

BAB I

PENDAHULUAN

1.1 Latar Belakang

Menurut laporan *We Are Social* disebutkan bahwa Indonesia termasuk salah satu negara dengan jumlah pengguna internet terbesar di dunia. Pada bulan Januari 2022, jumlah pengguna internet di Indonesia mencapai 204,7 juta orang. Tingkat rasio pengguna internet terhadap jumlah penduduk mencapai 73,7%, menunjukkan peningkatan penetrasi yang signifikan dari tahun ke tahun (Annur, 2022). Dari banyaknya pengguna internet di Indonesia, tidak sedikit dari pengguna yang mengeluh terhadap layanan yang diberikan oleh penyedia layanan. Banyak penyedia layanan internet yang menawarkan biaya relatif murah, akan tetapi pelanggan merasa kecewa karena terdapat beberapa masalah yang timbul (Jose, 2015). Keluhan yang dirasakan oleh pelanggan yaitu tidak memiliki kendali penuh untuk menikmati layanan data yang digunakan. Selain itu, terdapat batasan-batasan pemakaian data serta kuota yang terbagi-bagi. Untuk mengatasi masalah ini, Live.On hadir sebagai pendatang baru dalam penyediaan layanan digital ((kpl/gtr), 2020).

Live.On memanfaatkan *platform* Google Play untuk menerima umpan balik dari pelanggan dalam bentuk *rating* dan *review* dari aplikasi. Hal ini dapat membantu perusahaan memahami preferensi pelanggan dan meningkatkan layanan yang ditawarkan. Selain itu, dalam beberapa tahun terakhir metode komputasi untuk mempengaruhi opini, deteksi subjektif, dan analisis sentimen

telah menjadi minat baru (Speech, 2014). Dengan memanfaatkan teknologi dan *platform* yang tersedia, perusahaan dapat meningkatkan kualitas layanan dan memberikan pengalaman yang lebih baik bagi pelanggan (Septian et al., 2019). Pemanfaatan data *review* aplikasi dapat membantu perusahaan memahami sentimen konsumen terhadap merek. Oleh karena itu, pemanfaatan *Net Reputation Score* cukup penting untuk membangun citra positif dan meningkatkan loyalitas konsumen terhadap merek (*What Is Sentiment Analysis?*, 2023).

Berdasarkan penelitian telah dilakukan analisis sentimen aplikasi *provider* Live.On menggunakan metode Naïve Bayes dengan proses klasifikasi menggunakan metode *lexicon* yang mendapatkan nilai akurasi sebesar 87% dengan 1000 *record* data (Idli et al., 2022). Sedangkan, pada penelitian terdahulu mengenai analisis sentimen pengguna Indihome dengan data yang diambil dari Twitter selama 3 bulan, menunjukkan hasil bahwa metode Support Vector Machine memiliki nilai akurasi yang lebih akurat daripada Naïve Bayes. Dimana nilai akurasi SVM mencapai 84%, sedangkan Naïve Bayes yaitu sebesar 82% (Hashfi et al., 2022).

Terdapat penelitian lain, yang dilakukan untuk membandingkan metode Support Vector Machine, Random Forest, dan Decision Tree terhadap analisis sentimen *review* ponsel *Amazon Unlocked* dengan hasil bahwa metode Support Vector Machine memiliki nilai akurasi yang paling akurat yaitu sebesar 89% tidak jauh berbeda dengan Random Forest yaitu sebesar 88%, sedangkan Decision Tree 82%. Merek yang memiliki ulasan tertinggi pada penelitian ini adalah Samsung (Guia et al., 2019). Berdasarkan penelitian yang dilakukan oleh (Azizah et al.,

2015) terhadap analisis sentimen untuk mengetahui reputasi *brand provider* ponsel di Twitter menghasilkan bahwa Support Vector Machine memiliki nilai akurasi yang lebih baik daripada Naïve Bayes dan Decision Tree dengan reputasi merek tertinggi dimiliki oleh XL Axiata.

Beberapa penelitian sebelumnya telah menunjukkan adanya perbedaan hasil dalam penggunaan metode *machine learning* untuk analisis sentimen. Namun, belum ada perbandingan langsung antara metode Naïve Bayes dan Random Forest dengan menggunakan data dari Google Play. Algoritma Naïve Bayes dan Random Forest memiliki karakteristik yang berbeda. Naïve Bayes merupakan model probabilitas yang didasarkan pada asumsi bahwa semua fitur adalah independen satu sama lain, sedangkan Random Forest merupakan kumpulan pohon keputusan yang digunakan untuk menggabungkan hasil dari beberapa pohon. Selain itu, Naïve Bayes memiliki kecenderungan untuk *underfitting* karena asumsi sederhana yang dibuat dengan eksekusi yang cukup cepat, sedangkan Random Forest cenderung memiliki kemampuan dalam mengatasi *overfitting* karena diversitas pohon yang digunakan dalam ensemble dengan waktu eksekusi yang lebih lama (Kusrini & Luthfi, 2009). Berdasarkan hal tersebut, penelitian ini dilakukan untuk membandingkan kinerja kedua metode tersebut terhadap analisis sentimen reputasi *brand aplikasi provider* Live.On dengan menggunakan data dari Google Play. Setelah melakukan perbandingan antara kedua metode, dilakukan evaluasi menggunakan *confusion matrix*. Hasil akhirnya divisualisasikan dan diinterpretasikan menggunakan beberapa *library* di *Python*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, maka rumusan masalah pada penelitian ini adalah sebagai berikut:

- a. Bagaimana persentase sentimen pengguna *provider* Live.On dengan 3 kelas negatif, netral dan positif?
- b. Bagaimana hasil perbandingan performa antara algoritma Naïve Bayes dan Random Forest pada analisis sentimen *provider* Live.On dari komentar di Google Play?
- c. Bagaimana hasil perhitungan *Net Reputation Score* (NRS) dari *provider* Live.On berdasarkan performa metode Naïve Bayes dan Random Forest?

1.3 Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut:

- a. Membandingkan persentase sentimen pengguna *provider* Live.On dengan 3 kelas negatif, netral dan positif.
- b. Membandingkan performa algoritma Naïve Bayes dan Random Forest pada hasil analisis sentimen *provider* Live.On dari komentar di Google Play.
- c. Mengukur tingkat reputasi merek menggunakan perhitungan *Net Reputation Score* (NRS) *provider* Live.On berdasarkan performa metode Naïve Bayes dan Random Forest.

1.4 Batasan Masalah

Beberapa batasan masalah dari penelitian ini adalah sebagai berikut:

- a. Data yang diambil merupakan data ulasan bahasa Indonesia terbaru sebanyak 8.000 *record* data yang dihitung dari tanggal 15 Juli 2020 sampai dengan 24 Maret 2023 yang terdapat pada aplikasi Live.On di Google Play.
- b. Penelitian dilakukan dengan bahasa pemrograman *python* untuk tahap pengumpulan data, *data preprocessing*, klasifikasi, evaluasi dan visualisasi.
- c. Pelabelan sentimen dilakukan berdasarkan *rating* ulasan yang dilakukan menggunakan bahasa pemrograman *python* dan dikelompokkan menjadi 3 kelas, yaitu negatif, netral dan positif. *Rating* 1 dan 2 termasuk dalam kategori sentimen ‘Negatif’, *rating* 3 termasuk dalam kategori sentimen ‘Netral’, dan *rating* 4 dan 5 termasuk dalam kategori sentimen ‘Positif’.
- d. Tahapan *data preprocessing* yang dilakukan pada penelitian ini diantaranya: *filtering*, *cleansing*, *case folding*, *slangwords*, *stopword*, *tokenization*, dan *stemming*.
- e. Pembagian *data training* dan *data testing* menggunakan teknik *k-fold cross-validation* dengan nilai $k=10$.
- f. Untuk mengevaluasi performa algoritma digunakan metode *confusion matrix* dengan parameter kinerja yang terdiri dari nilai *precision*, *recall*, dan *f1-score* pada kedua algoritma. Sedangkan, untuk mengetahui tingkat kepuasan pengguna ditinjau dari hasil perhitungan nilai *Net Reputation Score* (NRS).
- g. Hasi akhirnya dataset tersebut akan divisualisasi dan diinterpretasikan menggunakan beberapa *library* di *python*.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

- a. Memperoleh informasi mengenai sentimen pengguna *provider* Live.On apakah mengarah ke sentimen positif, negatif atau netral.
- b. Memperoleh informasi mengenai perbandingan performa antara algoritma Naïve Bayes dan Random Forest pada data *review provider* Live.On di Google Play.
- c. Memperoleh informasi mengenai nilai reputasi *provider* Live.On di Google Play berdasarkan performa Naïve Bayes dan Random Forest, dimana nilai reputasi tersebut dapat dijadikan sebagai panduan untuk pelaku bisnis dalam mengambil keputusan terhadap kebijakan yang diterapkan pada *provider* Live.On.

1.6 Metodologi Penelitian

Metode penelitian yang dilakukan pada penelitian ini memiliki tahapan sebagai berikut:

- a. Studi Literatur

Melakukan tinjauan pustaka dari jurnal-jurnal atau sumber lainnya yang relevan dengan topik penelitian yang akan dilakukan.

- b. Pengumpulan Data

Proses pengumpulan data untuk kebutuhan penelitian berupa dataset.

- c. *Data Preprocessing*

Pembersihan terhadap data sebelum diolah ke proses selanjutnya yang bertujuan agar data terhindar dari gangguan, ketidaklengkapan, dan data yang tidak konsisten.

d. Klasifikasi

Proses pemodelan untuk membandingkan nilai akurasi dari metode Naïve Bayes dan Random Forest.

e. Evaluasi

Dilakukan evaluasi terhadap hasil performa model yang dipilih pada tahap klasifikasi.

f. Perbandingan

Membandingkan hasil nilai parameter kinerja yaitu *accuracy*, *precision*, *recall*, dan *f1-score* antara Naïve Bayes dan Random Forest.

g. Perhitungan NRS

Dilakukan perhitungan untuk melihat reputasi perusahaan di mata publik menggunakan perhitungan *Net Reputation Score* dengan data hasil prediksi kedua model.

h. Hasil Akhir

Pembuatan visualisasi dan interpretasi data menggunakan *library* di *python*.

1.7 Sistematika Penulisan

Sistematika penulisan yang digunakan penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bagian ini menjelaskan latar belakang dan dasar penelitian, rumusan permasalahan yang akan diteliti, tujuan penelitian, manfaat dari penelitian, metodologi yang digunakan dalam penelitian, serta sistematika penulisan untuk melaporkan hasil penelitian.

BAB II LANDASAN TEORI

Bagian ini memuat informasi mengenai studi-studi terkait yang relevan dengan topik penelitian yang akan dilakukan. Selain itu, bagian ini juga berisi penjelasan tentang teori-teori yang akan digunakan dalam penelitian.

BAB III METODOLOGI PENELITIAN

Bagian ini memuat informasi terkait dengan metodologi atau langkah-langkah penelitian yang akan dilakukan.

BAB IV HASIL DAN PEMBAHASAN

Bagian ini mencakup eksperimen penelitian yang dilakukan dan pembahasan hasil eksperimen. Selain itu, hasil eksperimen juga akan dibandingkan dengan eksperimen sebelumnya yang relevan dengan topik penelitian.

BAB V SIMPULAN DAN SARAN

Bagian ini berisi ringkasan kesimpulan dari hasil eksperimen yang telah dilakukan. Selanjutnya, bagian ini juga menyajikan saran-saran untuk penelitian selanjutnya yang akan membahas topik sejenis atau terkait.