BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan metode berbasis kecerdasan buatan (Artificial Intelligence/AI), khususnya deep learning, dalam beberapa tahun terakhir telah mendorong peningkatan kemampuan teknologi deepfake pada bidang sintesis suara. Kemampuan model AI modern dalam menghasilkan suara buatan yang semakin menyerupai manusia menimbulkan tantangan serius dalam proses verifikasi keaslian media digital, menjaga kepercayaan masyarakat, serta perlindungan keamanan siber [1]. Tingkat kemiripan antara suara sintetis dan suara manusia yang semakin tinggi menyebabkan metode deteksi konvensional berbasis pemeriksaan manual menjadi kurang efektif dalam membedakan keduanya secara akurat. Kondisi ini berpotensi memicu berbagai dampak negatif, seperti penyebaran informasi palsu, tindak penipuan berbasis suara, hingga penyalahgunaan identitas [2][3].

Model generatif terkini tidak hanya mampu mengubah teks menjadi suara (text-to-speech), tetapi juga melakukan voice cloning dengan meniru karakteristik unik pembicara dari sampel audio berdurasi singkat [4] [5]. Walaupun teknologi ini memberikan manfaat dalam industri kreatif dan layanan digital, pemanfaatannya secara tidak bertanggung jawab dapat menimbulkan ancaman keamanan yang signifikan.

Salah satu ancaman yang paling mengkhawatirkan adalah voice phishing (vishing), yaitu modus penipuan yang menggunakan rekayasa suara untuk memperoleh informasi sensitif, seperti data perbankan, kode verifikasi, atau kredensial pribadi [6]. Dengan memanfaatkan AI-generated speech, pelaku dapat meniru suara individu tertentu secara realistis sehingga meningkatkan kepercayaan korban dan memperbesar peluang keberhasilan serangan [7]. Laporan keamanan siber tahun 2023–2025 menunjukkan peningkatan signifikan penggunaan deepfake audio dalam skema penipuan finansial dan impersonasi identitas [8]. Meningkatnya kasus vishing berbasis AI-generated speech tersebut menegaskan pentingnya sistem deteksi yang mampu mengidentifikasi rekayasa suara secara otomatis dan akurat. Dalam konteks ini, penelitian mengenai spoofing detection dan synthetic speech detection menjadi semakin relevan untuk dikaji sebagai langkah mitigasi terhadap

ancaman tersebut.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang diuraikan, dapat dirumuskan masalah yang ingin diselesaikan yaitu:

a. Bagaimana kinerja model Wav2Vec 2.0 berbasis *transfer learning* dalam mendeteksi *AI-generated speech* dibandingkan dengan metode *baseline* yang menggunakan fitur akustik tradisional?

b. Seberapa baik kemampuan model dalam membedakan antara suara manusia (*human speech*) dan suara sintetis (*AI-generated speech*) berdasarkan metrik evaluasi klasifikasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*?

1.3 Batasan Permasalahan

Untuk menjaga agar penelitian tetap terfokus dan sesuai dengan ruang lingkup yang telah ditetapkan, maka batasan permasalahan dalam penelitian ini adalah sebagai berikut:

a. Dataset yang digunakan dalam penelitian ini dibatasi pada *The Fake-or-Real (FoR) Dataset*, khususnya sub-dataset *for-2sec* yang memiliki format berkas `.wav` dengan laju sampel 16 kHz dan audio mono. Dataset tersebut terdiri atas dua kelas, yaitu suara asli (*REAL*) dan suara sintetis (*FAKE*). Tahap prapemrosesan dibatasi pada proses *resampling* menjadi 16 kHz, konversi audio menjadi mono, serta normalisasi amplitudo menggunakan metode *Peak Normalization*.

b. Penelitian ini hanya melakukan klasifikasi audio ke dalam dua kategori (*binary classification*), yaitu suara asli (*human speech*) dan suara sintetis (*AI-generated speech*). Penelitian tidak membedakan jenis model kecerdasan buatan yang digunakan untuk menghasilkan suara sintetis maupun melakukan identifikasi terhadap identitas pembicara.

c. Penelitian ini tidak membahas implementasi sistem secara langsung pada perangkat keras telekomunikasi (*real-time deployment*) maupun integrasi fisik dengan sistem *Voice over Internet Protocol* (VoIP). Penelitian

difokuskan pada pengembangan model, proses pelatihan, pengujian berbagai skenario parameter, serta evaluasi kinerja model deteksi.

d. Evaluasi kinerja model dibatasi pada metrik evaluasi klasifikasi yang meliputi *Accuracy*, *Precision*, *Recall*, dan *F1-Score*.

1.4 Tujuan Penelitian

Penelitian ini memiliki tujuan sebagai berikut:

a. Menganalisis dan membandingkan kinerja model Wav2Vec 2.0 berbasis *transfer learning* dalam mendeteksi *AI-generated speech* dengan metode *baseline* yang menggunakan fitur akustik tradisional.

b. Mengevaluasi kemampuan model dalam membedakan antara suara manusia (*human speech*) dan suara sintetis (*AI-generated speech*) berdasarkan metrik evaluasi klasifikasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan beberapa manfaat sebagai berikut:

a. Memberikan kontribusi ilmiah dalam pengembangan metode deteksi *AI-generated speech* melalui penerapan *transfer learning* menggunakan model Wav2Vec 2.0.

b. Menyajikan gambaran implementasi yang terstruktur terkait proses pengolahan data audio, pelatihan model, serta evaluasi performa dalam sistem klasifikasi suara.

c. Menghasilkan sistem yang mampu membedakan antara suara manusia asli dan suara sintetis secara otomatis dengan tingkat akurasi yang dapat diukur.

d. Memberikan manfaat praktis dalam mendukung upaya pencegahan penyalahgunaan teknologi sintesis suara, serta meningkatkan kepercayaan terhadap informasi audio di media digital.

1.6 Sistematika Penulisan

Sistematika penulisan laporan penelitian ini disusun untuk memberikan gambaran mengenai struktur pembahasan pada setiap bab dalam penelitian. Adapun sistematika penulisan laporan penelitian ini adalah sebagai berikut:

a. Bab 1 PENDAHULUAN

   Bab ini berisi latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan laporan. Bab ini memberikan gambaran umum mengenai permasalahan yang diangkat dan arah penelitian yang dilakukan.

b. Bab 2 LANDASAN TEORI

   Bab ini memaparkan teori-teori yang menjadi dasar dalam penelitian, meliputi konsep-konsep yang berkaitan dengan kecerdasan buatan, *deepfake* audio, serta metode yang digunakan dalam penelitian seperti *transfer learning* dan model Wav2Vec 2.0. Landasan teori ini digunakan sebagai dasar dalam melakukan analisis dan pengembangan penelitian.

c. Bab 3 METODOLOGI PENELITIAN

   Bab ini menjelaskan metode penelitian yang digunakan, meliputi tahapan penelitian, dataset yang digunakan, proses pengolahan data, arsitektur model, serta metode evaluasi yang digunakan untuk mengukur performa sistem deteksi *AI-generated speech*.

d. Bab 4 HASIL DAN DISKUSI

   Bab ini menyajikan hasil dari proses eksperimen dan pengujian model yang telah dilakukan. Selain itu, bab ini juga membahas analisis terhadap hasil yang diperoleh serta interpretasi terhadap performa model dalam mendeteksi *AI-generated speech*.

e. Bab 5 KESIMPULAN DAN SARAN

   Bab ini berisi kesimpulan yang diperoleh dari hasil penelitian yang telah dilakukan serta saran yang dapat diberikan untuk pengembangan penelitian selanjutnya.