

BAB I

PENDAHULUAN

1.1. Latar Belakang

Ulasan yang diberikan pengguna di platform *Google Play Store* memainkan peran penting dalam mempengaruhi opini dan penilaian terhadap aplikasi di era digital yang terus berkembang. Sama halnya dengan aplikasi My Pertamina, aplikasi yang dikembangkan oleh PT Pertamina yang memberikan layanan pada pengguna berupa akses informasi barang dan jasa Pertamina. Untuk mendapatkan pemahaman mengenai sentimen pengguna terhadap aplikasi My Pertamina, analisis sentimen terhadap ulasan pengguna menjadi suatu hal yang penting. Analisis sentimen dari ulasan pengguna memiliki peranan penting untuk memahami sentimen dan pemikiran pengguna terhadap aplikasi MyPertamina.

Analisis sentimen adalah proses memahami dan memproses data teks secara otomatis untuk memperoleh informasi. Analisis sentimen dilakukan untuk mengidentifikasi opini tentang topik dan objek (orang, organisasi, produk, dll.) dalam kumpulan data. (Hasri & Alita, 2022). Algoritma yang dapat membantu dalam proses klasifikasi untuk analisis sentimen, antara lain *Naïve Bayes Classifier*, *Random Forest*, *K-Nearest Neighbor*, *Support Vector Machine*, dan masih banyak yang lainnya.

Algoritma *Naïve Bayes* merupakan algoritma klasifikasi sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan dan menggabungkan nilai dari kumpulan data yang diberikan (Hayuningtyas, 2019). Walaupun klasifikasi Algoritma *Naïve Bayes* dapat dikatakan sebagai algoritma klasifikasi yang sederhana, tetapi hasil yang didapat dari klasifikasi algoritma *Naïve Bayes* ini sering sekali mencapai performa yang serupa dengan algoritma klasifikasi lainnya seperti *Neural Network classifier & Decision tree*. (Indrayuni, 2019)

Salah satu penelitian yang menggunakan algoritma *Naïve Bayes* ialah penelitian Bintang Zulfikar Ramadhan dan rekan-rekannya yang melakukan

sebuah penelitian terhadap analisis sentimen ulasan aplikasi *e-commerce* (Ramadhan et al., 2022). Penelitian ini melakukan *crawling* data di 3 aplikasi *e-commerce* berbeda yaitu Shopee, Tokopedia, dan Lazada dengan masing-masing aplikasi menggunakan 500 data ulasan. Pembagian klasifikasi pada penelitian ini terbagi menjadi 2 yaitu ulasan positif dan ulasan negatif yang dilabeli secara manual. Pembagian data training dan testing terbagi menjadi 3 model yaitu 80:20, 70:30 dan 60:40. Hasil evaluasi algoritma *Naïve Bayes* menghasilkan hasil terbaik pada skenario 1 pada aplikasi Shopee (80% data training dan 20% data testing) yang menghasilkan *accuracy* sebesar 92%. Pada aplikasi Tokopedia menghasilkan hasil terbaik pada skenario 3 (60% data training dan 3401% data testing) menghasilkan *accuracy* 83,5%. Pada aplikasi Lazada menghasilkan hasil terbaik pada skenario pada 3 (60% data training dan 40% data testing) menghasilkan *accuracy* 79,5%.

Selain menggunakan *Naïve Bayes*, pada penelitian ini juga menggunakan algoritma kedua yang digunakan sebagai pembandingan yaitu algoritma *K-Nearest Neighbor*. Algoritma *K-Nearest Neighbor* adalah metode klasifikasi objek berdasarkan data pelatihan yang paling dekat dengannya (Tangguh Admojo, 2020). KNN juga merupakan metode algoritma *supervised learning*, di mana kelas yang paling banyak muncul (mayoritas) yang akan menjadi kelas hasil klasifikasi (Nurjanah et al., 2020).

Algoritma ini digunakan oleh Aluisius Dwiki Adhi Putra dan Safitri Juanita dalam sebuah penelitian analisis sentimen terhadap dua aplikasi investasi yaitu Bibit dan Bareksa dengan *K-Nearest Neighbor* sebagai algoritma klasifikasi. Penelitian ini menggunakan aplikasi *Rapid Miner* serta menggunakan data ulasan aplikasi Bibit dan Bareksa yang di-*crawling* pada 13 Februari 2021 hingga 20 April 2021. Klasifikasi sentimen pada penelitian ini terbagi menjadi sentimen positif dan sentimen negatif. Setelah itu dilakukan modelling menggunakan *split* data menjadi 3 model *training* dan *testing* yaitu 60:40, 70:30 dan 80:20. Hasil akurasi terbaik didapat dengan algoritma KNN

dengan model *training* dan *testing* 60:40, dengan hasil akurasi 85,14% untuk aplikasi Bibit dan 81,70% untuk aplikasi Bareksa (Adhi Putra & Juanita, 2021).

Adapun penelitian yang melakukan komparasi terhadap algoritma *Naïve Bayes* dan *K-Nearest Neighbor* dilakukan oleh Muhammad Syarifudin dengan menganalisis sentimen opini publik terhadap COVID-19 di Twitter. Penelitian yang dilakukan Syarifudin menggunakan *tweet* yang diambil dari API Twitter sebanyak 1098 opini dan dilakukan klasifikasi *tweet* dengan sentimen positif dan sentimen negatif. Didapatkan hasil sebanyak 610 *tweet* bersentimen positif dan 488 *tweet* bersentimen negatif. Nilai akurasi yang didapatkan dengan menggunakan metode klasifikasi *Naïve Bayes* sebesar 63.21% sedangkan metode klasifikasi *K-Nearest Neighbor* sebesar 58.10% (Syarifuddin, 2020).

Penelitian yang melakukan komparasi model algoritma juga dilakukan oleh Abdul Rozaq dan rekan-rekannya dengan judul “Analisis Sentimen Terhadap Implementasi Program Merdeka Belajar Kampus Merdeka Menggunakan *Naïve Bayes*, *K-Nearest Neighbors* Dan *Decision Tree*”. Dataset *tweet* di-*crawling* pada 1 Oktober 2021 hingga 18 Desember 2021 sebanyak 475 data kemudian dilabeli menjadi 3 yaitu positif, negatif, dan netral. Pembagian dataset untuk *training* dan *testing* dengan komposisi 80:20. Hasil klasifikasi komentar terkait implementasi program merdeka belajar kampus merdeka menunjukkan hasil bahwa akurasi sebesar 99,22% untuk *Naïve Bayes*, *K-Nearest Neighbors* sebesar 96,90% dan *Decision tree* sebesar 37,21% (Rozaq et al., 2022).

Pemilihan algoritma *Naïve Bayes* dan *K-Nearest Neighbor* pada penelitian ini karena merupakan algoritma yang sering digunakan dalam melakukan analisis sentimen. *Naïve Bayes* merupakan algoritma klasifikasi sederhana yang menghitung serangkaian probabilitas dengan menjumlahkan dan menggabungkan nilai dari kumpulan data yang diberikan. *Naïve Bayes* juga merupakan algoritma yang sederhana, cepat, serta juga memiliki akurasi yang relatif tinggi. Sementara itu, algoritma *K-Nearest Neighbor* merupakan algoritma adalah algoritma klasifikasi di mana kelas yang paling banyak muncul

(mayoritas) yang akan menjadi kelas hasil klasifikasi. *K-Nearest Neighbor* dipilih sebagai metode pembandingan karena merupakan metode yang umum digunakan, mudah diimplementasikan, dan memiliki sedikit *hyperparameter*.

Dalam penelitian ini, digunakan bahasa pemrograman Python dengan menerapkan algoritma *Naïve Bayes* dan *K-Nearest Neighbor* untuk menganalisis sentimen pada ulasan aplikasi MyPertamina. Ulasan pengguna diklasifikasikan ke dalam tiga kategori, yaitu positif, netral, dan negatif. Hasil analisis sentimen pada aplikasi MyPertamina diharapkan dapat memberikan masukan yang berharga untuk pengembang aplikasi dalam melakukan perbaikan fitur-fitur pada sentimen negatif, meningkatkan pengalaman pengguna, dan bermanfaat pada pengembangan aplikasi MyPertamina di masa yang akan datang.

Berdasarkan latar belakang diatas, maka tugas akhir ini akan mengimplementasikan algoritma *Naïve Bayes* dan *K-Nearest Neighbor* yang berjudul “Perbandingan Metode Analisis Sentimen Ulasan Aplikasi MyPertamina dengan Menggunakan Algoritma *Naïve Bayes* dan *K-Nearest Neighbor*”.

1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, maka dapat dibentuk beberapa poin rumusan masalah yaitu:

1. Bagaimana gambaran umum yang didapatkan pada ulasan aplikasi MyPertamina di Google Play Store?
2. Bagaimana hasil komparasi dari penggunaan algoritma *Naïve Bayes* dan algoritma *K-Nearest Neighbor* dalam menganalisis sentimen ulasan-ulasan aplikasi MyPertamina di *Google Play Store*?

1.3. Tujuan

Tujuan yang dilakukan oleh peneliti adalah untuk menjawab rumusan masalah yang telah dikemukakan sehingga tujuan dalam penelitian ini adalah sebagai berikut :

1. Mengetahui gambaran umum yang didapatkan pada ulasan aplikasi MyPertamina di *Google Play Store*.
2. Mengetahui hasil komparasi setelah pengujian algoritma Naïve Bayes dan *K-Nearest Neighbor* dalam menganalisis sentimen ulasan-ulasan aplikasi MyPertamina di *Google Play Store*.

1.4. Manfaat

Manfaat yang diharapkan pada penelitian ini ialah sebagai berikut :

1. Mengetahui gambaran umum yang didapatkan pada ulasan aplikasi MyPertamina di *Google Play Store*.
2. Hasil dari penelitian ini diharapkan dapat memudahkan pengembang aplikasi MyPertamina untuk menjadikan acuan dalam melakukan pemeliharaan kualitas dan melakukan evaluasi ke arah yang lebih baik dengan opini positif, netral, dan negatif yang telah diberikan pengguna.

1.5. Batasan Masalah

Berdasarkan rumusan masalah yang ada, maka permasalahan akan dibatasi oleh peneliti dengan batasan masalah sebagai berikut :

1. Metode yang digunakan untuk analisis sentimen ialah *Naïve Bayes* dan *K-Nearest Neighbor*.
2. Penelitian ini menggunakan 1500 buah data ulasan aplikasi MyPertamina di *Google Play Store*.
3. Data ulasan aplikasi MyPertamina yang digunakan pada penelitian ini menggunakan ulasan-ulasan berbahasa Indonesia.
4. Data ulasan aplikasi MyPertamina yang digunakan dipilih berdasarkan urutan data yang paling relevan.