

BAB I

PENDAHULUAN

1.1. Latar Belakang

Aplikasi SIAP UNDIP Mahasiswa merupakan sistem informasi akademik berbasis Android yang memudahkan mahasiswa dalam mengakses layanan akademik secara digital. Namun, aplikasi ini hanya tersedia untuk Android versi lama (4.4-10) dan tidak lagi diperbarui sejak Juli 2020. Ulasan pengguna yang diambil dari *Google Play Store* menunjukkan distribusi sentimen yang tidak seimbang, dimana ulasan negatif jauh lebih dominan daripada ulasan positif. Ketidakseimbangan ini menyulitkan analisis sentimen karena model cenderung bias terhadap kelas mayoritas dan kurang mampu mendeteksi keluhan atau masalah spesifik pengguna (Choudhary dan Shukla, 2022). Kondisi ini menegaskan pentingnya pengembangan metode analisis sentimen yang mampu mengatasi ketidakseimbangan data dan tetap menghasilkan prediksi yang akurat.

Review atau ulasan dari pengguna memiliki peran penting dalam meningkatkan kualitas dan efektivitas sebuah aplikasi. Ulasan ini memberikan masukan langsung mengenai kelebihan dan kekurangan sistem dari sudut pandang pengguna akhir, yang sangat bermanfaat untuk perbaikan dan pengembangan aplikasi lebih lanjut (Nufus, 2022). Selain disertai komentar, ulasan juga dilengkapi dengan *rating* bintang satu hingga lima yang menjadi indikator kualitas aplikasi. Oleh karena itu, analisis sentimen menjadi teknik yang penting untuk mengolah data ini dengan cara mengklasifikasikan opini pengguna ke dalam dua kategori utama, yaitu positif dan negatif (Stephenie dkk., 2020). Teknik ini membantu dalam mengidentifikasi aspek-aspek yang perlu diperbaiki dan mengungkap adanya inkonsistensi antara rating dan isi review yang mungkin terjadi.

Berbagai algoritma *machine learning* telah diterapkan dalam analisis sentimen, salah satunya adalah *Boosting Weighted Extreme Learning Machine* (BWELM). BWELM merupakan penggabungan antara *Weighted Extreme Learning Machine* (WELM) dan metode *ensemble AdaBoost* untuk meningkatkan akurasi klasifikasi, terutama pada data yang tidak seimbang (Choudhary dan Shukla, 2022). BWELM memanfaatkan kecepatan ELM dalam pelatihan serta kekuatan *boosting*

dalam mengenali pola tekstual yang kompleks pada ulasan pengguna (Gnanakumaran dkk., 2023). Akan tetapi, BWELM memiliki kelemahan, seperti sensitivitas terhadap dimensi data yang tinggi dan keterbatasan dalam proses paralelisasi karena penggunaan *pseudoinvers* (Rathod dan Wankhade, 2021).

Di sisi lain, *Random Forest* merupakan metode *ensemble* yang membangun banyak pohon keputusan secara paralel dengan pendekatan *bagging*. Model ini memberikan prediksi akhir berdasarkan voting mayoritas dan dikenal lebih stabil terhadap variasi data (Talekar dan Agrawal, 2020). *Random Forest* juga efektif dalam menangani fitur teks berdimensi tinggi dan mampu memberikan interpretasi melalui informasi *feature importance* (Özmen dan Gündüz, 2025).

Model *Hybrid* berbasis *machine learning* telah banyak digunakan untuk mengatasi berbagai kelemahan pada metode tunggal, seperti CNN-SVM, BERT-BiLSTM-CNN, serta SVM-RNN dalam analisis sentimen (Li dkk., 2024; Alawi dan Bozkurt, 2024; Srivastava dkk., 2023). Penggunaan *Boosting Weighted Extreme Learning Machine* (BWELM) telah terbukti efektif dalam menangani distribusi kelas yang tidak seimbang, sebagaimana ditunjukkan oleh Li dkk. (2014), Liu dkk. (2017), Ganesan (2024), dan Cheng dkk. (2020). Namun, beberapa studi menunjukkan bahwa BWELM masih memiliki keterbatasan dalam menghadapi data berdimensi tinggi dan distribusi fitur yang kompleks karena ketergantungannya pada bobot empiris yang bersifat tetap. Sementara itu, pendekatan *boosting* seperti *AdaBoost* telah dieksplorasi dalam studi seperti Feng (2019) dan Prabhakar dkk. (2019). Meskipun meningkatkan performa pada beberapa kasus, teknik ini cenderung kurang stabil pada data berukuran besar dan berisiko mengalami *overfitting*, terutama ketika digunakan pada data dengan *noise* tinggi. Sebaliknya, algoritma *Random Forest* telah menunjukkan performa yang kuat dan stabil pada tugas klasifikasi sentimen, khususnya dalam menangani data teks berdimensi tinggi dan *noisy*. Studi oleh Karthika dkk. (2019), Padhy dkk. (2024), dan Natras dkk. (2022) menunjukkan bahwa *Random Forest* unggul dalam menjaga kestabilan prediksi, toleransi terhadap variasi data, serta mampu memberikan interpretasi melalui fitur penting (*feature importance*). Keandalan *Random Forest* dalam berbagai konteks klasifikasi menjadikannya kandidat ideal

untuk dikombinasikan dalam model *Hybrid* guna mengimbangi kekurangan metode lain seperti BWELM.

Hasil studi sebelumnya menunjukkan bahwa BWELM efektif dalam menangani ketidakseimbangan data, tetapi kurang optimal pada data berdimensi tinggi dan rentan terhadap ketidakstabilan. Sebaliknya, *Random Forest* menunjukkan performa stabil dan kuat terhadap *overfitting*. Oleh karena itu, pendekatan *Hybrid* menjadi solusi yang relevan dengan menggabungkan kekuatan dari masing-masing metode. Contohnya, van de Kamp (2022) berhasil menggabungkan XGBoost dan *Random Forest* untuk meningkatkan akurasi prediksi GDP melalui pembagian peran seleksi fitur dan klasifikasi. Pendekatan serupa dapat diterapkan dalam konteks analisis sentimen terhadap aplikasi SIAP UNDIP Mahasiswa.

Penelitian ini diusulkan untuk menerapkan pendekatan model *Hybrid* berbasis *stacked generalization* dengan menggabungkan BWELM sebagai *base model* dan *Random Forest* sebagai *meta-classifier*. Dalam sistem ini, BWELM menghasilkan prediksi probabilistik terhadap sentimen yang kemudian digunakan sebagai fitur *input* bagi *Random Forest*. *Random Forest* selanjutnya mempelajari pola distribusi probabilitas tersebut untuk menghasilkan prediksi akhir. Pendekatan ini bertujuan untuk mengatasi kelemahan BWELM dalam menangani data ambigu, serta meningkatkan akurasi dan stabilitas model klasifikasi sentimen terhadap ulasan aplikasi SIAP UNDIP Mahasiswa dalam konteks akademik di Universitas Diponegoro.

1.2. Tujuan Penelitian

Penelitian bertujuan untuk menerapkan model *Boosting Weighted Extreme Learning Machine* (BWELM) dan *Random Forest* dalam bentuk *Hybrid* guna meningkatkan performa klasifikasi sentimen pada ulasan aplikasi SIAP UNDIP Mahasiswa di *Google Play Store*. Tujuan khusus dari penelitian ini dirumuskan sebagai berikut:

1. Meningkatkan akurasi klasifikasi sentimen ulasan pengguna terhadap aplikasi SIAP UNDIP Mahasiswa melalui penerapan model *Hybrid* BWELM dan

Random Forest, serta mengevaluasi kinerja model menggunakan berbagai metrik seperti akurasi, *precision*, *recall*, *F1-score*, dan ROC AUC.

2. Mengatasi ketidakseimbangan kelas, serta dimensi fitur teks yang tinggi pada dataset ulasan dengan memanfaatkan kekuatan kombinasi BWELM sebagai *base-learner* dan *Random Forest* sebagai *meta-classifier*.

1.3. Manfaat Penelitian

Penelitian diharapkan dapat memberikan manfaat sebagai berikut:

1. Manfaat Praktis

Memberikan wawasan kepada pihak Universitas Diponegoro mengenai persepsi dan kepuasan mahasiswa terhadap layanan aplikasi SIAP UNDIP Mahasiswa melalui hasil analisis sentimen ulasan pengguna. Informasi tersebut dapat digunakan sebagai bahan evaluasi dan dasar pengambilan keputusan dalam upaya peningkatan kualitas layanan sistem informasi akademik.

2. Manfaat Akademik

Memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang analisis sentimen dan *machine learning*, khususnya dalam penerapan model *Hybrid BWELM* dan *Random Forest*. Penelitian ini menunjukkan bagaimana kombinasi dua model dengan karakteristik berbeda dapat mengatasi permasalahan klasifikasi pada data teks, seperti ketidakseimbangan kelas dan dimensi fitur yang tinggi, serta meningkatkan akurasi dan stabilitas model klasifikasi sentimen.

SEKOLAH PASCASARJANA