

## BAB I PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan teknologi digital telah mendorong peningkatan penggunaan aplikasi kesehatan mental berbasis daring, seperti Riliv, yang menyediakan layanan konseling, meditasi, dan edukasi psikologis. Ulasan dari pengguna aplikasi-aplikasi tersebut merupakan sumber data penting yang dapat dianalisis untuk mengevaluasi kualitas layanan serta memahami persepsi publik. Riliv memiliki jumlah reviews dengan 3035 data ulasan pengguna dianalisis guna mendapatkan gambaran yang lebih komprehensif terkait tingkat kepuasan dan ketidakpuasan pengguna terhadap layanan tersebut. Analisis sentiment adalah metode efektif untuk mengekstraksi opini dari ulasan dan mengelompokkannya ke dalam kategori sentiment positif, negative, atau netral. Pendekatan ini sangat berguna dalam konteks aplikasi seperti Riliv untuk secara otomatis dan masif menangkap tingkat kepuasan atau ketidakpuasan pengguna.

Dalam penerapannya, algoritma *machine learning* seperti Naive Bayes dan Random Forest sering digunakan untuk analisis sentimen karena kemampuannya yang efektif dalam memproses data teks. Naive Bayes dikenal dengan kesederhanaan dan keandalannya dalam melakukan klasifikasi berbasis probabilitas, sedangkan Random Forest memiliki keunggulan dalam mengatasi masalah *overfitting* dan mampu menangani variabel yang kompleks dengan baik. Namun, kinerja kedua algoritma ini sangat dipengaruhi oleh pengaturan parameter yang digunakan. Oleh karena itu, optimasi *hyperparameter* menjadi langkah penting untuk meningkatkan akurasi dan efisiensi model dalam mengklasifikasikan sentimen dari ulasan pengguna [1].

Berbagai teknik optimasi *hyperparameter* seperti Grid Search, Random Search, dan Bayesian Optimization telah terbukti dapat meningkatkan performa algoritma klasifikasi di berbagai bidang analisis sentimen, termasuk pada teks berbahasa Indonesia [2]. Salah satu metode terbaru yang semakin banyak

digunakan adalah Optuna, sebuah kerangka optimasi *hyperparameter* yang menggabungkan fleksibilitas dan efisiensi melalui algoritma *Tree-structured Parzen Estimator* (TPE). Optuna secara otomatis mengeksplorasi ruang parameter dengan menerapkan strategi pemangkasan berdasarkan hasil eksperimen awal, sehingga secara signifikan memangkas waktu komputasi tanpa mengorbankan akurasi model. Berbagai penelitian telah menunjukkan bahwa Optuna mampu menghasilkan performa yang kompetitif, bahkan melampaui metode Bayesian Optimization konvensional dengan jumlah iterasi yang lebih sedikit [3].

Dengan menerapkan *hyperparameter* Optuna dapat mempertimbangkan pentingnya akurasi dalam klasifikasi sentimen aplikasi kesehatan untuk melihat perbandingan masing-masing algoritma dan metode optimasi, dalam penelitian ini bertujuan melakukan optimasi *hyperparameter* pada algoritma Naive Bayes dan Random Forest dalam menganalisis sentimen pengguna aplikasi Riliv. Peneliti mengharapkan dari penelitian ini dapat menjadikan acuan dalam pemilihan optimasi terbaik.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang penulis diatas, rumusan masalah dalam penelitian ini adalah sejauh mana penerapan teknik, TF-IDF, SMOTE dan *Hyperparameter Tuning* dengan Optuna dapat meningkatkan performa model klasifikasi *Random Forest* dan *Naive Bayes*.

## **1.3 Batasan Masalah**

Untuk menjaga fokus dan ruang lingkup penelitian tetap terarah, ditetapkan beberapa batasan masalah sebagai berikut:

1. Data yang diambil untuk penelitian ini hanya berasal dari ulasan pengguna aplikasi Riliv di Google Play Store.
2. Pengambilan data dilakukan dengan metode *Web Scraping* menggunakan Google Play Scraper API.
3. Sentimen yang dianalisis 3 kategori, yaitu positif, negatif, netral.

#### 1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk membandingkan efektivitas metode klasifikasi Random Forest dan Naive Bayes dalam menganalisis sentimen ulasan pengguna aplikasi Riliv di Google Play Store yang berkaitan dengan layanan kesehatan mental. Selain itu, penelitian ini juga bertujuan untuk mengkaji sejauh mana penerapan teknik *preprocessing* teks, ekstraksi fitur menggunakan TF-IDF, penanganan ketidakseimbangan data dengan SMOTE, serta optimasi *hyperparameter* menggunakan Optuna dapat meningkatkan performa model klasifikasi sentimen tersebut.

#### 1.5 Manfaat Penelitian

Berikut manfaat penelitian baik dari segi teoritis maupun praktis:

##### a. Manfaat Teoritis

1. Memberikan kontribusi dalam pengembangan ilmu tentang analisis sentimen.
2. Menambah pengetahuan mengenai penerapan algoritma Random Forest dan Naive Bayes untuk klasifikasi teks ulasan pada aplikasi kesehatan mental.
3. Menambah literatur terkait efektivitas teknik *preprocessing* teks dalam klasifikasi sentimen.
4. Mengkaji penggunaan TF-IDF sebagai metode ekstraksi fitur dalam analisis sentimen.
5. Menjelaskan penanganan ketidakseimbangan data menggunakan SMOTE untuk meningkatkan performa model.
6. Mengoptimasi *hyperparameter* model klasifikasi dengan menggunakan Optuna untuk meningkatkan akurasi dan kinerja.

##### b. Manfaat Praktis

1. Menjadi referensi bagi pengembang aplikasi dan praktisi *data science* dalam memilih metode klasifikasi sentimen yang paling tepat dan akurat.

2. Membantu pengembang aplikasi Riliv dalam meningkatkan kualitas pelayanan melalui pemahaman lebih baik terhadap opini dan kebutuhan pengguna.
3. Mendukung pengambilan keputusan berbasis data yang lebih efektif dalam pengembangan dan pengelolaan aplikasi kesehatan mental.
4. Mendorong penerapan teknik-teknik optimasi yang dapat meningkatkan performa model klasifikasi sentimen di dunia nyata.

## **1.6 Sistematika Penulisan**

Sistematika penulisan dari penelitian ini adalah:

### **BAB I PENDAHULUAN**

Bab ini berisi latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penelitian.

### **BAB II TINJAUAN PUSTAKA**

Bab ini berisi tinjauan pustaka, studi literatur, dan dasar-dasar teori yang digunakan termasuk dasar-dasar teori yang digunakan penelitian-penelitian sebelumnya yang relevan, serta penjelasan tentang metode *machine learning* yang diterapkan dalam penelitian ini.

### **BAB III METODE PENELITIAN**

Bab ini menjelaskan metode penelitian yang digunakan dalam proses penelitian ini. Terdapat uraian tentang objek penelitian, teknik pengumpulan data, proses *preprocessing* data, serta tahap-tahap pemodelan dan evaluasi model.

### **BAB IV HASIL DAN PEMBAHASAN**

Bab ini berisi terkait hasil penelitian, termasuk hasil analisis data, evaluasi performa model menggunakan metrik seperti akurasi, presisi, *recall*, dan *F1-score*.

### **BAB V PENUTUP**

Bab ini berisi kesimpulan dari penelitian yang telah dilakukan, termasuk jawaban atas rumusan masalah yang diajukan. Selain itu, bab ini memberikan saran saran yang dapat menjadikan rekomendasi untuk penelitian lebih lanjut.