PROKASDADIK: Prosiding Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan
Vol. 4, No. 2, (2026) E-ISSN: 3063-5845

Analisis Deteksi Stego pada Citra Digital

I Made Mustika Kerta Astawa[1], Efza Nur Agastya[2]
[1,2] Program Studi Teknik Informatika S-2, Program Pascasarjana, Universitas Pamulang, Tangerang Selatan, Banten
e-mail: made.mustika19@gmail.com[1], efzanuragastya@gmail.com[2]

Abstrak*— Perkembangan teknologi informasi telah meningkatkan kemudahan pertukaran data digital, namun juga membuka peluang penyalahgunaan teknik steganografi untuk menyembunyikan informasi pada citra digital. Teknik ini berpotensi dimanfaatkan dalam berbagai kejahatan siber, seperti penyebaran malware, komunikasi tersembunyi, dan pencurian informasi. Penelitian ini bertujuan menganalisis kinerja algoritma *Support Vector Machine* (SVM) dan *Random Forest* dalam mendeteksi citra stego berbasis penyisipan *Least Significant Bit* (LSB). Proses penelitian diawali dengan pengumpulan dataset yang terdiri atas citra normal dan citra stego, kemudian dilakukan ekstraksi fitur menggunakan InceptionV3 sebagai *image embedding* untuk menghasilkan representasi fitur berdimensi tinggi. Selanjutnya, proses klasifikasi dilakukan pada platform *Orange Data Mining* menggunakan algoritma SVM dan *Random Forest*. Evaluasi model dilakukan berdasarkan metrik *Area Under Curve* (AUC), *Classification Accuracy* (CA), *F1-Score*, *Precision*, *Recall*, *Matthews Correlation Coefficient* (MCC), serta analisis *Confusion Matrix*. Hasil penelitian menunjukkan bahwa *Random Forest* memberikan performa terbaik dengan nilai AUC sebesar 0,917, akurasi 91,7%, *F1-Score* 0,916, *Precision* 0,929, *Recall* 0,917, dan MCC 0,845. Sebaliknya, SVM hanya memperoleh AUC 0,139 dan akurasi 58,3%, yang menunjukkan ketidakmampuan model dalam membedakan citra normal dan citra stego secara optimal. Analisis *Confusion Matrix* memperlihatkan bahwa *Random Forest* mampu meminimalkan kesalahan *false alarm* dengan tingkat presisi deteksi stego yang sangat tinggi, sedangkan SVM mengalami bias klasifikasi akibat keterbatasannya dalam memodelkan pola non-linear yang dihasilkan oleh manipulasi bit LSB. Dengan demikian, *Random Forest* terbukti lebih efektif dan andal untuk mendukung proses steganalisis pada citra digital serta berpotensi diterapkan dalam investigasi forensik digital dan peningkatan keamanan informasi.

***Kata Kunci*— Steganalisis, Citra Digital, Least Significant Bit (LSB), InceptionV3, Random Forest.**

I. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi (TIK) telah meningkatkan kemudahan dalam pertukaran data digital, tetapi juga memunculkan berbagai ancaman keamanan siber. Termasuk kejahatan siber yang terjadi akhir-akhir ini dimana kasus scam memanfaatkan file gambar yang didalamnya terdapat malware. Salah satu ancaman yang semakin berkembang adalah penyalahgunaan teknik steganografi, yaitu metode menyembunyikan pesan atau file rahasia ke dalam media digital, seperti citra, sehingga sulit dideteksi oleh pengamatan visual. Teknik ini dapat dimanfaatkan untuk berbagai tindak kejahatan, seperti penyebaran malware melalui file gambar, komunikasi tersembunyi antar pelaku kejahatan siber, pencurian informasi, hingga distribusi konten ilegal. Kondisi tersebut menuntut adanya metode steganalisis yang mampu mendeteksi keberadaan pesan tersembunyi pada citra digital secara akurat dan efisien.

Penelitian ini mengimplementasikan teknik deteksi steganografi pada citra digital menggunakan InceptionV3 sebagai metode image embedding untuk mengekstraksi fitur visual yang merepresentasikan karakteristik citra ke dalam vektor berdimensi tinggi. Fitur yang dihasilkan kemudian diklasifikasikan menggunakan algoritma Support Vector Machine (SVM) dan Random Forest yang diimplementasikan pada Orange Data Mining. Pendekatan ini bertujuan membandingkan kinerja kedua algoritma dalam membedakan citra stego dan citra normal, sehingga diperoleh model deteksi yang lebih efektif untuk mendukung proses forensik digital dan meningkatkan keamanan informasi.

II. METODE PENELITIAN

Penelitian ini dilakukan melalui beberapa tahapan terstruktur yang diimplementasikan pada lingkungan *Orange Data Mining*, sebuah *platform* analisis berbasis visual workflows. Diagram alir metodologi eksperimen digambarkan pada metodologi di bawah ini:

2.1. Pengumpulan Data & Pre-processing

Dataset terdiri dari dua kelas citra digital dengan jumlah berimbang, yaitu:

1) Citra Asli (*Normal/Clean Image*): Citra digital standar tanpa modifikasi.
2) Citra Stego (*Stego Image*): Citra asli yang telah disisipi pesan teks rahasia pada bit LSB-nya, sehingga memiliki perbedaan mikroskopis akibat fluktuasi *noise* spasial.

PROKASDADIK: Prosiding Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan
Vol. 4, No. 2, (2026) E-ISSN: 3063-5845

Sebelum dilakukan pemodelan, dilakukan ekstraksi fitur berbasis nilai residu piksel untuk menangkap korelasi spasial lokal dan deviasi distribusi warna yang disebabkan oleh manipulasi bit LSB.

Sebelum dilakukan pemodelan, dilakukan ekstraksi fitur berbasis nilai residu piksel untuk menangkap korelasi spasial lokal dan deviasi distribusi warna yang disebabkan oleh manipulasi bit LSB dengan mekanisme image embadding menggunakan algoritma Inception V3. *Image embedding* dilakukan menggunakan InceptionV3 untuk mengekstraksi fitur visual secara otomatis dari citra digital. Model ini mengonversi setiap citra menjadi vektor fitur berdimensi tinggi yang merepresentasikan karakteristik visual penting, sehingga perubahan kecil akibat penyisipan pesan steganografi dapat terdeteksi dan digunakan sebagai masukan pada proses klasifikasi citra stego dan normal.

2.2. Konfigurasi Model Klasifikasi

Dua arsitektur pembelajaran mesin dilatih menggunakan data fitur hasil ekstraksi:

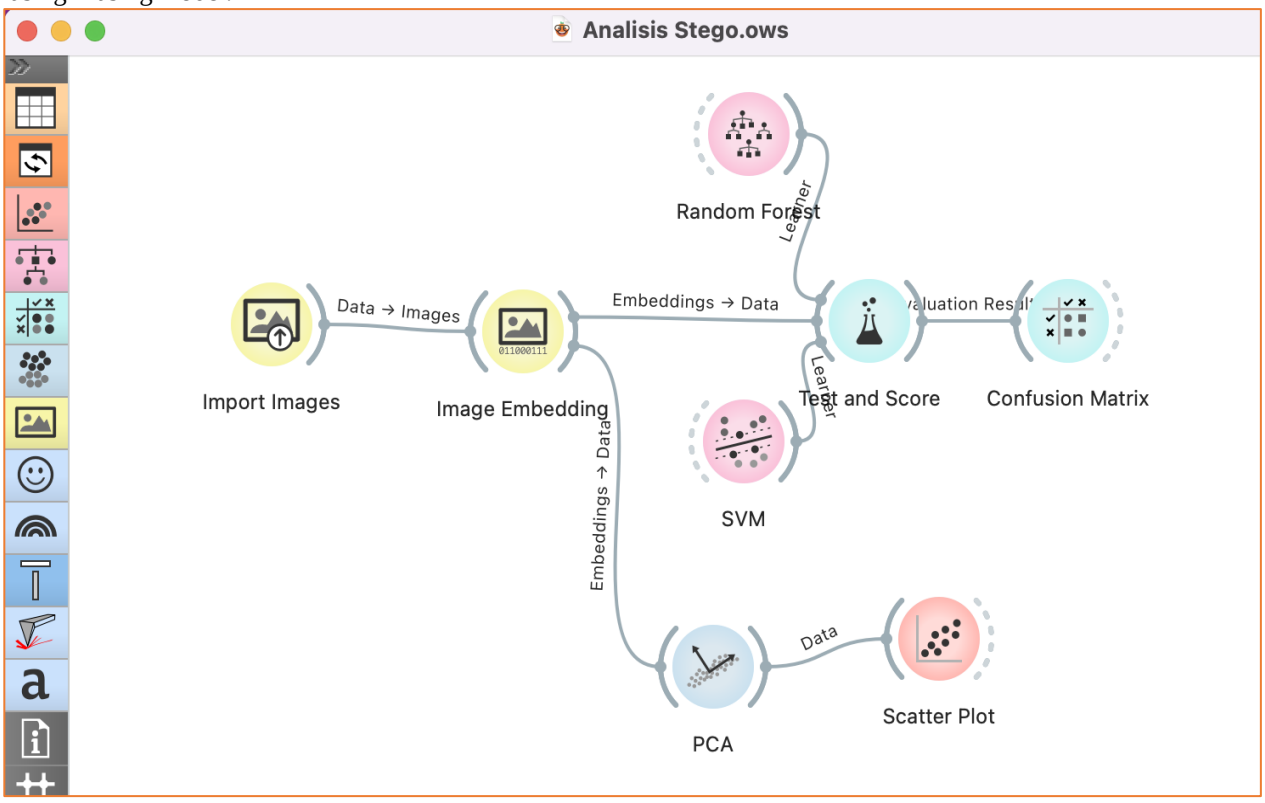
1) *Support Vector Machine* (SVM): Dikonfigurasi untuk memisahkan batas kelas *Clean* dan *Stego* menggunakan fungsi pemisah ruang vektor.
2) *Random Forest*: Dikonfigurasi sebagai pengklasifikasi berbasis himpunan pohon keputusan, di mana penentuan hasil akhir klasifikasi (Normal vs Stego) diambil melalui mekanisme *voting* mayoritas dari seluruh pohon yang terbentuk.

2.3. Metode Evaluasi

Model diuji menggunakan skenario pengujian prediktif (uji validasi). Kinerja model dinilai secara komparatif melalui enam metrik evaluasi utama:

1) *Area Under ROC* (AUC): Mengukur kemampuan diskriminasi model pada berbagai ambang batas.
2) *Classification Accuracy* (CA): Rasio prediksi benar terhadap keseluruhan data.
3) *F1-Score*: Rata-rata harmonik antara *Precision* dan *Recall*.
4) *Precision*: Tingkat keakuratan model dalam mendeteksi kelas positif dari seluruh prediksi positif.
5) *Recall*: Kemampuan model dalam menemukan kembali seluruh sampel kelas positif asli.
6) *Matthews Correlation Coefficient* (MCC): Tolok ukur kualitas klasifikasi biner dengan rentang -1 hingga +1, yang memberikan nilai tinggi hanya jika prediksi pada semua kuadran *Confusion Matrix* memberikan hasil yang baik.

Analisis kesalahan (*mis-detect* dan *false alarm*) diperdalam dengan membedah proporsi data pada bagan *Confusion Matrix* untuk masing-masing model.



Gambar 1. Implementasi Deteksi Stego dengan aplikasi Orange

PROKASDADIK: Prosiding Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan
Vol. 4, No. 2, (2026) E-ISSN: 3063-5845

III. PEMBAHASAN

Berdasarkan eksperimen yang dilakukan pada *Orange Data Mining*, hasil komparasi kuantitatif dari metrik evaluasi kedua model tersaji pada Tabel 1 di bawah ini:

Tabel 1. Perbandingan Metrik Evaluasi Model SVM dan Random Forest

Model Klasifikasi	AUC	CA	F1-Score	Precision	Recall	MCC
SVM	0.139	0.583	0.496	0.773	0.583	0.302
Random Forest	0.917	0.917	0.916	0.929	0.917	0.845

3.1. Analisis Kinerja Random Forest

Algoritma *Random Forest* menunjukkan performa yang sangat superior dalam studi kasus ini. Dengan nilai **AUC sebesar 0,917**, model ini terbukti memiliki kemampuan diskriminasi yang sangat andal untuk memisahkan citra normal dari citra stego. Nilai akurasi klasifikasi (CA = 0,917 atau 91,7%) mengindikasikan bahwa hampir seluruh sampel data berhasil dikelompokkan ke kelas yang benar. Keseimbangan model dalam mendeteksi kedua kelas juga tercermin dari nilai F1-Score (0,916) dan nilai MCC yang tinggi (0,845), menunjukkan bahwa performa tinggi ini bukan merupakan hasil dari bias tebakan acak.

3.2. Analisis Kegagalan Kinerja SVM

Berbanding terbalik dengan Random Forest, model *Support Vector Machine* (SVM) mengalami degradasi performa yang sangat signifikan. Skor AUC sebesar 0,139 berada jauh di bawah nilai ambang acak standar (0,5). Skor AUC yang mendekati 0 mengindikasikan bahwa model secara sistematis menghasilkan klasifikasi yang keliru atau terbalik dalam membedakan pola. Akurasi model yang hanya menyentuh angka 58,3% mempertegas bahwa performa SVM pada pengujian ini tidak jauh berbeda dengan metode spekulatif seperti pelemparan koin (kebetulan). Nilai F1-Score yang rendah (0,496) mengonfirmasi ketidakseimbangan klasifikasi yang masif.

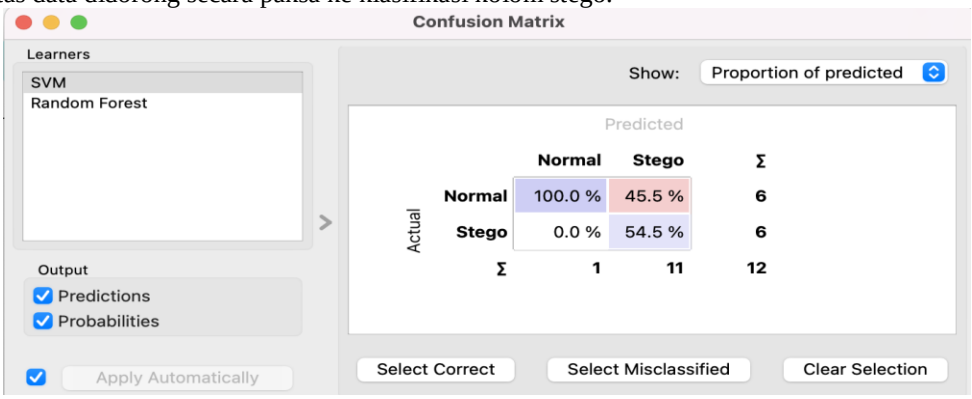
IV. ANALISIS

Untuk memahami anomali perilaku dan jenis kesalahan dari kedua model secara lebih mendalam, dilakukan analisis berbasis proporsi prediksi (*Proportion of predicted*) melalui *Confusion Matrix* yang dihasilkan oleh sistem.

4.1. Visualisasi Proporsi Prediksi SVM

Pada bagan *Confusion Matrix* untuk SVM, terlihat pola persebaran data sebagai berikut:

1) Kolom Prediksi Stego: Dari total sampel yang diprediksi oleh SVM sebagai citra "Stego", sebanyak 54,5% merupakan data Stego aktual, sedangkan 45,5% sisanya adalah citra Normal. Hal ini mengindikasikan adanya tingkat *False Alarm* (Positif Palsu) yang sangat tinggi; model sangat sering salah menuduh citra bersih sebagai citra stego.
2) Kolom Prediksi Normal: Skor 100% tercatat pada baris Normal di kolom ini. Namun, angka ini menipu karena jumlah akumulasi sampel (Σ) yang diprediksi ke arah kolom Normal sangat kecil (hanya 1 sampel). Hal ini membuktikan adanya fenomena *prediction bias*, di mana fungsi *hyperplane* SVM gagal mengonstruksi garis batas pemisah linear yang adil, sehingga mayoritas data didorong secara paksa ke klasifikasi kolom stego.

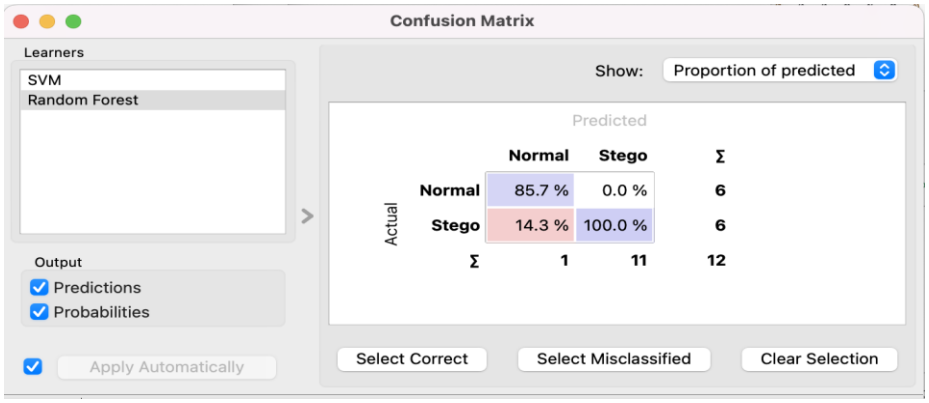


Gambar 2. Analisis Proporsi Prediksi SVM

4.2. Visualisasi Proporsi Prediksi Random Forest

Sebaliknya, *Confusion Matrix* milik *Random Forest* memaparkan hasil yang jauh lebih presisi:

1) Kolom Prediksi Stego: Menghasilkan tingkat kemurnian sempurna sebesar 100,0%. Artinya, model ini memiliki tingkat *Precision* yang mutlak pada kelas stego; tidak ada satu pun citra normal yang salah dideteksi sebagai stego. Jika *Random Forest* memberikan vonis bahwa sebuah citra adalah "Stego", maka tingkat kepastian kebenaran prediksi tersebut bernilai mutlak 100%.
2) Kolom Prediksi Normal: Tercatat nilai sebesar 85,7% untuk kesesuaian prediksi Normal aktual. Namun, terdapat celah kebocoran berupa *Miss-detect* (Negatif Palsu) sebesar 14,3%, di mana sistem mengidentifikasi citra yang sebenarnya mengandung pesan stego sebagai citra normal. Hal ini menunjukkan bahwa ada sebagian kecil variasi biner penyisipan LSB yang sangat halus sehingga lolos dari deteksi percabangan *decision tree*.



Gambar 3. Analisis Proporsi Prediksi Random Forest

4.3. Interpretasi Teoretis Perbedaan Algoritma

1) Kegagalan SVM dan keberhasilan *Random Forest* dalam mengekstrak parameter steganografi LSB dipengaruhi oleh struktur data spasial citra digital. Modifikasi bit pada LSB menghasilkan distorsi berupa *high-frequency noise* yang bersifat non-linear dan tersebar secara acak di seluruh matriks piksel citra.
2) SVM (terutama jika menggunakan kernel linear standar) mencoba menarik garis lurus fiktif untuk membagi ruang dimensi tinggi, suatu hal yang mustahil dilakukan secara akurat pada persebaran derau stego yang kompleks dan samar. Di sisi lain, *Random Forest* bekerja dengan melakukan sub-sampling fitur dan membangun ratusan pohon keputusan independen secara acak (*bootstrap aggregating*). Karakteristik *ensemble* ini memungkinkan *Random Forest* mendeteksi interaksi fitur non-linear yang sangat rumit dan mengisolasi anomali fluktuasi nilai piksel akibat steganografi LSB dengan jauh lebih efisien tanpa mengalami kegagalan *overfitting* atau bias kelas.

V. KESIMPULAN

Berdasarkan hasil eksperimen dan analisis mendalam yang telah dilakukan untuk deteksi stego pada citra digital, dapat disimpulkan bahwa algoritma Random Forest jauh lebih superior dan efektif dibandingkan dengan algoritma *Support Vector Machine* (SVM) dalam mendeteksi keberadaan steganografi LSB pada citra digital. *Random Forest* berhasil membukukan akurasi (CA) sebesar 91,7% dan tingkat presisi deteksi stego mutlak sebesar 100%, menjadikannya model yang sangat tangguh untuk meminimalkan kesalahan salah tuduh (*false alarm*) dalam investigasi forensik digital. Sebaliknya, SVM dinilai gagal memenuhi kualifikasi sistem steganalisis yang andal karena performanya yang buruk (CA 58,3% dan AUC 0,139) akibat ketidakmampuan model menangani kompleksitas fitur derau non-linear secara linear.

Untuk penelitian selanjutnya, disarankan melakukan optimasi penyetelan hiperparameter (*hyperparameter tuning*) pada fungsi kernel SVM (seperti menggunakan RBF atau kernel polinomial) untuk melihat apakah terdapat peningkatan performa. Selain itu, pengujian perlu diperluas menggunakan variasi *payload* kapasitas penyisipan pesan yang berbeda guna menguji batas sensitivitas terendah dari algoritma *Random Forest*..

DAFTAR PUSTAKA

[1] Fridrich, J. (2009). *Steganography in Digital Media: Principles, Algorithms, and Applications*. Cambridge University Press.
[2] Provos, N., & Honeyman, P. (2003). Hide and seek: An introduction to steganography. *IEEE Security & Privacy*, 1(3), 32-44.

[3] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
[4] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.
[5] Demidova, L., & Sokolova, Y. (2021). Using the Orange Data Mining Toolkit for Data Analysis in Classification Problems. *Journal of Physics: Conference Series*, 1828(1), 012015.
[6] Subhedar, M. S., & Mankar, V. H. (2014). Current status and key issues in image steganography: A survey. *Computers & Electrical Engineering*, 41, 151-173.