

BAB I

PENDAHULUAN

1.1 Latar Belakang Penelitian

Perkembangan teknologi digital telah membawa dampak besar bagi kehidupan masyarakat di seluruh dunia, termasuk Indonesia. Data dari DataReportal menunjukkan bahwa pada awal tahun 2024, Indonesia memiliki 278,7 juta penduduk dengan 185,3 juta pengguna internet, dengan tingkat penetrasi mencapai 66,5%. Sebanyak 139 juta pengguna media sosial aktif menjadikan media sosial sebagai bagian integral dari kehidupan sehari-hari masyarakat [1]. Dengan besarnya jumlah pengguna, konten digital yang tersebar di media sosial pun meningkat secara eksponensial, termasuk video, audio, dan gambar yang sering kali telah dimanipulasi secara canggih.

Manipulasi konten digital, seperti *deepfake*, menjadi tantangan baru dalam menjaga keamanan dan validitas informasi di ruang digital. *Deepfake* adalah teknologi berbasis kecerdasan buatan (AI) yang dapat menghasilkan video, gambar, atau suara yang tampak sangat meyakinkan namun sebenarnya telah dipalsukan [2]. Data dari *Yahoo!Finance* menunjukkan bahwa konten *deepfake* di internet tumbuh dengan kecepatan 400 persen setiap tahunnya [3]. Fenomena ini bukan hanya sekadar fenomena sosial, tetapi juga membawa risiko signifikan, terutama dalam hal penipuan, pencemaran nama baik, dan penyebaran hoaks [4].

Bahkan, menurut laporan Tempo, manipulasi semacam ini telah digunakan untuk menipu masyarakat dengan menyalahgunakan nama figur publik di Indonesia, seperti pada kasus pemalsuan video Dokter Terawan untuk promosi obat palsu [5]. Kejadian serupa juga ditemukan di dunia internasional, di mana *deepfake* sering digunakan untuk tujuan propaganda politik, penipuan siber, atau bahkan serangan terhadap individu yang disasar secara personal.

Lebih jauh, *deepfake* tidak hanya digunakan untuk tujuan penipuan online bermotif keuangan, tetapi juga untuk fitnah, pencemaran nama baik, hoaks, dan ujaran kebencian. Salah satu contohnya adalah ketika teknologi *deepfake* digunakan untuk meniru tokoh terkenal, seperti Mark Zuckerberg, dalam video yang berisi

pesan manipulatif mengenai kontrol data pribadi. Video ini menghebohkan masyarakat dan menambah konotasi negatif terhadap teknologi *deepfake* [6]. Modus manipulasi seperti ini memperburuk kepercayaan publik terhadap platform digital dan berpotensi menyebabkan kerugian besar bagi individu atau organisasi yang disalahgunakan namanya.

Dalam upaya mengatasi masalah ini, teknologi deep learning muncul sebagai solusi potensial untuk mendeteksi konten *deepfake*. Deep learning memungkinkan komputer mempelajari pola kompleks dari data, termasuk fitur-fitur unik pada wajah manusia yang mungkin sulit dideteksi oleh mata manusia biasa. Salah satu teknik deep learning yang efektif dalam mendeteksi *deepfake* adalah Convolutional Neural Network (CNN). CNN mampu mengenali perbedaan halus pada piksel gambar yang mungkin tidak disadari oleh pengguna biasa, sehingga dapat mengidentifikasi apakah suatu gambar merupakan hasil manipulasi atau tidak. Dengan algoritma CNN, sistem dapat dilatih untuk mengenali ciri-ciri khas dari gambar *deepfake*, memberikan hasil deteksi yang lebih akurat dan cepat.

Beberapa penelitian sebelumnya telah menunjukkan perkembangan signifikan dalam bidang deteksi *deepfake*. Misalnya, Fatwa Akbar Jiwani dan Bagus Satrio Waluyo Poetro (2024) mengembangkan model deteksi *deepfake* berbasis DenseNet-121 dengan teknik watermarking LSB [7], sedangkan Mu et al. (2024) menggunakan CNN dengan arsitektur VGG dan batch normalization yang mampu mencapai akurasi hingga 91% [8]. Pendekatan hibrida juga mulai diterapkan, seperti kombinasi CNN dan Extreme Learning Machine (ELM) oleh Alvian Dwi Sanjaya (2024) [9], serta CNN dan SVM oleh Jacob Mallet et al. (2023) [10]. Selain itu, pendekatan berbasis Transformer dan frequency masking mulai diadopsi untuk meningkatkan generalisasi model terhadap dataset baru [10][12][16].

Berbeda dari penelitian-penelitian tersebut, penelitian ini tidak hanya berfokus pada pengujian performa model CNN dalam mendeteksi wajah palsu, tetapi juga mengimplementasikan model ResNet-50 ke dalam sistem aplikasi berbasis web menggunakan REST API. Dengan pendekatan ini, pengguna dapat mengunggah gambar wajah dan langsung memperoleh hasil klasifikasi secara *real-*

time. Selain itu, penelitian ini menekankan pada eksperimen konfigurasi hyperparameter seperti batch size dan learning rate, serta mengevaluasi ketahanan model terhadap domain shift melalui pengujian tambahan pada dataset primer. Dengan demikian, penelitian ini diharapkan memberikan kontribusi tidak hanya dalam aspek akurasi, tetapi juga dalam hal penerapan praktis dan ketahanan model terhadap kondisi dunia nyata.

Penelitian ini memanfaatkan OpenForensics, Dataset yang merupakan kumpulan data berskala besar yang dirancang untuk mendeteksi dan melakukan segmentasi pemalsuan wajah dalam situasi dunia nyata (in-the-wild). Dataset ini mencakup lebih dari 115.000 gambar dengan sekitar 334.000 wajah manusia, di mana sebagian telah dimanipulasi menggunakan berbagai teknik *deepfake* modern seperti FaceShifter, FSGAN, dan DF-VAE. Setiap wajah dalam gambar dianotasi secara rinci, mencakup label *real* atau *fake*, kotak pembatas wajah, peta batas pemalsuan, segmentasi piksel wajah palsu, dan titik-titik landmark wajah. Tidak seperti dataset lain, OpenForensics menantang karena satu gambar dapat mengandung banyak wajah dengan hanya sebagian yang dipalsukan, sehingga menuntut model deteksi yang lebih canggih dan tahan terhadap variasi kontekstual seperti pencahayaan, ekspresi, dan latar belakang. Dataset ini sangat berguna untuk pengembangan model deteksi dan segmentasi *deepfake* yang akurat dan robust di lingkungan dunia nyata[10]. Dataset ini tersedia secara publik di platform Kaggle, yang merupakan salah satu repository terpercaya. Sebelum digunakan, dataset ini akan di *sampling* sesuai dengan rasio dataset yang akan digunakan untuk eksperimen penelitian lalu digunakan untuk melatih, memvalidasi, dan menguji model ResNet50, salah satu arsitektur CNN yang diandalkan untuk mendeteksi manipulasi wajah.

Model yang telah dilatih akan diimplementasikan ke dalam sistem berbasis website untuk memungkinkan mengunggah gambar dan mendapatkan hasil analisis keasliannya secara *real-time*. Sistem ini diharapkan tidak hanya membantu meningkatkan literasi digital, tetapi juga memberikan solusi praktis dalam mendeteksi dan menangkal penyebaran konten manipulatif di media sosial. Dengan

pendekatan ini, penelitian ini berpotensi berkontribusi secara signifikan dalam menciptakan ekosistem digital yang lebih aman dan terpercaya di Indonesia.

1.2 Rumusan Masalah

Berdasarkan latar belakang permasalahan diatas, maka dapat dirumuskan permasalahan pada penelitian ini adalah sebagai berikut :

1. Bagaimana mengimplementasikan algoritma Covolutional Neural Network (CNN) dengan arsitektur Resnet-50 untuk *deepfake* face detection?
2. Bagaimana kinerja algoritma Convolutional Neural Network (CNN) dengan arsitektur Resnet-50 untuk *deepfake* dace detection?
3. Bagaimana cara membangun aplikasi *deepfake* face detection yang diintegrasikan dengan model berupa REST API?

1.3 Tujuan Penelitian

Tujuan yang hendak dicapai dalam penelitian ini adalah sebagai berikut :

1. Untuk mengimplementasikan algoritma Convolutional Neural Network (CNN) dengan arsitektur Resnet-50 untuk *deepfake* face detection.
2. Untuk mengetahui kinerja algoritma Convolutional Neural Network (CNN) dengan arsitektur Resnet-50 untuk *deepfake* face detection.
3. Untuk membangun sebuah aplikasi *deepfake* face detection yang diintegrasikan dengan model berupa REST API sehingga dapat digunakan pada berbagai platform.

1.4 Manfaat Penelitian

Penelitian ini diharapkan mampu memiliki manfaat sebagai berikut :

1. Membantu untuk mengembangkan penelitian *deepfake* face detection secara *real-time*.
2. Memberikan pengetahuan mengenai *deepfake* face detection menggunakan algoritma Convolutional Neural Network dengan arsitektur Resnet-50.

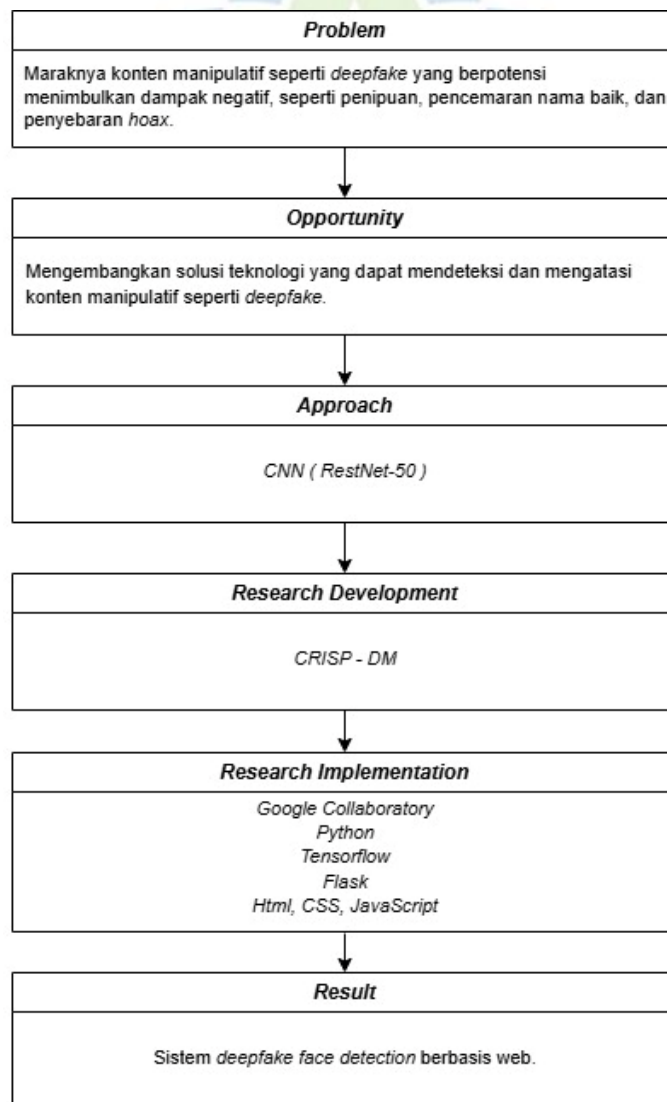
1.5 Batasan Masalah

Penelitian ini menetapkan beberapa batasan masalah yang akan peneliti terapkan, yaitu :

1. Metode untuk mendeteksi gambar *deepfake* menggunakan algoritma CNN dengan arsitektur *Resnet-50*.
2. Media yang dideteksi hanya berupa gambar spesifik wajah manusia tidak berupa audio atau video.
3. Sumber data atau dataset didapatkan dari *OpenForensic* dari *Kaggle*, dataset awal terdiri dari 190.334 data lalu *disampling* sebesar 10%.
4. Penelitian ini hanya sampai pada proses identifikasi pada aplikasi web.

1.6 Kerangka Pemikiran

Kerangka pemikiran merupakan dasar pemikiran penelitian untuk menjelaskan hubungan antar konsep dan teori. Kerangka pemikiran penelitian ini ditunjukkan pada gambar 1.1.



Gambar 1. 1 Kerangka Pemikiran

1.7 Sistematika Penulisan

Sistematika penulisan perangkat lunak terdiri dari lima bab, masing-masing dengan cakupan sebagai berikut:

BAB I PENDAHULUAN

Menjelaskan latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, kerangka penelitian, dan sistematika penulisan. Bab ini berfungsi sebagai pengantar untuk memberikan pemahaman tentang konteks dan urgensi penelitian.

BAB II TINJAUAN PUSTAKA

Membahas literatur dan penelitian terdahulu, termasuk konsep, teori, dan model yang mendasari perancangan sistem. Analisis dalam bab ini membantu mengidentifikasi kelemahan dan kesenjangan penelitian sebelumnya sebagai dasar pengembangan lebih lanjut.

BAB III METODOLOGI PENELITIAN

Menguraikan tahapan sistematis dalam memperoleh data secara objektif dan valid untuk menjawab pertanyaan penelitian. Bab ini menjelaskan metode yang digunakan serta implementasi dalam penelitian.

BAB IV HASIL DAN PEMBAHASAN

Menyajikan hasil penelitian serta analisisnya untuk menjawab rumusan masalah. Pembahasan dilakukan berdasarkan data yang diperoleh guna memperjelas temuan penelitian.

BAB V PENUTUP

Menarik kesimpulan dari hasil penelitian serta memberikan saran untuk penelitian selanjutnya guna meningkatkan kualitas studi di bidang terkait.