

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi di bidang kecerdasan buatan (*Artificial Intelligence*) dalam beberapa tahun terakhir menunjukkan peningkatan yang sangat pesat, khususnya pada pengembangan *Generative Model*. Model ini dapat menciptakan sebuah citra sintetis yang memiliki karakteristik seperti gambar asli yang sulit dibedakan secara langsung oleh indera penglihatan manusia. Kemampuan menghasilkan citra sintetis yang realistis menyebabkan model tersebut banyak dimanfaatkan dalam berbagai bidang, antara lain seni digital, industri kreatif, hiburan, serta memunculkan fenomena *Deepfake* yang semakin populer di masyarakat khususnya pada penyalahgunaan manipulasi citra gambar wajah hasil *Generative model* (Hardiyanto, 2025).

Penyalahgunaan teknologi tersebut berpotensi dimanfaatkan untuk tujuan penipuan seperti memalsukan identitas seseorang, pemerasan, dan penyalahgunaan data pribadi (Prayoga & Tuasikal, 2024). Dalam konteks hukum Indonesia masih belum ada undang-undang yang mengatur tentang penggunaan *Generative AI* ini, namun perbuatan penyalahgunaan teknologi tersebut dapat dikaitkan dengan UU ITE Pasal 27 ayat (3) yaitu melarang distribusi konten elektronik yang melanggar kesusilaan atau menyesatkan, yang dapat mencakup citra wajah sintetis untuk pemalsuan identitas. Selain itu, UU Perlindungan Data Pribadi (PDP) menekankan perlindungan data pribadi, di mana manipulasi citra wajah dapat dianggap sebagai pelanggaran hak privasi, dengan sanksi pidana hingga 6 tahun penjara. Hal ini turut diperkuat oleh PP No. 71 Tahun 2019 tentang PSTE dengan mewajibkan penyelenggara sistem elektronik untuk mencegah penyalahgunaan, termasuk deteksi konten palsu. Ketidaklengkapan regulasi ini menimbulkan celah hukum, di mana deteksi otomatis citra *Generative AI* menjadi krusial untuk mendukung penegakan hukum.

Saat ini, membedakan citra wajah asli dan palsu hanya dengan pengamatan visual secara langsung terbilang sangat sulit. Penelitian oleh (Velásquez-salamanca et al., 2025), membuktikan bahwa manusia secara umum tidak mampu mengidentifikasi sumber gambar secara akurat, di mana hampir 40% citra hasil *Generative AI* gagal dikenali dan justru diklasifikasikan sebagai foto asli. Hal ini disebabkan oleh perkembangan *Generative AI* yang mampu menghasilkan gambar dengan kualitas tinggi dan tampak realistis. Meskipun tampak sempurna secara visual, namun secara teoritis citra wajah hasil *Generative AI* tetap memiliki anomali dan distribusi statistik yang berbeda dari citra asli. Menurut penelitian oleh (Osińska et al., 2025), menyebutkan bahwa karakteristik citra wajah asli dapat dikenali dari adanya ketidaksempurnaan kulit yang alami, detail bayangan yang konsisten, serta naturalitas pada mimik dan mata manusia. Sebaliknya, citra hasil *Generative AI* sering kali meninggalkan jejak kecacatan digital, seperti tekstur kulit yang terlalu mulus, detail tepi yang terlampaui tajam, serta kejanggalan pada pantulan cahaya mata, garis luar pupil, hingga kelainan pada transisi objek yang rumit seperti rambut.

Berbagai penelitian terdahulu telah berupaya merancang sistem otomatis untuk mendeteksi citra wajah sintetis ini. Pendekatan yang paling dominan digunakan saat ini adalah arsitektur *Deep Learning*, khususnya *Convolutional Neural Network (CNN)*. Penelitian oleh (Mu et al., 2024), menunjukkan bahwa model *CNN* dengan arsitektur menyerupai *VGG* dengan tambahan *Batch Normalization* terbukti mampu mengklasifikasikan antara citra wajah asli dan *Deepfake* dengan tingkat akurasi mencapai 91%. Meskipun model *Deep Learning* seperti *CNN* terbukti mampu mencapai tingkat akurasi yang tinggi, namun pendekatan ini memiliki beberapa kelemahan teknis. Menurut (Barus et al., 2025), Algoritma ini cenderung membutuhkan sumber daya komputasi yang besar serta waktu pelatihan yang relatif lama. Selain itu, kompleksitas model membuatnya rentan terhadap risiko *overfitting* jika tidak didukung oleh volume data yang sangat besar dan berkualitas serta kemampuannya dalam mengekstrak fitur secara otomatis yang justru menyebabkan *CNN* memiliki sifat *black-box*. Hal ini berakibat pada proses pengambilan keputusan yang sangat sulit diinterpretasikan, sehingga sulit

untuk mengetahui secara pasti fitur spesifik apa, baik tekstur maupun warna yang digunakan oleh model dalam mendeteksi citra tersebut.

Sebagai upaya untuk mengatasi keterbatasan teknis dan kendala interpretasi dari arsitektur *Deep Learning* tersebut, diperlukan pendekatan alternatif yang lebih efisien, transparan, dan ringan secara komputasi. Sebagai solusi, penelitian ini mengusulkan penggunaan metode klasifikasi berbasis *Machine Learning* dengan mengandalkan fusi fitur dan algoritma *eXtreme Gradient Boosting (XGBoost)*. Algoritma ini menawarkan performa yang sangat tangguh dalam mengklasifikasikan data tabular berdimensi tinggi yang dihasilkan dari proses ekstraksi fitur citra.

Keunggulan penggunaan algoritma *XGBoost* dalam ranah deteksi citra sintesis ini telah divalidasi oleh penelitian sebelumnya. Sebagai contoh penelitian yang dilakukan oleh (Kusnaeni et al., 2025), mengusulkan pendekatan hybrid untuk membedakan konten asli dan konten buatan AI, di mana arsitektur *ResNet50* digunakan untuk mengekstrak fitur visual, dan algoritma *XGBoost* digunakan sebagai model pengklasifikasinya. Hasil eksperimen dari penelitian tersebut menunjukkan bahwa *XGBoost* sangat efektif dalam mengklasifikasikan vektor fitur citra dengan tingkat akurasi mencapai 94,6%. Selain akurasi yang tinggi, model ini terbukti menawarkan keseimbangan yang sangat baik antara performa prediksi dan efisiensi komputasi.

Berdasarkan urgensi permasalahan dalam latar belakang ini dan celah dalam penelitian terdahulu, pengembangan sistem yang efisien dan transparan menjadi sebuah keharusan. Melalui fusi ekstraksi fitur tekstur dan warna yang diklasifikasikan menggunakan algoritma *XGBoost*. Penelitian ini diharapkan mampu menghasilkan sistem pendeteksi citra wajah hasil *Generative AI* yang tidak hanya akurat dan ringan secara komputasi, tetapi juga mampu memberikan interpretasi yang jelas mengenai landasan keputusannya.

Berdasarkan penjelasan uraian latar belakang yang telah disampaikan di atas, maka penulis tertarik untuk mengangkat permasalahan beserta solusi tersebut ke dalam sebuah penelitian skripsi dengan judul " ***Klasifikasi Citra Wajah Asli Dan***

Wajah Hasil Generative AI Menggunakan XGBoost Berbasis Fitur Tekstur Dan Warna".

1.2 Rumusan Masalah

Berdasarkan ulasan latar belakang di atas, rumusan masalah dalam penelitian ini diuraikan sebagai berikut:

1. Bagaimana cara mengekstraksi fitur tekstur dan warna untuk membedakan karakteristik citra wajah asli dan citra wajah hasil *Generative AI*?
2. Bagaimana penerapan algoritma *XGBoost* dalam mengklasifikasikan citra wajah asli dan citra wajah hasil *Generative AI* berdasarkan data hasil ekstraksi fitur tersebut?
3. Bagaimana mengevaluasi model algoritma *XGBoost* dalam melakukan klasifikasi citra wajah asli dan citra wajah hasil *Generative AI*?

1.3 Tujuan Penelitian

Berdasarkan uraian rumusan masalah di atas, tujuan penelitian ini diuraikan sebagai berikut:

1. Untuk menjelaskan cara ekstraksi fitur tekstur dan warna pada citra wajah melalui tahapan konversi citra ke grayscale untuk metode *GLCM*, konversi citra *RGB* ke ruang warna *YCbCr*, serta pengambilan nilai statistik fitur yang digunakan sebagai pembeda antara citra wajah asli dan citra wajah hasil *Generative AI*.
2. Untuk menjelaskan cara penerapan algoritma *XGBoost* dalam membangun model klasifikasi biner, dimulai dari pembagian dataset, normalisasi vektor fitur, pelatihan model, hingga prediksi kelas citra wajah asli dan citra wajah hasil *Generative AI*.
3. Untuk menjelaskan cara mengevaluasi performa model *XGBoost* dengan menggunakan *confusion matrix* serta menghitung metrik *accuracy*, *precision*, *recall*, dan *F1-score* pada data uji.

1.4 Manfaat Penelitian

Adapun beberapa manfaat dari penelitian yang dilakukan adalah sebagai berikut:

1. Penelitian ini membantu memahami bagaimana sebuah citra wajah memiliki nilai fitur yang dapat digunakan sebagai variabel untuk proses klasifikasi citra menggunakan model *Machine Learning*.
2. Penelitian ini diharapkan dapat menghasilkan sebuah sistem yang mempermudah proses pengecekan keaslian foto wajah sehingga dapat digunakan untuk mengurangi kejahatan siber.
3. Hasil penelitian dapat menjadi referensi untuk penelitian-penelitian selanjutnya yang berfokus pada klasifikasi citra dengan objek berbeda maupun dengan metode yang lebih baik.

1.5 Ruang Lingkup Dan Batasan penelitian

Adapun ruang lingkup dan batasan penelitian adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini hanya berupa citra gambar wajah diam manusia, bukan berupa video atau suara.
2. Dataset yang digunakan dalam penelitian ini berjumlah 750 citra wajah yang terbagi secara seimbang ke dalam dua kelas, yaitu 375 citra wajah asli dan 375 citra wajah hasil *Generative AI*. Setiap kelas dibagi menjadi 80% data latih dan 20% data uji, sehingga masing-masing kelas terdiri atas 300 citra data latih dan 75 citra data uji.
3. Data citra wajah asli diperoleh dari dataset terbuka, seperti *Kaggle*, media sosial, Platform *Pinterest*, dan sumber daring yang relevan dengan kebutuhan penelitian, serta hasil pengambilan gambar secara langsung menggunakan kamera. Adapun data citra wajah hasil *Generative AI* diperoleh melalui proses generate gambar menggunakan beberapa model dan platform *image generator*, meliputi *GPT Image 4.2* pada platform *ChatGPT* yang dikembangkan oleh *OpenAI*, *Nano Banana* pada platform *Gemini* yang dikembangkan oleh *Google*, *Leonardo AI*, *Canva*, serta *Dreamina AI* yang dikembangkan oleh *ByteDance*.

4. Citra wajah yang dianalisis pada proses klasifikasi hanya terbagi menjadi dua kelas yaitu citra wajah asli manusia dan citra wajah hasil *Generative AI*.
5. Ciri-ciri atau fitur yang diambil dari citra wajah hanya dibatasi pada dua hal, yaitu fitur tekstur dan fitur warna.
6. Metode yang digunakan untuk ekstraksi fitur tekstur hanya menggunakan *Gray Level Co-occurrence Matrix (GLCM)*, yaitu metode analisis hubungan spasial antar piksel keabuan, sedangkan fitur warna diekstraksi menggunakan ruang warna *YCbCr*, yaitu ruang warna yang memisahkan komponen *luminansi (Y)* dan *krominansi (Cb dan Cr)*.
7. Algoritma yang digunakan untuk mengklasifikasikan citra tersebut hanya difokuskan pada algoritma *eXtreme Gradient Boosting (XGBoost)*.