dimana penelitian ini memiliki 6 tahapan yang akan digunakan yaitu:

1) Business Understanding

Tahapan pertama ini akan memiliki tujuan untuk mengetahui penelitian ini harus dilakukan, manfaat dari penelitian ini, dan berbagai masalah yang ingin diselesaikan, penelitian ini dilakukan memiliki tujuan untuk melihat perkembangan dari aplikasi WeTV mulai dari tahun 2021 hingga 2024 awal. Selain itu, hasil dari analisis sentimen ini dapat menjadi rekomendasi untuk calon pelanggan yang ingin menggunakan aplikasi *streaming* dari WeTV dan berlangganan.

2) Data Understanding

Dalam *data understanding* ini akan memahami data yang sudah diperoleh secara keseluruhan seperti pengumpulan data, eksplorasi data, pengecekan kualitas dari data dan mendeskripsikan data yang akan digunakan [41]. Dalam penelitian ini berfokus pada *review* dan ulasan dari pengguna WeTV di Google Playstore. Dan data ini akan didapatkan dengan menggunakan

data scraping yaitu *google-play-scraper*. Dalam penelitian ini, beberapa kolom data ulasan dari Google Play Store telah diidentifikasi, tetapi karena fokusnya adalah pada analisis sentimen, hanya kolom '*content*' yang berisi ulasan dan '*score*' yang merupakan rating pengguna terhadap aplikasi WeTV akan digunakan menjadi pelabelan dalam proses pengukuran dari *accuracy*, *Precision*, *Recall* dan *F1-Score*. Kolom '*content*' atau *review* akan digunakan untuk proses dari tahapan Data Preparation dan permodelan, Data ini untuk mendapatkan hasil perhitungan dari akurasi dari setiap algoritma yang digunakan.

3) Data Preparation

Dalam bagian ini merupakan persiapan data agar dapat diproses, dalam tahapan ini akan dilakukan *data cleaning* dengan menggunakan bahasa pemrograman Python, dalam *data cleaning* ini akan merubah huruf kapital menjadi huruf kecil, menghapus tanda baca, menghapus spasi yang berlebih dan menghapus angka. Adanya pelabelan dari angka *ratings* seperti *rating* 1 dan 2 akan dilabelkan negatif, *rating* 4 dan 5 akan dilabelkan positif. Setelah dilakukan *data cleaning* masuk ke bagian *data pre – processing* dimana data yang sudah rapi dan dilabelkan akan dilakukan seperti *tokenizing*, *stemming* dan *stop words*.

4) Data Modeling

Dalam tahapan ini akan melakukan pemisahan dataset menjadi dua bagian yaitu *training* dan *testing* yang akan menggunakan *TF – IDF* yang akan berfungsi untuk memberikan beban terhadap frekuensi dari kata – kata yang sudah terbagi. Berikut merupakan alur dari Data Splitting yaitu Data Split Training Setelah dilakukan pembagian atau Splitting Data ini akan masuk ke bagian permodelan dengan menggunakan metode Naïve Bayes, Decision Tree dan Random Forest untuk menentukan hasil dari data yang sudah dikelola.

5) Evaluation

Dalam tahapan ini berfungsi untuk melakukan evaluasi dari hasil dan performa dari model – model yang sudah terbentuk dari berbagai tahapan yang sudah di modeling. Dalam tahapan ini akan menggunakan *confusion matrix* agar mendapatkan hasil dari *accuracy*, *Precision*, *Recall* dan *F1 - Score* dari algoritma Naïve Bayes, Decision Tree dan Random Forest agar dapat dipertimbangkan dan hasilnya mendapatkan *precision*. Setelah Evaluation akan dilakukan masuk kedalam tahapan Deployment.

6) Deployment

Dalam tahapan terakhir ini akan dilakukan *deployment* pada website sederhana, dalam website ini menggunakan satu *model* yang terbaik dari ketiga algoritma yang sudah dihitung dan dibandingkan. Satu *model* tersebut akan melakukan perhitungan dari kalimat yang sudah dimasukkan oleh calon pengguna.

Dengan website sederhana para pengguna dan calon pengguna dapat melihat dan mencoba dari *model* yang sudah mendapatkan hasil untuk melakukan prediksi dari sentimen ulasan dan *review* apakah masuk dalam bagian positif atau negatif.

3.3 Teknik Pengumpulan Data

3.3.1 Populasi dan Sampel

Dalam populasi ini akan diambil untuk penelitian adalah *review* dan ulasan dari pengguna di Google Playstore pada pengguna aplikasi WeTV. Dalam hal ini menggunakan teknik dari Purposive Sampling. *Purposive sampling* merupakan sebuah teknik yang dilakukan untuk menentukan terlebih dahulu tujuan dari penelitian, sehingga data yang diperoleh dan dikelola akan menjadi lebih optimal dibandingkan dengan teknik lainnya [42]. Adanya sampel yang diambil dalam penelitian ini dari Google Playstore menggunakan *google-play-scraper* dan menggunakan bahasa Python dan *software* dari Jupyter Notebook.

3.3.2 Periode Pengambilan Data

Dalam periode pengambilan data dari ulasan pada aplikasi WeTV ini di Google Playstore berada rentang 1 januari 2021 hingga 31 januari 2024, sehingga dalam perhitungan dan analisis sentimen ini akan diharapkan lebih seimbang dan memberikan dampak dalam penggunaan aplikasi WeTV. Total terdapat 5001 data ulasan yang masih mentah dan sudah diambil dari Google Play Store.

3.4 Teknik Analisis Data

Dalam penelitian ini, memiliki sifat kualitatif dikarenakan data yang diambil dan dikelola merupakan ulasan dan *review* oleh pengguna. Dalam analisis data ini menggunakan Python dan Jupyter Notebook untuk mengambil data atau *data scraping*. Setelah mendapatkan data akan masuk ke dalam *data pre-processing* dan dilakukan TF-IDF dan analisis dari metode Naïve Bayes, Decision Tree dan Random Forest. Setelah itu penelitian ini akan dilanjutkan ke metode perhitungan akurasi dari ketiga algoritma dan dilanjutkan pada *confusion matrix*. Tahapan terakhir akan dilakukan perbandingan akurasi yang sudah didapatkan antar ketiga algoritma.

