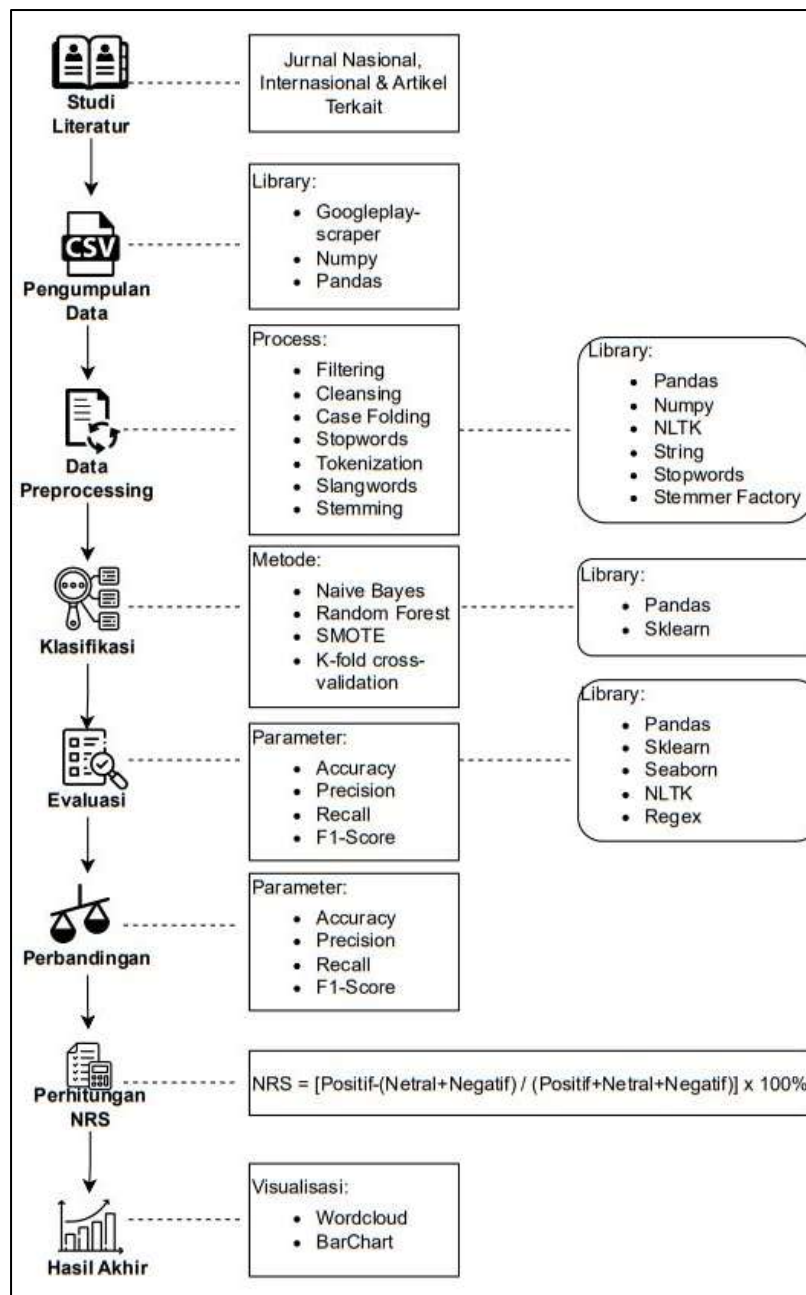


BAB III

METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Gambar 3.1 menyajikan rangkaian tahapan penelitian secara keseluruhan.



Gambar 3.1 Tahapan Penelitian

Tahapan-tahapan penelitian dijelaskan secara rinci sebagai berikut:

a. Studi Literatur

Studi literatur yang digunakan sebagai referensi dalam penelitian ini terdiri dari jurnal-jurnal ilmiah nasional dan internasional, serta sumber lain yang berkaitan dengan analisis sentimen, algoritma Naïve Bayes, dan algoritma Random Forest. Literatur-literatur ini akan menjadi panduan utama dalam memfasilitasi penelitian. Informasi lebih rinci mengenai literatur dapat ditemukan pada SUB BAB 2.

b. Pengumpulan Data

Data yang digunakan sebagai objek penelitian adalah dataset yang terdiri dari data *review* pengguna terhadap aplikasi layanan *provider* Live.On. Dataset ini diperoleh melalui penggunaan *library python* bernama *google-play-scraper* dan terdiri dari 8.000 *record* data.

c. *Data Preprocessing*

Data yang telah dikumpulkan diolah untuk dibersihkan agar terhindar dari gangguan, ketidaklengkapan, dan data yang tidak konsisten sebelum maju ke tahap selanjutnya. Untuk *preprocessing* data digunakan modul sastrawi *python* yang menyediakan berbagai fungsi dan algoritma untuk melakukan *preprocessing* pada teks dalam bahasa Indonesia. Berikut merupakan tahapan pada *data preprocessing* pada penelitian ini:

- *Filtering*, untuk proses klasifikasi diambil atribut ‘*content*’ dan ‘sentimen’.
- *Cleansing data*, pada tahap ini dilakukan penghapusan URL, tanda baca, (`(?|$.|.|@#%^/&*!=!_:")(-+,,)`), nomor (0-9), *white space* di awal dan akhir.

Contohnya “Sudahkah kamu melihat link <https://> pada pukul 10?” menjadi “Sudahkah kamu melihat link pada pukul”.

- *Case Folding*, merupakan pengubahan semua teks pada atribut ‘*content*’ menjadi huruf kecil untuk menghindari kebingungan antara kata yang sebenarnya tetapi ditulis dengan huruf besar atau kecil yang berbeda. Contohnya kata “Rumah” menjadi “rumah” yang sebenarnya merujuk pada hal yang sama.
- *Slangwords*, merupakan perubahan pada kata-kata atau bahasa informal menjadi bahasa formal atau standar. Contoh kata seperti “gue”, “lu” diubah menjadi “saya”, “kamu”. Perubahan kata-kata informal tersebut diterapkan menggunakan kamus *slangwords* yang tersedia.
- *Stopwords*, pada tahap ini dilakukan penghilangan kata-kata kurang berbobot dimana kata-kata tersebut disesuaikan dengan modul sastrawi dan *library nltk* (*natural language tool kit*) yang terdapat pada *python*. Contoh kata yang dihapus antara lain “yang”, “di”, “pada”, “dalam”, dan lain-lain.
- *Tokenization*, merupakan proses pemotongan kata per-kata dengan *library nltk* (*natural language tool kit*) yang terdapat pada *python* dengan algoritma *regex* (*regular expression*). Contoh pada kalimat “saya suka makan” menjadi “saya”, “suka”, “makan”.
- *Stemming*, pada tahap ini dilakukan penghilangan kata berimbuhan yang disesuaikan dengan modul sastrawi. Contoh kata seperti “memakan”, “makanan”, dan “makanan-makanan” diubah menjadi kata dasar “makan”.

d. Klasifikasi

Pada tahap klasifikasi, data yang telah melewati proses *preprocessing* dibagi menjadi *data testing* dan *data training* menggunakan metode *k-fold cross-validation*. Setelah itu, model algoritma klasifikasi Naïve Bayes dan Random Forest diterapkan untuk menghitung nilai rata-rata akurasi dari kedua model tersebut.

e. Evaluasi

Tahap evaluasi merupakan proses pengujian model untuk mengevaluasi performa model yang dihasilkan. Evaluasi dilakukan dengan menggunakan teknik *confusion matrix*, yang mencakup parameter evaluasi seperti, *precision*, *recall*, dan *f1-score*. Pemilihan *confusion matrix* pada tahap evaluasi karena memberikan informasi mengenai performa model klasifikasi, termasuk jumlah data yang diklasifikasikan dengan benar maupun salah, serta jenis kesalahan yang dibuat oleh model.

f. Perbandingan

Tahap perbandingan merupakan tahap membandingkan nilai *accuracy*, *precision*, *recall* dan *f1-score* antara Naïve Bayes dan Random Forest.

g. Perhitungan NRS

Tahap ini merupakan tahap perhitungan *Net Reputation Score* untuk mengukur reputasi perusahaan, dengan data hasil prediksi kedua model. Dengan perhitungan tersebut perusahaan dapat membuat kebijakan di masa depan dengan meningkatkan reputasi jika sentimen negatif terdeteksi, atau memanfaatkan

sentimen positif untuk memperkuat citra dan reputasi brand. Dimana semakin tinggi nilai NRS maka, semakin baik reputasi *brand* tersebut.

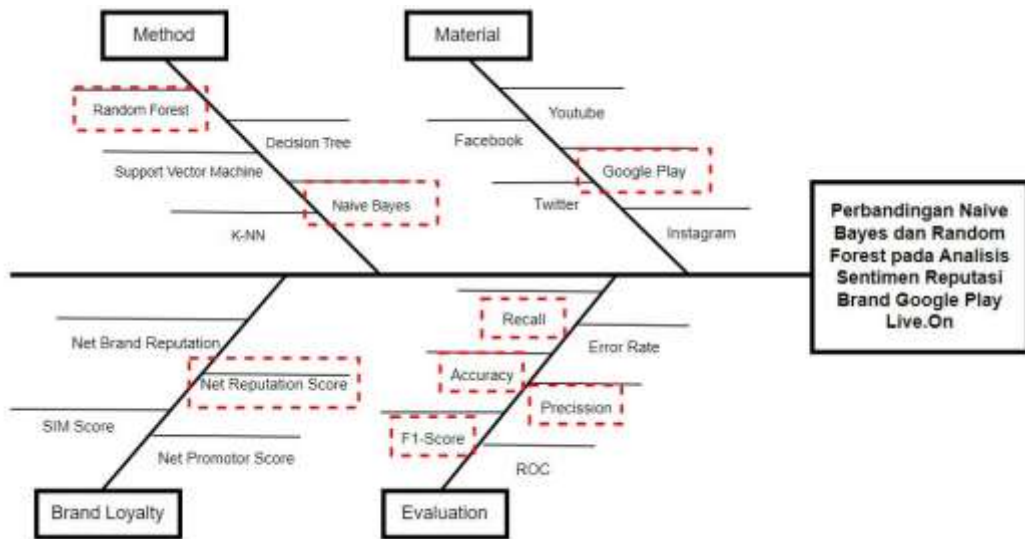
h. Hasil Akhir

Pada tahap ini dibuat hasil visualisasi dan interpretasi data dengan memanfaatkan beberapa *library* yaitu *pandas*, *matplotlib*, *seaborn*, dan *scikit-learn*

agar hasil dari analisis dapat ditangkap dengan mudah oleh pemangku kepentingan.

3.2 Fishbone Diagram

Gambar 3.2 adalah *fishbone diagram* dari penelitian ini yang menunjukkan bahwa *method*, *material*, *brand loyalty* dan *evaluation* metode dalam analisis sentimen dapat diselesaikan dengan banyak cara. Oleh karena itu, dalam penelitian ini metode yang digunakan ditandai dengan kotak merah pada *fishbone diagram* diantaranya yaitu Random Forest, Naïve Bayes, Google Play, *Net Reputation Score*, *Accurary*, *Recall*, *F1-Score*, dan *Precision*.



Gambar 3.2 *Fishbone Diagram*