Deteksi Serangan Audio (*Deepfake*) Mengguanakn *Time-Base* dan *Cepstral Domain Feature* dengan *Stacking Classifier*

Zefanya Darma Putri[1], Vera Suryani[2]

[1,2]Fakultas Informatika, Universitas Telkom, Bandung
[1]zefanyaputri@students.telkomuniversity.ac.id,[2]verasuryani@telkomuniversity.ac.id

Abstrak

Audio *deepfake*, atau menipulasi suara, meniru atau mengubah suara asli, dapat digunakan untuk penipuan dan pencemaran nama baik. Tujuan dari penelitian ini adalah untuk meningkatkan akurasi deteksi audio *deepafake* dengan menggunakan metode *stacking classifier* dengan parameter terbaik dati SVM, *random forest* dan *logistic regression* sebagai *base learner* dari *stacking classifier*. Pada penelitian ini digunakan 6 jenis fitur pada audio seperti *Mel-Frequency Cepstral Coefficients (MFCC), Spectral Rolloff, Spectral Contrast, Bandwidth, Zero-Crossing Rate (ZCR)* dan *Root Mean Square (RMS)*. Penulis menggunakan dataset *The Fake or Real*, dataset ini dibuat menggunakan model *text-to-speech* dan dibagi menjadi empat sub-dataset: for-rerec, for-2sec, for-norm dan for-original. Hasil eksperimen sistem yang telah dilakukan memiliki akurasi pengujian 98-99% dan akurasi validasi 97-99%. Penelitian ini membuktikan efektifitas dari pendekatan *stacking classifier* dalam mendeteksi audio *deepfake* asli atau palsu dan telah mengalami peningkatan dari penelitian sebelumnya.

Kata kunci: audio *deepfake*, *stacking classifier, machine learning,* fitur audio, spektral, berbasis waktu

Abstract

Audio deepfake, or the manipulation of sound, it imitates or change the original voice, it can be used for fraud and defamation. The goal of this research is to improve the accuracy of deepfake audio detection by using stacking classifier with the best parameter of SVM, random forest and logistic regression as base learners. This research used 6 types of features such as Mel-Frequency Cepstral Coefficients (MFCC), Spectral Rolloff, Spectral Contrast, Bandwidth, Zero-Crossing Rate (ZCR) and Root Mean Square (RMS). The author used The Fake or Real dataset, this dataset is created using a text-to-speech model and is divided into four sub-datasets: for-rerec, for-2-sec, for-norm, and for-original. The experimental result of this system has 98-99% accuracy of testing and 97-99% accuracy of validation. This research proves the effectiveness of stacking classifier approaches in detecting real or fake audio deepfake and has improved from the previous research.

Keywords: audio deepfake, stacking classifier, machine learning, audio features, spectral, time based

1. Pendahuluan

Latar Belakang

Era digital telah membawa terobosan teknologi yang sangat maju, hal ini telah mengubah kehidupan sehari-hari. Inovasi kecerdasan buatan adalah salah satunya. Meskipun inovasi ini memiliki dampak positif, inovasi ini juga dapat digunakan untuk tindakan kriminal, salah satunya adalah teknologi *deepfake* yang merupakan bentuk paling berbahaya dari kecerdasan buatan. *Deepfake* adalah konten atau materi yang dibuat atau dimanipulasi secara sintetis menggunakan metode kecerdasan buatan (AI) agar tampak seolah-olah nyata [1]. Salah satu jenisnya adalah audio *deepfake*, di mana *deepfake* digunakan untuk membuat atau memanipulasi audio untuk tujuan tertentu. *Deepfake* audio memiliki dampak yang signifikan, termasuk ancaman terhadap keuangan melalui penipuan suara, penyebaran informasi yang salah serta merusak reputasi atau menimbulkan keresahan sosial. Teknologi ini juga mengancam keandalan dalam sistem verifikasi suara dan menimbulkan masalah etika dalam media dan hiburan, seperti pelanggaran hak cipta. Selain itu, *deepfake* audio dapat menyebarkan ketidak percayaan terhadap bukti audio dan komunikasi publik, yang memungkinkan individu untuk menyangkal fakta dengan mengklaim bahwa itu adalah *deepfake* [1]. Hal ini menekankan pentingnya mengembangkan teknologi deteksi pada audio *deepfake*.

Berbagai penelitian telah dilakukan untuk mengidentifikasi solusi dari permasalahan ini. Namun, masing-masing penelitian tersebut memiliki keterbatasan yang perlu diatasi. Salah satu penelitian yang dilakukan oleh [5], diketahui menghasilkan tingkat akurasi antara 70 hingga 98 persen dengan hanya menggunakan fitur MFCC dan *machine learning* akan tetapi model tersebut tidak mampu

melakukan kalkulasi akan varian data pada for-original dataset, dimana hal ini menunjukkan bahwa diperlukan mesin dengan kapabilitas yang lebih optimal, serta penambahan fitur yang bervariasi untuk dilakukan analisis oleh mesin. Kemudian ada penelitian yang dilakukan oleh [6], dengan pernyataan bahwa menggunakan lebih dari satu jenis fitur, tingkat akurasi meningkat hingga menjadi 94% dengan menggunakan MFCC, LFCC dan Chroma- STFT, penelitian tersebut menggunakan mesin yang telah mereka buat sendiri yaitu MFAAN. Penelitian lainnya yang dilakukan oleh [7] menggunakan lebih dari satu fitur seperti MMFCC, spectral rolloff, spectral centroid, spectral contrast, spectral bandwidth dan zero crossing rate. Pada penelitian ini, tingkat akurasi berkisar antara 70-98% dengan model *machine learning* SVM menjadi yang tertinggi. Penelitian sebelumnya yang hanya menggunakan satu fitur cenderung memiliki akurasi yang tidak optimal dibandingkan dengan penelitian yang menggunakan kombinasi fitur lainnya. Tidak hanya pengaruh fitur saja, jika hanya menggunakan *machine learning* tradisional cenderung tidak mampu memberikan hasil yang terbaik.

Untuk mengatasi keterbatasan tersebut, penelitian ini mengusulkan sistem deteksi audio *deepfake* yang lebih akurat dengan memanfaatkan beragam fitur audio serta menerapkan *Stacking Classifier*. *Stacking classifier* meningkatkan performa klasifikasi dengan menggabungkan beberapa model dasar, yaitu *Support Vector Machine* (SVM), *Random Forest*, dan *Logistic Regression*, sehingga masing-masing model dapat menangkap aspek berbeda dari data. SVM efektif dalam menangani ruang fitur berdimensi tinggi, *Random Forest* tangguh terhadap noise dan mampu menangkap hubungan non-linear, sementara *Logistic Regression* memberikan prediksi berbasis probabilitas. Dengan menggabungkan keluaran dari ketiga model melalui meta classifier, *stacking* meningkatkan generalisasi model dan mengurangi kesalahan yang disebabkan oleh kelemahan masing-masing model individu. Pendekatan *ensemble* ini berhasil mengatasi kelemahan yang diamati dalam penelitian sebelumnya [7], terutama dalam menangani variabilitas data, serta mencapai akurasi yang lebih tinggi dibandingkan metode berbasis fitur sebelumnya. Oleh karena itu, penelitian ini menyoroti efektivitas *stacking classifier* dalam meningkatkan deteksi *deepfake*, sekaligus mengatasi tantangan dalam pemilihan fitur dan ketahanan model terhadap variasi data.

Topik dan Batasannya

Berdasarkan latar belakang masalah yang telah diuraikan pada Bab Pendahuluan, penelitian dalam tugas akhir ini berfokus pada pengembangan sistem deteksi audio *deepfake* yang lebih optimal menggunakan *stacking classifier*. Penelitian ini juga mengevaluasi performa sistem yang telah dibangun serta mengeksplorasi strategi peningkatan akurasi dalam deteksi audio deepfake melalui pemanfaatan *feature extraction* dan kombinasi model *machine learning*.

Adapun batasan masalah dalam penelitian ini terletak pada penggunaan dataset yang berasal dari sumber *open source* yaitu Kaggle. Dataset yang digunakan adalah *The Fake or Real*, dalam dataset ini terdiri dari empat jenis dataset yaitu *for-rerec, for-2sec, for-norm,* dan *for-original* yang diambil dengan empat cara berbeda dalam pengambilan audio. Penelitian ini hanya fokus menggunakan satu kombinasi dari beberapa fitur yang digabungkan dan tidak termasuk analisis terhadap konten yang terdapat pada audio.

Tujuan

Penelitian ini bertujuan untuk mengatasi keterbatasan metode sebelumnya dengan mengembangkan sistem pendekatan serta bertujuan untuk melakukan evaluasi performansi deteksi audio *deepfake* dengan memanfaatkan kombinasi fitur domain yang berbeda dan menggunakan *stacking classifier* sebagai algoritma yang digunakan untuk dibandingkan dengan model *machine learning traditional* yang dijadikan sebagai *base learner*.

Organisasi Tulisan

Organisasi penulisan dalam penelitian ini diantaranya adalah bagian ke-1 merupakan pendahuluan dengan sub-bagian berupa latar belakang, topik dan batasan, serta tujuan. Bagian ke-2 akan merupakan studi terkait. Bagian ke-3 akan berisi penjelasan mengenai sistem yang dibangun diawali dengan alur diagram dimulai dari proses ekstraksi fitur, pengolahan data, penentuan *base learner*, pelatihan & evaluasi *stacking classifier*, dan parameter evaluasi. Bagian ke-4 akan menjelaskan mengenai hasil dan analisis pengujian. Bagian ke-5 akan akan membahas kesimpulan.

2. Studi Terkait

Terdapat penelitian terkait yang telah dilakukan dengan menggunakan dataset yang sama maupun berbeda. Penelitian pertama yang dilakukan oleh Ameer Hamza [5] memanfaatkan berbagai mofel *machine learning* untuk mengklasifikasikan *deepfake* pada dataset for-2sec, for-norm dan for-rerec. Pada penelitian yang