

BAB I

PENDAHULUAN

1.1 Latar Belakang

Gangguan kesehatan mental, khususnya depresi, telah menjadi perhatian global yang signifikan, mempengaruhi jutaan individu dan berdampak pada kualitas hidup mereka (Fikry & Inoue, 2023).

Dalam implementasi *Random Forest*, teknik pembagian data (*splitting data*) memegang peranan penting dalam membangun pohon keputusan yang efektif. Pemilihan kriteria pemisahan yang tepat dapat meningkatkan akurasi model dan efisiensi komputasi. Beberapa penelitian sebelumnya telah mengeksplorasi pengaruh teknik pembagian data terhadap kinerja model. Sebagai contoh, menganalisis kriteria pemisahan pada *Random Forest* untuk klasifikasi evaluasi kendaraan, menunjukkan bahwa pemilihan kriteria pemisahan mempengaruhi nilai akurasi model (Nugroho, 2022).

Menunjukkan bahwa dalam sistem pengambilan keputusan berbasis data *real-time*, seperti deteksi banjir menggunakan sensor ultrasonik, pemilihan metode pemrosesan dan validasi sistem sangat menentukan akurasi dan respons sistem. Pendekatan ini sejalan dengan pentingnya teknik *splitting data* yang tepat untuk meningkatkan efisiensi model dalam pengolahan sinyal visual seperti gerak wajah dan kepala (Pohan et al., 2024).

Selain itu, membandingkan teknik pembagian *dataset* untuk klasifikasi wajah pada citra CCTV, menemukan bahwa teknik *k-Fold Cross Validation* memberikan performa yang lebih stabil dan akurasi tinggi dibandingkan teknik lainnya (Nurhopipah & Hasanah, 2020). Penelitian lain mengusulkan metode pengenalan gerak tubuh bagian atas berdasarkan *frame* kunci dan regresi *Random Forest*, menunjukkan efektivitas pendekatan tersebut dalam pengenalan gerak (Li et al., 2020).

Mengembangkan metode *subtyping* benturan kepala berbasis pembelajaran mesin menggunakan kepadatan spektral dari kinematika kepala yang terukur, menunjukkan bahwa klasifikasi *Random Forest* dapat mencapai akurasi median

96% dalam membedakan jenis benturan kepala (Alizadeh et al., n.d. 2021). Selanjutnya, mengusulkan *Multiscale Adaptive-switch Random Forest* (MARF) untuk deteksi kaki menggunakan pemindai laser 2D, yang menunjukkan peningkatan kinerja dibandingkan detektor kaki lainnya (Roth et al., n.d. 2022). Selain itu, memperkenalkan *Pairwise Conditional Random Forests* (PCRF) *multi-view* untuk pengenalan ekspresi wajah yang tahan terhadap variasi *pose*, menunjukkan peningkatan kinerja dibandingkan metode *Random Forest* standar (Dapogny et al., n.d.).

Meskipun berbagai penelitian telah dilakukan, optimalisasi teknik pembagian data pada algoritma *Random Forest* khususnya dalam konteks pengolahan data gerak wajah dan kepala masih belum banyak dieksplorasi. Oleh karena itu, penelitian ini bertujuan untuk mengkaji dan mengoptimalkan teknik pembagian data pada algoritma *Random Forest* guna meningkatkan akurasi dan efisiensi dalam analisis data gerak wajah dan kepala.

Oleh karena itu, penelitian ini bertujuan untuk mengeksplorasi dan mengoptimalkan metode pembagian data guna meningkatkan performa *Random Forest* dalam mendeteksi gerak wajah dan kepala, baik dari segi akurasi maupun efisiensi.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana pengaruh teknik pembagian data terhadap kinerja algoritma *Random Forest* dalam pengolahan data gerak wajah dan kepala?
2. Apa teknik pembagian data yang paling optimal untuk meningkatkan akurasi dan efisiensi algoritma *Random Forest*?
3. Apakah variasi parameter dalam algoritma *Random Forest* mempengaruhi efektivitas teknik pembagian data dalam analisis gerak wajah dan kepala?
4. Bagaimana implementasi teknik pembagian data yang *optimal* dapat diterapkan dalam aplikasi nyata yang memanfaatkan analisis gerak wajah dan kepala?

1.3 Tujuan Penelitian

Manfaat yang diharapkan dari penelitian ini adalah:

1. Mengkaji bagaimana berbagai metode pembagian data dapat memengaruhi performa algoritma *Random Forest* dalam menganalisis data pergerakan wajah dan kepala.
2. Mengidentifikasi metode pembagian data yang paling efektif dalam meningkatkan tingkat akurasi serta efisiensi pemrosesan pada algoritma *Random Forest*.
3. Menelusuri pengaruh perubahan parameter dalam algoritma *Random Forest* terhadap kinerja teknik pembagian data dalam proses klasifikasi data gerakan wajah dan kepala.
4. Menyusun panduan implementasi teknik pembagian data yang optimal untuk diterapkan pada sistem berbasis analisis pola gerakan wajah dan kepala di lingkungan teknologi sekarang.

1.4 Manfaat Penelitian

Penelitian ini bertujuan untuk:

1. Menganalisis pengaruh berbagai teknik pembagian data terhadap kinerja algoritma *Random Forest* dalam pengolahan data gerak wajah dan kepala.
2. Mengidentifikasi teknik pembagian data yang optimal untuk meningkatkan akurasi dan efisiensi algoritma *Random Forest* dalam konteks tersebut.
3. Mengkaji pengaruh variasi parameter dalam algoritma *Random Forest* terhadap efektivitas teknik pembagian data dalam analisis gerak wajah dan kepala.
4. Menyusun rekomendasi implementasi teknik pembagian data yang optimal dalam aplikasi nyata yang memanfaatkan analisis gerak wajah dan kepala.

1.5 Ruang Lingkup dan Batasan Penelitian

Untuk menjaga fokus dan ruang lingkup penelitian ini agar tetap terarah serta dapat diselesaikan secara efektif dan efisien, maka penelitian ini dibatasi pada hal-hal berikut:

1. Data yang digunakan dalam penelitian ini bersumber dari *Beyond Smile Challenge*, yang dapat diakses melalui situs <https://sites.google.com/view/beyondsmilechallenge/data>. *Dataset* tersebut terdiri dari sekitar 130.000 entri data yang diperoleh dari 20 *partisipan* yang berinteraksi dalam berbagai skenario dan ekspresi wajah.
2. Format data yang digunakan adalah *JSON*, yaitu hasil keluaran dari perangkat lunak *FacePsychology (FacePsy)*. Data *JSON* ini memuat informasi numerik terkait gerakan wajah dan kepala, seperti landmark wajah, sudut rotasi kepala, intensitas ekspresi, dan atribut afektif lainnya.
3. Penelitian ini hanya akan memanfaatkan fitur-fitur numerik dari *file JSON* yang relevan untuk keperluan pelatihan dan pengujian algoritma *Random Forest*. Atribut lain seperti metadata, teks deskriptif, atau informasi *non-numerik* tidak dianalisis dalam ruang lingkup ini.
4. Penelitian hanya difokuskan pada pengaruh teknik pembagian data (*splitting data*) terhadap kinerja algoritma *Random Forest*, tanpa membandingkannya dengan algoritma lain seperti *Naïve Bayes*, *Decision Tree*, *XGBoost*, *KNN*, *Neural Network* juga lainnya.
5. Penelitian tidak membahas faktor eksternal seperti kualitas video, pencahayaan, atau ketidak konsistenan ekspresi karena dianggap sudah terstandarisasi dalam proses pengambilan data oleh penyedia *dataset*.

Evaluasi kinerja algoritma hanya akan dilakukan menggunakan metrik umum seperti akurasi, presisi, *recall*, dan *F1-score*, serta waktu komputasi dalam proses pelatihan dan prediksi.