

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan hoaks di internet merupakan permasalahan serius yang dapat memanipulasi pola pikir masyarakat dengan informasi yang salah dan tidak dapat diverifikasi kebenarannya [1]. Namun, seiring dengan meningkatnya penyebaran hoaks melalui teks, masyarakat menjadi semakin kritis terhadap informasi berbasis teks yang mereka terima [2]. Peningkatan literasi digital ini mendorong berbagai inisiatif untuk mendidik masyarakat dalam mengenali dan menanggulangi hoaks, sehingga dampak negatifnya dapat diminimalkan.

Seiring dengan kemajuan teknologi, bentuk penyebaran hoaks tidak lagi terbatas pada teks saja, melainkan telah merambah ke media lain, termasuk audio. Teknologi deepfake, khususnya dalam bidang audio, telah mengalami perkembangan pesat dalam beberapa tahun terakhir. Deepfake audio adalah hasil sintesis suara yang dibuat menggunakan model generatif seperti Generative Adversarial Networks (GANs) atau Text-to-Speech (TTS), yang memungkinkan pembuatan suara yang hampir identik dengan suara asli, bahkan dengan durasi panjang dan intonasi natural. Meskipun teknologi ini memiliki potensi positif, seperti dalam industri hiburan dan restorasi audio, penerapannya untuk tujuan manipulasi dan penipuan menimbulkan ancaman serius. Menurut laporan dari Deeptrace Labs, jumlah konten deepfake di internet meningkat sebesar 84% dalam waktu kurang dari satu tahun, yang menunjukkan peningkatan signifikan dalam penyalahgunaan teknologi ini [3].

Transisi dari hoaks berbasis teks ke deepfake audio menandai pergeseran lanskap disinformasi digital. Di masa lalu, hoaks teks banyak beredar melalui media sosial dan pesan instan, namun kini teknologi deepfake audio memungkinkan penjahat siber untuk menciptakan rekaman suara palsu yang sangat meyakinkan. Beberapa kasus internasional telah menunjukkan bagaimana penipuan berbasis deepfake audio dapat menimbulkan kerugian finansial dan mengganggu kestabilan kepercayaan publik. Misalnya, pada tahun 2019, penipu menggunakan kecerdasan buatan untuk meniru suara CEO sebuah perusahaan energi di Jerman, yang kemudian mengarahkan anak perusahaan di Inggris untuk mentransfer dana sebesar €220.000 [4]. Kemudian, pada tahun 2020, metode serupa digunakan dalam skema

yang meyakinkan manajer cabang untuk mentransfer \$35 juta dengan memanipulasi suara seorang direktur perusahaan [5]. Lebih lanjut, pada tahun 2023, sebuah rekaman audio deepfake yang meniru suara pemimpin Partai Buruh Inggris, Keir Starmer, diluncurkan pada hari pertama konferensi partai, yang menunjukkan betapa mudahnya opini publik dapat dimanipulasi melalui media audio [6]. Selain itu, studi dari University College London (UCL) mengungkapkan bahwa dari 529 partisipan, mereka bisa membedakan deepfake dengan 73% angka keyakinan, bahkan setelah mengikuti pelatihan singkat [7].

Tantangan teknis dalam mendeteksi deepfake audio semakin kompleks seiring dengan kemajuan model-model generatif seperti WaveNet dan Tacotron 2, yang mampu menghasilkan audio dengan kualitas hampir sempurna. Studi oleh ASVspoof menunjukkan bahwa akurasi deteksi menggunakan metode konvensional hanya mencapai 60-70%, yang masih jauh dari tingkat keandalan yang diharapkan [8]. Kondisi ini memperlihatkan bahwa pendekatan deteksi manual atau berbasis aturan sudah tidak memadai untuk mengatasi tingkat kecanggihan deepfake saat ini.

Dengan semakin kompleksnya penyalahgunaan deepfake audio, kebutuhan akan sistem deteksi yang akurat dan efisien menjadi semakin mendesak. Pengembangan model deteksi berbasis Generative Adversarial Networks (GANs) diharapkan dapat memberikan solusi yang lebih baik dalam mengidentifikasi audio palsu. Selain digunakan untuk menghasilkan data sintetis, GANs memiliki potensi untuk membangun model yang mampu mempelajari pola-pola spesifik dari deepfake audio sehingga dapat meningkatkan akurasi deteksi [9]. Untuk memungkinkan model menganalisis audio, sinyal suara satu dimensi (1D) perlu diubah menjadi representasi visual dua dimensi (2D) yang kaya akan fitur. Salah satu representasi yang paling umum dan efektif untuk tugas ini adalah Mel-spektrogram [10]. Mel-spektrogram adalah visualisasi dari spektrum frekuensi audio seiring berjalannya waktu, yang dimodifikasi menggunakan skala Mel agar sesuai dengan persepsi pendengaran manusia. Dengan mengubah audio menjadi gambar spektrogram, model berbasis Convolutional Neural Network (CNN), yang merupakan inti dari arsitektur GANs dalam penelitian ini, dapat secara efektif mengekstraksi fitur-fitur tekstural untuk membedakan antara audio asli dan palsu. Penelitian ini bertujuan untuk mengembangkan model deteksi deepfake audio berbasis GANs yang dapat diandalkan, sehingga diharapkan dapat mengurangi dampak negatif dari penyalahgunaan teknologi deepfake dalam konteks disinformasi digital.

Penelitian ini juga akan mengkaji aspek teknis dan etis yang muncul akibat penyalahgunaan deepfake audio. Pengembangan teknologi deteksi bukan hanya penting untuk melindungi sektor industri dan keuangan dari penipuan, tetapi juga untuk menjaga kepercayaan publik terhadap media dan institusi pemerintahan. Dengan demikian, hasil penelitian diharapkan tidak hanya memberikan kontribusi pada pengembangan teknologi keamanan siber, tetapi juga mendukung upaya global dalam menjaga integritas informasi di era digital.

1.2 Rumusan Masalah

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

1. Bagaimana cara konversi audio menjadi Mel-Spectrogram?
2. Bagaimana cara membangun model Deep Convolutional Generative Adversarial Networks (DCGANs) untuk mendeteksi deepfake audio file?
3. Bagaimana cara mengevaluasi hasil dari model DCGANs dalam membedakan audio *fake* dengan yang *real*?

Tabel 1.1. Tabel Rumusan Masalah

RM	Rumusan Masalah
RM1	Bagaimana cara konversi audio file menjadi Mel-Spectrogram?
RM2	Bagaimana cara membangun model Deep Convolutional Generative Adversarial Networks (DCGANs) untuk mendeteksi deepfake audio?
RM3	Bagaimana cara mengevaluasi hasil dari model DCGANs dalam membedakan audio <i>fake</i> dengan yang <i>real</i> ?

1.3 Batasan Permasalahan

Pada bagian ini dijabarkan batasan yang diterapkan dalam penelitian ini agar pelaksanaan penelitian menjadi lebih terfokus kepada aspek-aspek berikut:

1. Penelitian hanya berfokus pada deteksi deepfake audio.
2. Dataset yang digunakan terbatas pada audio manusia (speech) dalam Bahasa Inggris.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah tertulis, berikut merupakan tujuan dari penelitian ini.

1. Konversi audio file menjadi Mel-Spectrogram.
2. Membangun model Deep Convolutional Generative Adversarial Networks (DCGANs) untuk mendeteksi deepfake audio.
3. Mengevaluasi model DCGANs dalam membedakan audio real dan audio fake.

Tabel 1.2. Tabel Tujuan Penelitian

RM	TP	Tujuan Penelitian
RM1	TP1	Konversi audio file menjadi Mel-Spectrogram
RM2	TP2	Membangun model Deep Convolutional Generative Adversarial Networks (DCGANs) untuk mendeteksi deepfake audio
RM3	TP3	Mengevaluasi model DCGANs dalam membedakan audio <i>real</i> dan audio <i>fake</i>

1.5 Manfaat Penelitian

Manfaat penelitian ini dibagi menjadi tiga, yaitu:

1. Manfaat Akademis, Memberikan kontribusi dalam pengembangan metode deteksi deepfake audio dengan menggunakan DCGANs.

2. Manfaat Praktis, Menyediakan alat deteksi deepfake audio yang dapat digunakan untuk mencegah penyalahgunaan teknologi deepfake.
3. Manfaat Sosial, Meningkatkan kesadaran masyarakat tentang bahaya deepfake audio dan pentingnya deteksi dini.

1.6 Sistematika Penulisan

Berikut merupakan sistematika dari penulisan mengenai Pengembangan Model Deteksi Deepfake Audio Berbasis Algoritma Deep Convolutional Generative Adversarial Networks (DCGANs).

Sistematika penulisan laporan adalah sebagai berikut:

1. Bab 1 PENDAHULUAN

Bab ini berisi mengenai penjelasan latar belakang masalah atas meningkatnya ancaman siber dalam bentuk deepfake audio dan juga menegaskan seberapa pentingnya sistem deteksi deepfake guna untuk mengatasi ancaman siber tersebut.

2. Bab 2 LANDASAN TEORI

Bab ini secara komprehensif menguraikan landasan teoretis yang menjadi dasar fundamental bagi penelitian ini. Pembahasan akan dimulai dari konsep-konsep umum dalam Artificial Intelligence dan Keamanan Siber (Cyber Security), kemudian mengerucut secara progresif ke konsep yang lebih spesifik, yaitu Deep Convolutional Generative Adversarial Networks (DCGANs) dan penerapannya dalam mendeteksi audio deepfake. Disertakan juga riset-riset terkait yang menjadi basis dari penelitian ini.

3. Bab 3 METODOLOGI PENELITIAN

Bab ini berisi metode yang digunakan dalam penelitian, mulai dari *preprocessing* dataset 'Fake or Real (FoR)', konfigurasi model DCGAN untuk mengklasifikasi audio, proses training yang dilakukan, dan juga pengujian serta analisis hasil model.

4. Bab 4 HASIL DAN DISKUSI

Bab ini berisi seluruh hasil dari penelitian yang telah dilakukan terkait pelatihan model DCGAN untuk klasifikasi audio. Mulai dari implementasi metode dalam bentuk kode, permasalahan dalam overfitting yang telah diatasi, dan juga analisis beserta diskusi dari hasil penelitian.

5. Bab 5 KESIMPULAN DAN SARAN

Bab ini menyimpulkan penelitian secara menyeluruh. Mulai dari latar belakang masalah yang muncul, tujuan penelitian yang terpenuhi, pembahasan singkat mengenai analisis hasil penelitian, dan juga saran untuk para peneliti selanjutnya.

