

STEGANALISIS UNTUK FILE AUDIO BERFORMAT MP3 DENGAN METODE *LEAST SIGNIFICANT BIT* (LSB) PADA KLASIFIKASI *PRINCIPAL COMPONENT ANALYSIS* (PCA)

Vivin Fauziah Ramadhani ¹⁾, Bambang Hidayat, ²⁾, Azizah. ³⁾

^{1),2),3)} Teknik Telekomunikasi, Universitas Telkom
Jl. Telekomunikasi No. 1, Bandung
Email : vivin_fauziah@ymail.com

Abstrak . Steganalisis merupakan suatu ilmu dan seni untuk mendeteksi pesan tersembunyi dari sebuah objek pesan yang dikirimkan dari pengirim kepada penerima. Munculnya teknik steganografi untuk menyembunyikan pesan tersebut, maka dibutuhkan suatu deteksi yang dapat menganalisis ada tidaknya pesan yang dikirimkan memiliki informasi tersembunyi. Pesan tersembunyi yang telah disisipkan melalui suatu objek secara baik tidak dapat dirasakan oleh panca indra manusia. Oleh karena itu, dibutuhkanlah suatu perangkat yang dapat mendeteksi keberadaan pesan tersembunyi tersebut. Dalam penelitian ini, dibuat suatu perangkat lunak yang mampu mendeteksi keberadaan pesan tersembunyi dalam sebuah file MP3 menggunakan metode statistik. Dalam metode ini digunakan klasifikasi PCA (Principal Component Analysis). PCA merupakan salah satu teknik mengambil atau menarik fitur-fitur dari suatu data pada sebuah skala dengan dimensi tinggi. MP3 menjadi objek yang tidak bisa dilepaskan dari kehidupan seseorang. Selain itu, penyembunyian pesan melalui file berformat MP3 merupakan salah satu format yang baik digunakan. Hal ini dikarenakan file berformat MP3 mempunyai kualitas yang bagus dan tidak mengalami proses cropping sebagaimana pesan disembunyikan dalam sebuah gambar. Penelitian ini diharapkan bisa digunakan bagi pihak-pihak yang berkepentingan dan mempunyai kebutuhan mengenai informasi dan pesan rahasia seperti kepolisian dan Badan Intelijen Negara yang memiliki tingkat keakurasian yang lebih baik dari penelitian sebelumnya.

Kata Kunci : Steganalisis, Steganografi, Least Significant Bit, Audio, MP3.

1. Pendahuluan

Perkembangan teknologi yang sangat pesat memberikan banyak kemudahan komunikasi kepada setiap orang. Internet menjadi sarana untuk saling bertukar pesan dan mendapatkan informasi yang dibutuhkan. Semakin banyaknya pengguna teknologi komunikasi seperti pesan singkat, media sosial, maka semakin berkembang pula jenis komunikasi yang terjadi. Adanya pihak-pihak yang memiliki kepentingan tertentu yang tidak menginginkan pesannya dilihat maupun disadap karena memiliki pesan rahasia mengharuskan mereka untuk menyembunyikan pesannya. Teknik yang menggunakan penyembunyian pesan ini dinamakan steganografi.

Seiring berkembangnya steganografi dalam menyampaikan pesan rahasia, sering dimanfaatkan oleh pihak-pihak tertentu untuk menyampaikan pesan yang berisi kejahatan. Penggunaan teknik steganografi oleh pihak tersebut akan menimbulkan kejahatan yang sulit diketahui oleh pihak berwajib. Hal ini dikarenakan beberapa pesan tertentu mengandung informasi yang sangat penting untuk pihak berwajib untuk diolah dan dianalisis.

Dalam melihat suatu informasi yang mengandung pesan rahasia dibutuhkan suatu teknik yang dapat menganalisa keberadaan pesan tersebut, salah satunya adalah steganalisis. Steganalisis merupakan ilmu dan seni untuk mendeteksi pesan rahasia dalam suatu objek pesan. Dalam perkembangannya steganalisis mulai banyak digunakan seiring berkembangnya steganografi. Steganalisis diimplementasikan dalam berbagai objek pesan seperti gambar, audio, dan video. Melalui teknik steganalisis ini pesan rahasia mampu dideteksi dan diketahui ada tidaknya pesan tersembunyi tersebut.

Dalam penelitian ini, steganalisis akan mendeteksi suatu objek pesan berupa audio berformat MP3. Teknik steganalisis ini belum bisa mendeteksi segala bentuk objek pesan yang disisipi pesan rahasia. Oleh karena itu dibutuhkan teknik steganalisis dengan berbagai metode dan klasifikasi sehingga pesan

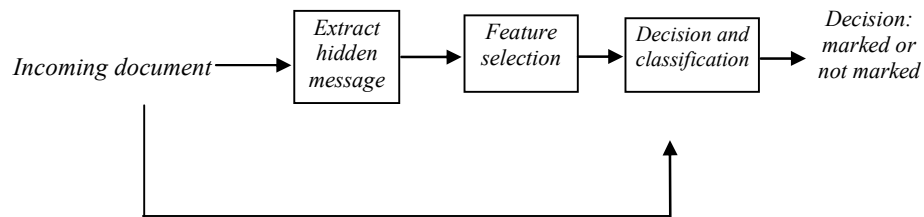
yang dideteksi bisa diketahui keberadaan pesan rahasianya dengan tingkat keakuratan yang tinggi. Dengan tingkat keakuratan yang tinggi, objek tersebut bisa dipastikan memiliki pesan rahasia atau tidak.

Penelitian ini mengacu terhadap referensi-referensi yang terkait dengan penelitian yang telah ada sebelumnya, dimana masing-masing penulis menggunakan metode penelitian yang berbeda tergantung atas permasalahan yang akan dikaji. Pada penelitian [1] dirancang suatu sistem untuk mendeteksi pesan tersembunyi pada *file* mp3 dengan melihat beberapa ciri statistic seperti mean, standar deviasi, skewness, dan kurtosis. Sedangkan penelitian [2] dengan menggunakan format file audio yang sama, ekstraksi ciri yang dipakai adalah MFCC (*Mel-Frequency Cepstrum Coefficients*). Penggunaan beberapa referensi ini akan digunakan untuk membedakan pembahasan yang dibahas penulis dengan penelitian-penelitian yang telah ada sebelumnya.

Pengembangan arah penelitian yang dilakukan dalam Tugas Akhir ini adalah melakukan deteksi pada audio mp3 menggunakan ekstraksi ciri Principal Component Analysis (PCA) untuk mereduksi kekompleksan pada data audio sehingga didapatkan ciri statistic yang sederhana dan mudah untuk dideteksi.

Steganalisis

Steganalisis adalah seperangkat teknik yang bertujuan untuk membedakan antara penutup-benda dan stego-benda. Dalam metode steganalisis yang diusulkan, seperti yang disajikan pada Gambar 1, ide yang mendasari adalah untuk pertama mengisolasi stego-sinyal dengan mengurangi perkiraan sinyal asli dari sinyal tes yang diberikan. Sinyal asli itu sendiri diperkirakan oleh beberapa skema *de-noising* atau dengan algoritma model berbasis. Dalam kasus aditif pesan *embedding*, algoritma denoising menjadi alat yang efektif untuk estimasi asli.



Gambar 1. Diagram blok dari metode steganalisis

Principal Component Analysis (PCA)

PCA adalah suatu transformasi linier yang biasa digunakan pada kompresi data. PCA adalah sebuah teknik statistika yang berguna pada bidang pengenalan, klasifikasi, dan kompresi data citra. PCA juga merupakan teknik yang umum digunakan untuk menarik fitur-fitur dari data pada sebuah skala berdimensi tinggi. Dengan cara mentransformasikan citra ke dalam eigenfaces secara linier, proyeksikan citra ke dalam bentuk skala berdimensi n , yang menampakkan properti dari sampel yang paling jelas sepanjang koordinat. Dengan kata lain, PCA adalah transformasi linear untuk menentukan sistem koordinat yang baru dari dataset. Teknik PCA dapat mengurangi dimensi dari dataset tanpa tidak menghilangkan informasi penting dari dataset.

Least Significant Bit (LSB)

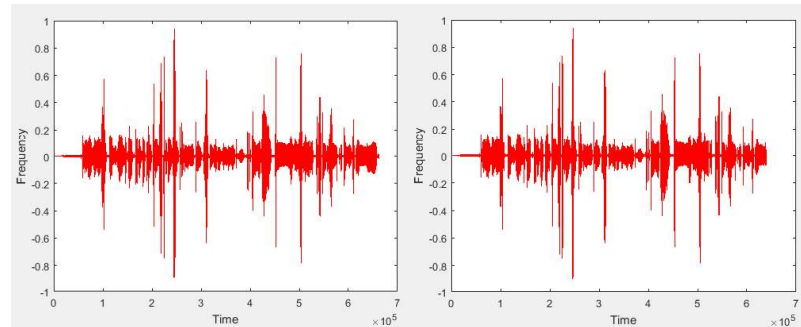
LSB adalah bit yang mempunyai nilai paling rendah, atau bit yang berada pada posisi paling kanan. Penyisipan LSB dilakukan dengan memodifikasi bit terakhir dalam satu *byte* data. Pada teknik steganografi, penyisipan pesan atau informasi rahasia biasanya dilakukan pada bit terakhir suatu objek. Bit yang diganti adalah LSB karena perubahan pada LSB hanya menyebabkan perubahan nilai *byte* satu lebih tinggi atau satu lebih rendah.

2. Pembahasan

Segmentasi dan *Framing*

Segmentasi merupakan proses pengelompokan berdasarkan karakteristik tertentu. Sedangkan, framing berfungsi untuk membagi data menjadi bagian yang sama besar.

Berikut adalah gambar data audio sebelum dan sesudah dilakukan proses segmentasi dan *framing* :



Gambar 2. Sinyal audio mp3 sebelum dan sesudah proses *framing*

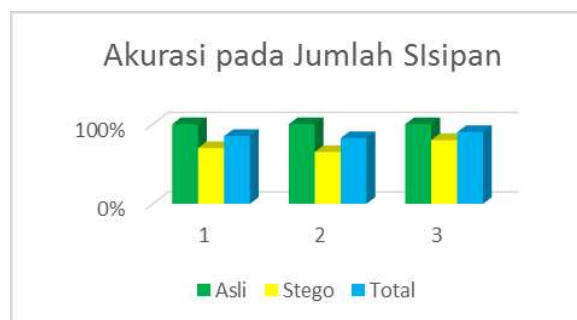
Pengujian Berdasarkan Ukuran Dimensi Objek Pesan yang Disisipkan

Pengujian ini dilakukan untuk melihat pengaruh dan tingkat akurasi pada sistem. Berkas pengujian ada 3 jenis, yaitu audio yang disisipkan dengan objek pesan 20x20, 30x30, dan 40x40. Ketiga jenis berkas tersebut akan dibandingkan dengan nilai pada ekstraksi ciri berupa mean, standar deviasi, *skewness*, dan *kurtosis*. Berdasarkan hasil pengujian pada sistem, berikut perbandingan dan nilai akurasi yang disajikan dalam tabel dibawah ini.

Tabel 1. Akurasi pada beberapa jumlah sisipan

Ciri	Jenis File Audio Mp3	Ukuran Dimensi Objek Pesan yang Disisipkan		
		20x20	25x25	30x30
Mean, Standar Deviasi, Skewness, dan Kurtosis	Asli	100%	100%	100%
	Stego	70%	65%	80%
	Total	85%	82,5%	90%

Data keakurasian sistem akan disajikan juga dalam bentuk grafik berikut ini.



Gambar 3. Diagram akurasi pada jumlah sisipan

Berdasarkan pengujian pada parameter ini dimana diujikan pada 60 data latih dan 60 data uji, didapatlah akurasi sistem 85% pada audio yang disisipkan 20x20 dimensi pesan (*bit*), 82,5% pada audio yang disisipkan 25x25 dimensi pesan, dan 90% untuk audio yang disisipkan pesan 30x30 dimensi pesan. Hal ini menunjukkan bahwa audio yang disisipkan dengan pesan yang ukurannya besar, memiliki tingkat akurasi yang baik dan dapat dideteksi sebagai audio stego. Sedangkan untuk

penyisipan yang lebih kecil, akurasi menunjukkan hasil yang lebih rendah namun masih dikategorikan baik.

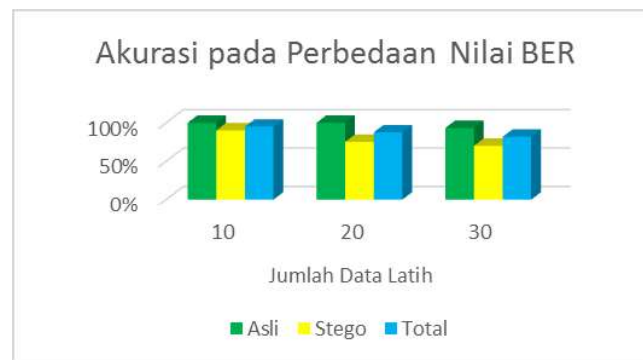
Pengujian Berdasarkan Nilai BER

Pengujian pada parameter ini diberikan nilai BER data audio yang bernilai variatif. Pada tahap pengujian ini dilakukan untuk melihat pengaruh BER terhadap kualitas audio serta batas nilai BER maksimal agar pesan tersembunyi yang disisipkan tidak dapat dideteksi menggunakan metode ini. Nilai BER yang diberikan berkisar antara 0,001-0,5. Pengujian dilakukan dalam beberapa kelompok untuk mengkhususkan nilai BER.

Tabel 2. Akurasi pada beberapa nilai BER

Jenis File Audio MP3	Nilai BER		
	BER < 0,1	0,1 < BER < 0,3	BER > 0,3
Asli	100%	100%	100%
Stego	40%	90%	100%
Total	60%	95%	100%

Berikut ini nilai akurasi yang ditampilkan dalam bentuk diagram untuk mempermudah proses pemahaman.



Gambar 4. Diagram akurasi pada nilai BER

Berdasarkan tabel dan diagram ditinjau dari total akurasi, dapat diambil kesimpulan bahwa nilai BER menunjukkan kualitas audio yang dihasilkan oleh proses steganografi. Nilai BER yang tinggi mengindikasikan audio tersebut memiliki suara yang sedikit tidak mirip dengan aslinya atau terdapat bunyi-bunyi yang tidak diinginkan. Semakin rendah nilai BER, audio stego akan sulit dideteksi, ini dibuktikan dengan simulasi menggunakan nilai BER dibawah 0,1. Untuk nilai BER yang berada pada rentang tertentu atau pada kondisi nilai BER yang tidak terlalu buruk, audio stego dapat dideteksi dengan baik.

Pengujian Berdasarkan Jenis atau *Genre* Musik

Pada tugas akhir ini dilakukan pengujian kedalam beberapa jenis musik yang biasa didengarkan oleh orang lain. Pada tahapan ini, jenis atau *genre* musik yang digunakan sebanyak 5 jenis, yaitu lagu dengan *genre* pop, jazz, reggae, rnb, dan rock. Pengujian terhadap beberapa jenis musik ini dilakukan untuk mengetahui keakurasian sistem terhadap jenis musik dengan format mp3 yang biasa digunakan sebagai format pada lagu. Lagu pada data latih dan uji dibatasi selama 10 detik dengan besar memori atau kapasitas yang sama.

Tabel 3. Akurasi pada jenis atau *genre* musik

Jenis atau <i>genre</i> Musik	Jenis File Audio MP3		
	Asli	Stego	Total
<i>Pop</i>	100%	70%	85%
<i>Rock</i>	90%	80%	85%
<i>Reggae</i>	100%	80%	90%
<i>Jazz</i>	100%	70%	85%

Perbedaan untuk jenis atau *genre* musik juga akan disajikan dalam bentuk diagram berikut ini.



Gambar 5. Diagram akurasi pada perbedaan jenis atau *genre* musik

Berdasarkan tabel dan diagram di atas, dapat disimpulkan bahwa untuk jenis musik rock dan reggae memiliki tingkat akurasi yang tinggi dengan nilai yang sama, begitu juga dengan jenis musik pop dan jazz memiliki tingkat akurasi yang sama yaitu 75%, namun lebih kecil dibandingkan jenis musik rock dan reggae. Akan tetapi pada pengujian audio asli pada rock memiliki tingkat akurasi yang paling kecil dari semua jenis musik.

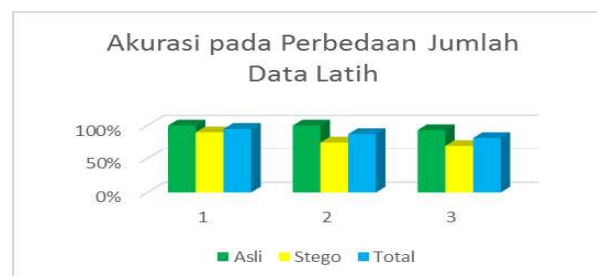
Pengujian Berdasarkan Jumlah Data Latih

Perbedaan jumlah data latih dilakukan dengan berbagai variasi jumlah data. Data latih diujikan dalam tiga kelompok, yaitu dengan jumlah 5 data latih, 10 data latih, dan 20 data latih. Ketiga kategori tersebut akan diukur tingkat akurasinya untuk pada ekstraksi ciri yang telah ditentukan sebelumnya.

Tabel 4. Akurasi pada perbedaan jumlah data latih

Jenis File Audio MP3	Jumlah Data Latih		
	10	20	30
Asli	100%	100%	93,33%
Stego	90%	75%	70%
Total	95%	87,5%	81,665%

Perbedaan untuk jenis atau *genre* musik juga akan disajikan dalam bentuk diagram berikut ini.



Gambar 6. Akurasi pada perbedaan jumlah data latih

Berdasarkan tabel dan diagram dari parameter tersebut, maka jumlah data latih menunjukkan bahwa data latih yang banyak memberikan banyak ciri dan kemiripan yang sama, sehingga audio stego yang memiliki selisih nilai yang sangat sedikit akan dideteksi sebagai audio asli. Akan tetapi, hal ini memperlihatkan bahwa banyaknya jumlah data latih akan membantu sistem untuk mendapatkan ciri yang mirip.

3. Simpulan

1. Steganalisis pada file audio berformat *mp3 dilakukan dengan membandingkan nilai ciri atau karakteristiknya, yaitu nilai mean, standar deviasi, *skewness*, dan *kurtosis*.
2. SVM (Support Vector Machine) digunakan sebagai klasifikasi pada proses steganalisis untuk penelitian ini untuk mengetahui keberadaan pesan yang telah disisipkan pesan lainnya.
3. Beberapa parameter dapat mempengaruhi perbedaan akurasi menggunakan klasifikasi yang sama. Akurasi pada sistem didapatkan dengan rentang 50%-90%.
4. Sistem dapat membedakan audio asli dan audio stego dengan baik walaupun hasil yang didapatkan belum sepenuhnya maksimal. Akan tetapi, secara pendengaran manusia perbedaan kedua audio tersebut tidak dapat didengar.

Daftar Pustaka

- [1]. Ardhyana, A. S., & Juarna, A. (t.thn.). Aplikasi Steganografi pada Mp3 Menggunakan Teknik LSB.
- [2]. Muzaqiyah, M. (2014). Steganalisis Pada Audio Mp3 Dengan Metode LSB (Least Significant Bit) Matching . *Universitas Telkom*.
- [3]. Tarigan, C. D. (2014). Steganografi Pada File Audio Mp3 Untuk Pengamanan Data Menggunakan Metode Least Significant Bit (LSB). *Pelita Informatika Budi Dharma*.
- [4]. Altun, O., Sarma, G., Celik, M., Sterling, M., Titlebaum, E., & Bocko, M. (2005). Morphological Steganalysis Of Audio Signals And The Principle Of Diminishing Marginal Distortions. *IEEE*.
- [5]. Anonym. (t.thn.). Principal Component Analysis (PCA). Dalam *STAT 505*. United State: The Pennsylvania State University.
- [6]. Anonym. (t.thn.). Steganography and Steganalysis Method. *Chapter 2*.
- [7]. Badr, S. I., Salama, G. I., Selim, G. M., & Khalil, A. H. (2014). A Review on Steganalysis Technique: From Image Format Point of View. *International Journal of Computer Applications*.
- [8]. Ozer, H., Avcibas, I., Sankur, B., & Memon, N. (t.thn.). Steganalysis of Audio Based on Audio Quality Metrics.
- [9]. Prasetyo, G. B. (2013). Kompresi File Audio Wave Menggunakan Algoritma Huffman Shith Coding. *Universitas Brawijaya*.