Journal of Information Technology and Computer Science (INTECOMS)
Volume 8 Nomor 3, Tahun 2025
e-ISSN : 2614-1574
p-ISSN : 2621-3249

APPLICATION OF DATA SCIENCE TO PREDICTE INVALID TRANSACTIONS IN BLOCKCHAIN BASED VOTING SYSTEMS

PENERAPAN DATA SCIENCE UNTUK MEMPREDIKSI TRANSAKSI TIDAK VALID PADA SISTEM VOTING BERBASIS BLOCKCHAIN

Daniel Simamora[1], Yusuf Deardo Purba[2], Kristianos Zamili[3], Finky Notavianus[4], Maya Sofhia[5]
Universitas Prima Indonesia[1,2,3,4,5]
danielsimamora830@gmail.com[1], yusufdeardo@gmail.com[2], kristianozamili@gmail.com[3], finkywaruwu@gmail.com[4], mayasofhia@unprimdn.ac.id[5]

ABSTRACT

Security in blockchain-based electronic voting systems is a major concern in the digital era. This study aims to build a prediction model for invalid transactions in blockchain-based e-voting systems using a data science approach. The two algorithms used are Random Forest and Support Vector Machine (SVM). The dataset used contains more than 500,000 blockchain transaction entries, with a resampling process using SMOTE to overcome the class imbalance between valid and invalid transactions. The initial model shows that both Random Forest and SVM fail to recognize invalid transactions, with a recall value of 0.00. After parameter tuning and SMOTE application, the model performance increases significantly. The SVM model produces the highest recall in the minority class of 57%, although the total accuracy decreases. In contrast, Random Forest shows a recall of 42%, with higher accuracy. These results indicate that the machine learning approach is able to increase the sensitivity of detection of anomalous transactions in e-voting systems, although there is still a trade-off in precision and general accuracy. This study provides an initial basis for the development of a security support system for electronic voting with the help of data-based predictions.
Keywords: *E-Voting, Blockchain, Random Forest, SVM, Data Mining, Data Imbalance, System Security*

ABSTRAK

Keamanan pada sistem pemungutan suara elektronik berbasis blockchain menjadi perhatian utama dalam era digital. Penelitian ini bertujuan untuk membangun model prediksi terhadap transaksi tidak valid pada sistem e-voting berbasis blockchain menggunakan pendekatan data science. Dua algoritma yang digunakan adalah Random Forest dan Support Vector Machine (SVM). Dataset yang digunakan berisi lebih dari 500.000 entri transaksi blockchain, dengan proses *resampling* menggunakan SMOTE untuk mengatasi ketidakseimbangan kelas antara transaksi valid dan tidak valid. Model awal menunjukkan bahwa baik Random Forest maupun SVM gagal mengenali transaksi tidak valid, dengan nilai recall sebesar 0.00. Setelah dilakukan tuning parameter dan penerapan SMOTE, performa model meningkat secara signifikan. Model SVM menghasilkan recall tertinggi pada kelas minoritas sebesar 57%, meskipun akurasi total menurun. Sebaliknya, Random Forest menunjukkan recall sebesar 42%, dengan akurasi yang lebih tinggi. Hasil ini menunjukkan bahwa pendekatan machine learning mampu meningkatkan sensitivitas deteksi terhadap transaksi anomali dalam sistem e-voting, meskipun masih terdapat trade-off terhadap precision dan akurasi umum. Penelitian ini memberikan dasar awal untuk pengembangan sistem pendukung keamanan pada voting elektronik dengan bantuan prediksi berbasis data.

Kata Kunci: E-Voting, Blockchain, Random Forest, SVM, Data Mining, Ketidakseimbangan Data, Keamanan Sistem

PENDAHULUAN

Dalam era digital saat ini, teknologi blockchain telah menjadi inovasi signifikan yang menawarkan transparansi, desentralisasi, dan keamanan dalam berbagai aplikasi, termasuk sistem pemungutan suara elektronik (e-voting). Di Indonesia, penerapan e-voting berbasis blockchain dianggap sebagai solusi potensial untuk meningkatkan integritas dan kepercayaan dalam proses pengambilan keputusan yang melibatkan banyak pihak, seperti pemilihan ketua organisasi, rapat umum pemegang saham, hingga kegiatan voting dalam komunitas digital. Namun, tantangan terkait keamanan data dan privasi pemilih tetap menjadi perhatian utama.

Blockchain dengan kemampuan menciptakan catatan transaksi yang tidak dapat diubah dan didistribusikan secara desentralisasi menjanjikan peningkatan

transparansi dalam sistem voting. Hal ini sangat penting untuk menjawab kekhawatiran masyarakat tentang potensi kecurangan dalam proses tradisional. Namun, seperti teknologi lainnya, blockchain juga memiliki risiko dan kerentanan tertentu. Penelitian oleh Ningrum (2023) menunjukkan bahwa pengembangan sistem keamanan data berbasis blockchain untuk aplikasi voting memerlukan desain yang cermat, algoritma kriptografi yang kuat, dan mekanisme konsensus yang andal. Studi ini menekankan bahwa meskipun blockchain memberikan solusi keamanan yang menjanjikan, implementasinya memerlukan evaluasi mendalam terkait ancaman yang mungkin muncul.

Selain itu, studi lain yang dilakukan di Indonesia juga menunjukkan bahwa penerapan e-voting berbasis blockchain dapat memberikan berbagai manfaat, seperti perlindungan terhadap integritas suara, transparansi proses voting, dan pengurangan biaya operasional. Misalnya, penelitian oleh Ajib Susanto (2020) dengan judul "Implementation of Smart Contracts Ethereum Blockchain in Web-Based Electronic Voting (e-voting)" menyoroti bagaimana penerapan smart contract dalam jaringan blockchain publik dapat meningkatkan efisiensi proses voting, dengan setiap suara terekam secara otomatis dan diverifikasi melalui mekanisme konsensus.

Untuk mendukung evaluasi keamanan sistem blockchain dalam konteks voting, teknologi data science memainkan peran penting. Teknik-teknik data science seperti algoritma Random Forest dan Support Vector Machine (SVM) dapat digunakan

untuk menganalisis data transaksi blockchain, mendeteksi anomali, dan mengidentifikasi potensi ancaman keamanan. Random Forest dikenal sebagai metode pembelajaran ensemble yang efektif dalam menangani data yang kompleks dan besar, sementara SVM memiliki kemampuan untuk memisahkan data dengan margin maksimal, sehingga sangat cocok untuk mendeteksi pola-pola yang mencurigakan dalam data blockchain.

Penggunaan algoritma ini memungkinkan evaluasi sistem keamanan blockchain secara lebih mendalam dan terukur. Dengan menerapkan Random Forest dan SVM, penelitian ini dapat membantu mengidentifikasi kerentanan dalam sistem voting berbasis blockchain serta memberikan rekomendasi perbaikan untuk mengurangi risiko. Hal ini relevan mengingat tantangan teknis seperti kebutuhan akan infrastruktur jaringan yang memadai, kemampuan teknologi untuk menangani skala besar dalam berbagai jenis voting, serta risiko terhadap privasi pemilih yang menjadi isu penting di Indonesia.

Di sisi lain, tantangan non-teknis mencakup kesiapan regulasi, adopsi teknologi oleh masyarakat, dan pendidikan tentang keamanan teknologi blockchain itu sendiri. Oleh karena itu, kebutuhan untuk menilai keamanan sistem blockchain dalam konteks voting menjadi semakin mendesak, mengingat pentingnya memastikan kepercayaan masyarakat terhadap proses pengambilan keputusan yang adil dan transparan.

Dengan memahami tantangan dan peluang yang ada, penelitian ini diharapkan dapat memberikan kontribusi signifikan terhadap pengembangan teknologi voting berbasis blockchain yang lebih baik di Indonesia. Integrasi data science dan blockchain dalam penelitian ini menawarkan pendekatan baru untuk membangun sistem voting elektronik yang aman, transparan, dan terpercaya di berbagai sektor.

Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, rumusan masalah yang akan dibahas dalam penelitian ini meliputi:

1. Bagaimana tingkat keamanan yang dapat dicapai oleh sistem voting berbasis blockchain dalam menghadapi

potensi ancaman keamanan?
2. Bagaimana algoritma data science, seperti Random Forest dan Support Vector Machine (SVM), dapat digunakan untuk mengevaluasi dan meningkatkan keamanan sistem voting berbasis blockchain?

Tujuan Dan Manfaat

Tujuan Penelitian

Tujuan yang diharapkan dari penelitian ini yaitu:

1. Penelitian ini bertujuan untuk mengukur sejauh mana sistem voting berbasis blockchain dapat melindungi data pemilih dan memastikan integritas hasil pemilihan.
2. Penelitian ini akan menerapkan algoritma seperti Random Forest dan SVM untuk menganalisis data transaksi blockchain dan mendeteksi potensi ancaman terhadap keamanan sistem.
3. Berdasarkan evaluasi dan analisis, penelitian ini akan memberikan rekomendasi untuk meningkatkan keamanan dan keandalan sistem voting berbasis blockchain.
4. Penelitian ini juga bertujuan mengembangkan sistem simulasi untuk memvisualisasikan dan menguji bagaimana blockchain dapat diterapkan dalam sistem voting elektronik.

Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini yaitu:

1 Penelitian ini membantu meningkatkan sistem voting berbasis blockchain dengan mengidentifikasi ancaman dan memberi solusi untuk mengurangi risikonya.
2 Hasil penelitian dapat memberikan solusi untuk mengatasi masalah yang ada dalam sistem voting berbasis blockchain.
3 Penelitian ini dapat meningkatkan transparansi dan kepercayaan masyarakat terhadap sistem voting berbasis blockchain.

Batasan Masalah

Berikut ini akan diuraikan batasan masalah pada penelitian ini agar penelitian ini berjalan sesuai dengan perencanaan yaitu:

1. Penelitian ini hanya akan fokus pada aspek keamanan dalam sistem voting berbasis blockchain, dengan tujuan untuk mengidentifikasi potensi ancaman dan kerentanan yang ada dalam sistem tersebut
2. Penelitian ini terbatas pada penggunaan algoritma data science Random Forest dan Support Vector Machine (SVM) untuk menganalisis dan mendeteksi anomali dalam data blockchain yang berkaitan dengan sistem voting.
3. Penelitian ini hanya akan mengidentifikasi dan menganalisis kerentanan yang ada pada sistem blockchain yang digunakan dalam pemilihan elektronik, tanpa melibatkan sistem blockchain di luar konteks voting.

Keterbaruan Penelitian

Berikut adalah contoh keterbaruan penelitian mengenai penggunaan data science untuk menilai keamanan blockchain pada sistem voting elektronik berbasis blockchain:

1. Dalam penelitian Setiawan Restu Aji & Wahyuningdiah Trisari Harsanti Putri. (2023), yang berjudul " Implementasi Teknologi Blockchain dalam Aplikasi E-Voting Berbasis Mobile", dijelaskan bahwa penerapan blockchain berbasis Ethereum dalam sistem e-voting berbasis web dapat meningkatkan transparansi dan keamanan. Penelitian ini menemukan bahwa penggunaan blockchain mengurangi risiko manipulasi suara dan meningkatkan integritas pemilihan.
2. Dalam penelitian oleh A. I. Khan et al. (2025) yang berjudul "AI and Blockchain in Cybersecurity: A Sustainable Approach to Protecting Digital Assets", dibahas bagaimana teknologi blockchain dapat digunakan

untuk meningkatkan keamanan data, termasuk dalam konteks sistem seperti e- voting. Penelitian ini menyoroti bahwa blockchain mampu menjamin integritas dan transparansi data, serta mengurangi risiko manipulasi dalam proses digital.

3. Menurut Prabakar dan Kanchana (2023), dalam penelitiannya yang berjudul "E-Voting Based Blockchain Mechanism Using Feature Selection Based Machine Learning", dikaji bagaimana sistem pemungutan suara elektronik berbasis blockchain dapat diperkuat dengan algoritma machine learning untuk memilih fitur keamanan secara optimal. Penelitian ini menunjukkan bahwa integrasi machine learning dalam proses verifikasi suara dan autentikasi berbasis QR-code mampu meningkatkan keamanan dan efisiensi pemilihan secara digital..
4. Dalam penelitian oleh Hasan dkk. (2024), yang berjudul "Detecting Anomalies in Blockchain Transactions using Machine Learning Classifiers and Explainability Analysis", algoritma machine learning digunakan untuk menganalisis transaksi blockchain guna mendeteksi anomali dan potensi ancaman. Meskipun tidak secara langsung pada sistem e-voting, hasil penelitian ini menunjukkan bahwa pendekatan data science berbasis klasifikasi dan analisis penjelas dapat diterapkan untuk meningkatkan keamanan dan transparansi dalam sistem berbasis blockchain, termasuk voting digital.
5. Menurut Cholevas dkk. (2024), dalam studinya yang berjudul "Anomaly Detection in Blockchain Networks Using Unsupervised Learning: A Survey", dijelaskan berbagai metode unsupervised learning seperti clustering dan density-based analysis dalam mendeteksi aktivitas tidak wajar di jaringan blockchain. Penelitian ini memberikan gambaran bahwa teknik data science memiliki potensi kuat untuk digunakan dalam sistem e-voting berbasis blockchain, terutama dalam mendeteksi manipulasi data suara atau aktivitas jahat yang membahayakan integritas proses pemilihan.

METODE

Jenis Penelitian

Pada penelitian ini, digunakan jenis penelitian kuantitatif eksperimental untuk menguji keamanan sistem voting elektronik berbasis blockchain. Penelitian ini dilakukan dengan menganalisis data transaksi pada sistem blockchain menggunakan algoritma data science Random Forest dan Support Vector Machine (SVM). Pendekatan ini bertujuan untuk mendeteksi anomali dan mengidentifikasi kerentanan keamanan dalam sistem tersebut.

Penelitian ini juga berfokus pada pengumpulan dan pengolahan data yang dihasilkan dari simulasi sistem voting berbasis blockchain. Hasilnya diharapkan dapat memberikan gambaran yang lebih jelas dan terukur mengenai sejauh mana teknologi blockchain dapat menjaga integritas suara dan transparansi dalam proses voting. Pendekatan ini dilakukan agar evaluasi keamanan yang dihasilkan lebih akurat dan dapat digunakan sebagai acuan untuk pengembangan sistem di masa depan.

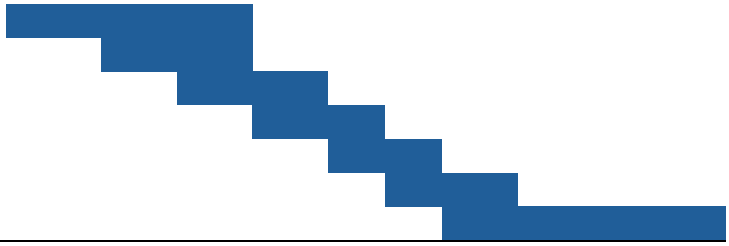
Waktu dan Tempat Pelaksanaan Penelitian

Penulis melakukan penelitian dari Bulan April 2024 di Universitas Prima Indonesia Berikut adalah jadwal pelaksanaan penelitian yang dilakukan.

Tabel 1. Jadwal Penelitian

No	Kegiatan	Bulan									
		2024		2025							
		11	12	1	2	3	4	5	6	7	8
1	Diskusi topik Penelitian										
2	Pencarian Jurnal										

3	Pengumpulan Data
4	Pembuatan Proposal
5	Diskusi Proposal
6	Pembuatan Aplikasi
7	Pembuatan Laporan
8	Publikasi Jurnal
9	Diseminasi

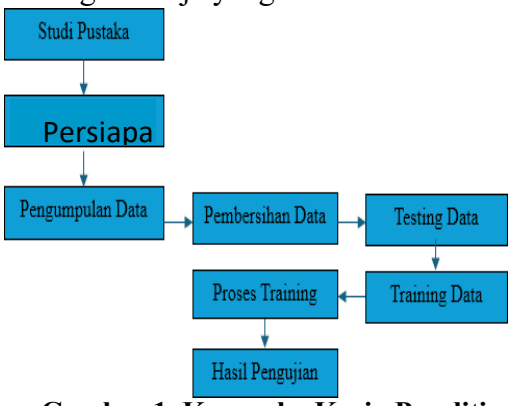


Prosedur Kerja

Langkah-langkah yang dilakukan dalam penelitian ini dapat dijabarkan sebagai berikut:

1. Mempelajari jurnal internasional dan nasional yang berhubungan dengan teknologi blockchain, keamanan pada sistem voting elektronik, serta penerapan algoritma data science seperti Random Forest dan SVM untuk analisis keamanan.
2. Mengidentifikasi masalah utama yang dihadapi dalam sistem voting berbasis blockchain, khususnya mengenai ancaman terhadap keamanan data
3. Menjelaskan sistem simulasi blockchain yang akan digunakan dalam penelitian ini, termasuk memilih dan menerapkan algoritma untuk evaluasi keamanan.
4. Melakukan pengujian terhadap dataset untuk menilai akurasi algoritma dalam mendeteksi potensi ancaman dan kerentanan sistem.
5. Menyusun laporan penelitian yang mencakup analisis hasil pengujian dan memberikan rekomendasi untuk meningkatkan keamanan sistem voting berbasis blockchain.

Kerangka Kerja yang Diusulkan :



Gambar 1. Kerangka Kerja Penelitian

Alat dan Bahan

Alat

Dalam melakukan penelitian ini terdapat beberapa alat dan bahan yang digunakan adalah sebagai berikut:

Nama Alat	Spesifikasi	Deskripsi Penggunaan
Laptop ASUS ROG Strix G512LV	Prosesor: Intel Core i7 10875H GPU: NVIDIA GeForce RTX 2060 6GB VRAM RAM: 16GB SSD: 2TB	Laptop ini digunakan untuk menjalankan simulasi blockchain, melatih model algoritma data science, dan menganalisis data penelitian.
Python	Versi 3.8 atau lebih tinggi	Bahasa pemrograman utama yang digunakan untuk implementasi algoritma seperti Random Forest dan SVM, serta pengolahan data.
Jupyter Notebook	Versi terbaru	Alat untuk melakukan manipulasi data dan eksperimen analisis algoritma dengan cara yang interaktif dan terstruktur.
VS Code	Versi terbaru	IDE yang digunakan untuk menulis, menjalankan, dan mengelola kode pemrograman selama penelitian berlangsung.
GitHub/Git	Versi terbaru	Platform ini digunakan untuk mengelola versi kode, menyimpan proyek, serta berkolaborasi.

Internet	Kecepatan 50Mbps	Koneksi internet diperlukan untuk mengunduh dataset, library, dan melakukan riset literatur sebagai pendukung penelitian.

Tabel 2. Alat Penelitian

digunakan dalam penelitian ini:

Bahan

Berikut adalah bahan-bahan yang

Tabel 3. Bahan Penelitian

Nama Bahan	Deskripsi	Penggunaan
Dataset Blockchain	Dataset yang dihasilkan dari simulasi blockchain atau dataset publik seperti transaksi Ethereum.	Dataset ini digunakan untuk melatih dan menguji algoritma seperti Random Forest dan SVM dalam penelitian.
Library Python	**Scikit-learn:** Untuk menerapkan algoritma Random Forest dan SVM. **Pandas &** **NumPy:** Untuk manipulasi data.	Library ini mempermudah pengolahan data, pembuatan model prediksi, dan evaluasi kinerja algoritma.
Visualisasi Data	**Matplotlib & Seaborn:** Untuk membuat grafik dan visualisasi dari hasil analisis data.	Digunakan untuk menyajikan hasil analisis data secara visual agar lebih mudah dipahami.
Dokumentasi Penelitian	Dokumen penelitian dari jurnal atau konferensi yang relevan dengan blockchain dan data science (2020 ke atas).	Sebagai referensi untuk mendukung teori penelitian dan membandingkan metode yang digunakan.

Metode

Dalam penelitian ini, metode yang digunakan adalah simulasi blockchain, Random Forest, dan Support Vector Machine (SVM) untuk menganalisis keamanan sistem voting berbasis blockchain. Ketiga metode ini digunakan untuk mengidentifikasi dan menganalisis potensi ancaman dalam sistem voting berbasis blockchain yang disimulasika

HASIL DAN PEMBAHASAN

Data dan Sumber Data

Sumber Data

Sumber Dataset Dataset yang digunakan dalam penelitian ini diperoleh dari platform berbagi data Kaggle https://www.kaggle.com/datasets/ardodev/ blockchain- voting-transactions-jurnal/ . Dataset tersebut dipublikasikan pada bulan November 2024, namun saya upload ulang karena dataset tersebut terhapus dari kaggle, berisi sebanyak 500.000 data transaksi terkait pemungutan suara elektronik berbasis teknologi blockchain. Pemilihan dataset ini didasarkan pada kesesuaiannya dengan tujuan penelitian, yaitu untuk menguji keamanan sistem voting berbasis blockchain dengan menggunakan algoritma Random Forest dan Support Vector Machine (SVM).

Struktur Data

Dataset ini mencakup berbagai informasi penting terkait proses pemungutan suara berbasis blockchain, dengan beberapa fitur utama sebagai berikut: Trasaction_ID : ID unik untuk setiap transaksi voting, Voter_ID : Identitas pemilih (disamarkan untuk menjaga privasi), Timestamp : Waktu pemungutan suara dilakukan, Candidate_ID : ID kandidat yang dipilih oleh pemilih, Blockchain_Hash : Hash unik untuk setiap transaksi voting yang tersimpan di blockchain, Block_Number : Nomor blok tempat transaksi voting disimpan, Verification_Status : Status verivikasi transaksi voting (Verified, Rejected atau Pending), Smart_Contract_Address :

Alamat kontak pintar yang

digunakan untuk mengelola proses voting. Fitur-fitur tersebut memberikan informasi penting mengenai keabsahan dan integritas dari proses voting yang dilakukan secara digital. Data ini sangat relevan untuk mengidentifikasi pola anomali dan potensi risiko keamanan pada sistem voting berbasis blockchain. Penggunaan dataset ini memberikan dasar yang kuat untuk melakukan analisis keamanan sistem e- voting. Dataset telah mencakup berbagai atribut yang memungkinkan untuk deteksi terhadap potensi anomali dan validasi transaksi voting.

Pengolahan Data dan Eksperimen Algoritma

Pra-pemrosesan Data

Langkah pertama dalam pengolahan data adalah melakukan pembersihan (data cleaning) dan transformasi data agar siap digunakan dalam pelatihan model. Beberapa tindakan penting yang dilakukan pada tahap ini antara lain**:**

a. Menghapus kolom yang tidak relevan untuk analisis machine learning, seperti Blockchain_Hash dan Smart_Contract_Address yang bersifat unik per transaksi namun tidak memberi nilai prediktif.
b. Mengubah nilai kategorikal seperti Verification_Status menjadi format numerik.
c. Mengecek dan menangani nilai kosong atau duplikat, jika ada.
d. Membuat label untuk klasifikasi: misalnya menjadikan Verification_Status sebagai target, di mana kelas Verified dianggap valid, dan Rejected serta Pending dianggap sebagai indikasi potensi ancaman atau anomali.

Pembagian Dataset (Train-Test Split)

Langkah selanjutnya adalah membagi dataset menjadi dua subset, yaitu data pelatihan (training data) dan data pengujian (testing data). Tujuan dari pembagian ini adalah untuk melatih model pada sebagian data, lalu menguji performanya terhadap data yang belum pernah dilihat sebelumnya, guna mengukur kemampuan generalisasi model. Pembagian dilakukan dengan rasio 80:20, di mana 80% dari data digunakan untuk pelatihan dan 20% sisanya untuk pengujian. Teknik stratified splitting diterapkan agar proporsi kelas pada label tetap seimbang di kedua subset. Ini penting karena dataset memiliki distribusi kelas yang tidak seimbang antara transaksi yang diverifikasi dan tidak diverifikasi.

```
Jumlah fitur (kolom): 9
Total data (baris): 500000

Ukuran data pelatihan (X_train): (400000, 9)
  -> Jumlah data pelatihan: 400000
  -> Jumlah fitur: 9

Ukuran data pengujian (X_test): (100000, 9)
  -> Jumlah data pengujian: 100000
  -> Jumlah fitur: 9
```

Gambar 2. Pembagian Dataset

Dalam proses praproses data, dilakukan penghapusan terhadap tiga fitur yaitu Blockchain_Hash, Smart_Contract_Address, dan Transaction_ID. Ketiga atribut tersebut merupakan penanda unik (identifier) yang tidak memberikan informasi bermakna dalam proses klasifikasi. Atribut seperti ini dapat mengganggu proses pembelajaran model karena bersifat acak dan tidak memiliki pola yang dapat dipelajari. Selain itu, atribut Timestamp yang awalnya berupa data waktu lengkap, diubah menjadi enam komponen terpisah yang merepresentasikan tanggal, bulan, tahun, jam, menit, dan detik. Transformasi ini dilakukan agar informasi temporal dari waktu transaksi dapat dimanfaatkan secara optimal oleh algoritma pembelajaran mesin dalam menemukan pola terhadap status verifikasi transaksi. seperti: 'Timestamp_Day' 'Timestamp_Month' 'Timestamp_Year' 'Timestamp_Hour' 'Timestamp_Minute' 'Timestamp_Second'. Transformasi ini bertujuan untuk mengubah data waktu dalam format yang lebih terstruktur, sehingga model dapat menangkap pola berdasarkan waktu dalam data. Setelah proses preprocessing, dataset

yang digunakan untuk pelatihan model memiliki 9 fitur. Fitur-fitur ini termasuk data yang relevan untuk memprediksi status verifikasi pemilih, dan sudah bebas dari kolom yang tidak berkontribusi secara langsung terhadap target prediksi.

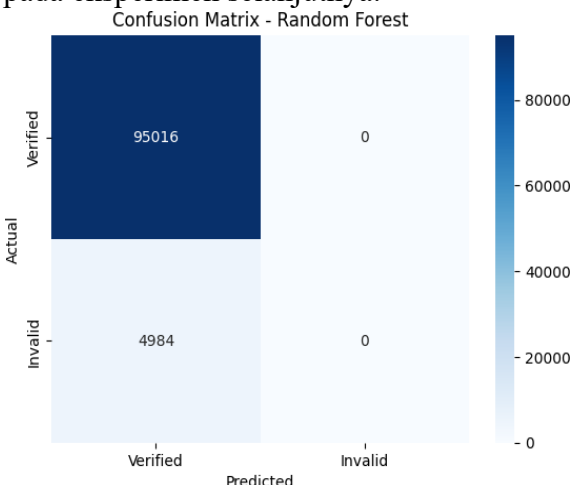
```
Jumlah fitur sebelum drop kolom: 10
Nama kolom setelah drop kolom: Index(['Voter_ID', 'Candidate_ID', 'Block_Number', 'Verification_Status',
       'Timestamp_Day', 'Timestamp_Month', 'Timestamp_Year', 'Timestamp_Hour',
       'Timestamp_Minute', 'Timestamp_Second'],
      dtype='object')
Jumlah fitur setelah drop 'Verification_Status': 9
Jumlah fitur data pelatihan: 9
Jumlah fitur data pengujian: 9
```

Gambar 3. Pembagian Fitur Dataset

Implementasi dan Eksperimen Algoritma

Eksperimen 1: Random Forest Model Awal

Berdasarkan hasil evaluasi, dilakukan pengujian terhadap data uji untuk mengetahui akurasi awal model dalam memprediksi transaksi valid maupun tidak valid. Selanjutnya, model menghasilkan metrik evaluasi berupa akurasi, precision, recall, dan F1-score. Confusion matrix juga digunakan untuk memperjelas jumlah prediksi benar dan salah pada masing-masing kelas. Hasil dari eksperimen ini menjadi acuan dasar sebelum dilakukan peningkatan performa pada eksperimen selanjutnya.



Gambar 4. Confusion Matrix Random Forest Model Awal

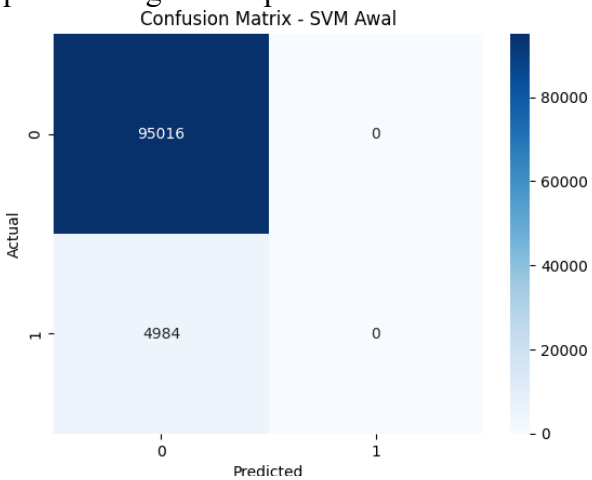
Model awal Random Forest menunjukkan hasil yang belum memuaskan, terutama dalam mengidentifikasi transaksi yang tidak valid. Hal ini ditandai dengan rendahnya nilai recall dan precision untuk kelas "Invalid", bahkan dalam beberapa kasus model tidak berhasil mendeteksi sama sekali. Hasil ini menjadi indikator penting bahwa dibutuhkan langkah optimasi dan penyesuaian model lebih lanjut, yang akan dilakukan pada eksperimen selanjutnya.

```
=== Evaluasi Random Forest (Model Awal) ===
              precision    recall  f1-score   support

    Verified       0.95      1.00      0.97     95016
     Invalid       0.00      0.00      0.00      4984

    accuracy                           0.95    100000
   macro avg       0.48      0.50      0.49    100000
weighted avg       0.90      0.95      0.93    100000
```

Gambar 5. Evaluasi Random Forest Tahap Awal

Eksperimen 2: SVM Model Awal

Setelah menguji model awal Random Forest, pada tahap ini kita melakukan eksperimen menggunakan algoritma Support Vector Machine (SVM) dengan kernel Radial Basis Function (RBF) secara default tanpa tuning. Tujuannya adalah untuk mengetahui performa awal SVM dalam mendeteksi transaksi yang tidak valid, sebagai pembanding terhadap Random Forest.



Gambar 6. Confusion Matrix SVM Awal

Tingkat akurasi keseluruhan yang dicapai adalah 95%, namun apabila ditinjau lebih dalam, model sepenuhnya gagal mengenali transaksi tidak valid (kelas 1). Hal ini ditunjukkan oleh nilai recall dan precision untuk kelas 1 yang sama-sama nol. Artinya, semua prediksi diarahkan ke kelas mayoritas (valid), sedangkan seluruh data dari kelas minoritas

(tidak valid) justru tidak terdeteksi.

Hasil Evaluasi SVM Model Awal:

	precision	recall	f1-score	support
0	0.95	1.00	0.97	95016
1	0.00	0.00	0.00	4984
accuracy			0.95	100000
macro avg	0.48	0.50	0.49	100000
weighted avg	0.90	0.95	0.93	100000

Gambar 7. Evaluasi SVM Tahap Awal

Eksperimen awal menggunakan SVM belum menunjukkan hasil yang memuaskan dalam mendeteksi transaksi bermasalah. Hal ini menunjukkan bahwa diperlukan tahapan lanjutan seperti optimasi parameter dan penanganan ketidakseimbangan kelas agar model dapat berperforma lebih baik terhadap seluruh kelas. Oleh karena itu, pada bagian selanjutnya akan dilakukan upaya optimasi terhadap kedua algoritma yang telah diuji.

Eksperimen 3: Optimasi Random Forest

Setelah evaluasi terhadap model Random Forest awal menunjukkan keterbatasan dalam mengenali kelas minoritas, dilakukan tahap optimasi guna meningkatkan kinerja klasifikasi, khususnya dalam mendeteksi transaksi yang tidak valid. Strategi optimasi yang digunakan meliputi:

a. Penyesuaian parameter model (hyperparameter tuning) seperti n_estimators, max_depth, dan class_weight.
b. Penanganan class imbalance dengan menambahkan parameter class_weight='balanced' untuk memaksa model memberikan bobot yang proporsional pada kelas minoritas.

Hasil Evaluasi Random Forest Setelah Optimasi:

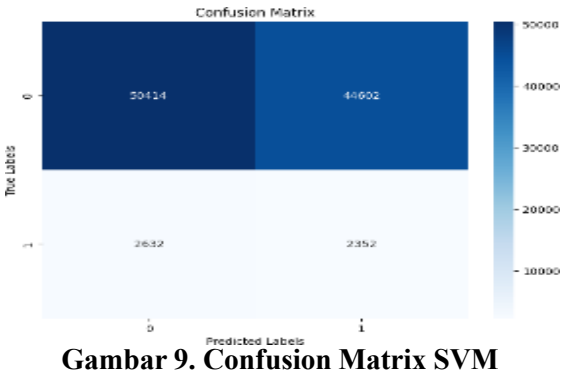
	precision	recall	f1-score	support
0	0.9502	0.9998	0.9743	95016
1	0.0000	0.0000	0.0000	4984
accuracy			0.9500	100000
macro avg	0.4751	0.4999	0.4872	100000
weighted avg	0.9028	0.9500	0.9258	100000

Gambar 8. Evaluasi Random Forest Optimized

Meskipun telah dilakukan proses optimasi terhadap model Random Forest dengan menyesuaikan jumlah estimators, kedalaman pohon, serta bobot kelas (class_weight='balanced'), hasil evaluasi menunjukkan bahwa model masih belum berhasil mengenali transaksi bermasalah secara efektif. Hal ini terlihat dari nilai recall dan precision untuk kelas minoritas yang berada pada angka nol. Kondisi tersebut mengindikasikan bahwa permasalahan ketidakseimbangan kelas memiliki pengaruh signifikan terhadap kinerja model. Oleh karena itu, pada tahap selanjutnya akan dilakukan pendekatan tambahan melalui teknik sampling dan tuning parameter lebih lanjut untuk meningkatkan sensitivitas model terhadap kelas minoritas.

Eksperimen 4: Optimasi SVM

Pada eksperimen ini, dilakukan optimasi terhadap algoritma Support Vector Machine (SVM) dengan pendekatan kernel non-linear (RBF), pemberian bobot seimbang terhadap kelas (melalui class_weight='balanced'), serta pengaturan parameter kompleksitas margin (C=10). Tujuan dari langkah ini adalah meningkatkan sensitivitas model terhadap data kelas minoritas, yaitu transaksi yang tidak valid.



Gambar 9. Confusion Matrix SVM Optimized

Berdasarkan hasil evaluasi, terjadi peningkatan signifikan dalam kemampuan model untuk mendeteksi kelas 1. Recall untuk kelas minoritas mencapai 0.47, artinya sekitar 47% dari transaksi tidak valid berhasil dikenali oleh model, jauh lebih baik dibandingkan eksperimen awal yang gagal sepenuhnya. Namun, peningkatan recall ini menyebabkan trade-off pada kelas mayoritas, di mana akurasi keseluruhan menurun menjadi 53%. Hal ini

dapat dimaklumi, mengingat model kini mencoba lebih "berani" dalam mengklasifikasikan data minoritas.

Hasil Evaluasi SVM Setelah Optimasi:

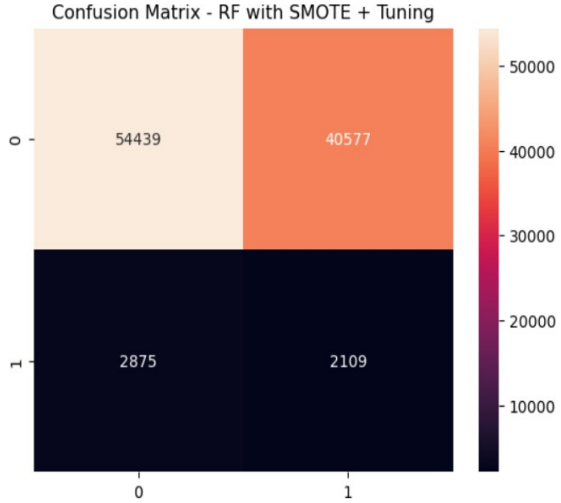
	precision	recall	f1-score	support
0	0.95	0.53	0.68	95016
1	0.05	0.47	0.09	4984
accuracy			0.53	100000
macro avg	0.50	0.50	0.39	100000
weighted avg	0.91	0.53	0.65	100000

Gambar 10. Evaluasi SVM Optimized

Berdasarkan hasil eksperimen sebelumnya, terlihat bahwa pendekatan class_weight saja belum cukup untuk menangani ketidakseimbangan kelas. Oleh karena itu, pada tahap ini dilakukan pendekatan resampling menggunakan teknik SMOTE sebelum pelatihan model, dan tuning parameter dilakukan melalui GridSearchCV.

Tuning Model Random Forest dengan SMOTE dan GridSearchCV

Pada tahap ini, dilakukan tuning terhadap model Random Forest untuk meningkatkan kemampuan dalam mendeteksi transaksi tidak valid. Mengingat ketidakseimbangan kelas yang signifikan dalam dataset, digunakan pendekatan resampling dengan metode SMOTE (Synthetic Minority Oversampling Technique), yang dikombinasikan dalam pipeline bersama algoritma Random Forest. Proses tuning parameter dilakukan menggunakan GridSearchCV dengan validasi silang lima lipatan (5-fold cross-validation), serta menggunakan metrik recall sebagai fokus utama. Parameter yang disetel meliputi jumlah pohon (n_estimators), kedalaman maksimum (max_depth), dan bobot kelas (class_weight). Model diuji menggunakan keseluruhan dataset tanpa pengambilan sampel. Hasil GridSearchCV menunjukkan bahwa konfigurasi terbaik untuk model adalah sebagai berikut: Best Params: {'rf_____class_weight': 'balanced', 'rf max_depth': 10, 'rf__n_estimators': 100}.



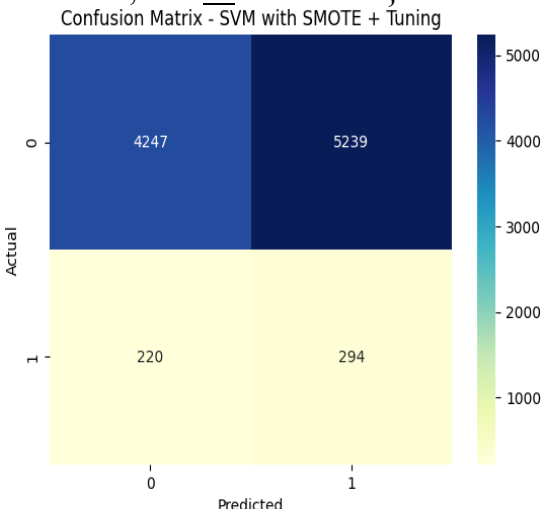
Gambar 11. Tuning Model Random Forest dengan SMOTE dan GridSearchCV

Peningkatan recall ini menunjukkan bahwa model menjadi lebih sensitif dalam mengidentifikasi transaksi yang tidak valid, yang merupakan aspek kritis dalam konteks keamanan sistem e-voting. Trade-off dalam bentuk penurunan precision dapat ditoleransi, mengingat hasil prediksi masih dapat diverifikasi melalui proses validasi manual.

Tuning Model SVM dengan SMOTE dan GridSearchCV

Eksperimen selanjutnya dilakukan terhadap algoritma Support Vector Machine (SVM) dengan tujuan meningkatkan deteksi transaksi tidak valid. Sama seperti pendekatan sebelumnya, digunakan metode resampling SMOTE (Synthetic Minority Oversampling Technique) untuk menangani ketidakseimbangan kelas. Teknik ini dikombinasikan dalam sebuah pipeline bersama model SVM, dan proses tuning dilakukan melalui GridSearchCV. Untuk menghindari kendala memori yang terjadi pada saat pemrosesan data dalam skala penuh, eksperimen SVM ini dilakukan menggunakan sampel acak sebanyak 50.000 data transaksi, yang diambil dari dataset asli. Pengambilan sampel dilakukan dengan acakan tetap (random_state) guna menjaga reprodusibilitas. Parameter yang ditelusuri melalui GridSearchCV mencakup nilai C sebagai kontrol kompleksitas margin, jenis

kernel, serta bobot kelas (class_weight). Proses tuning menggunakan 5-fold cross-validation dan metrik recall sebagai prioritas utama, sesuai dengan tujuan mendeteksi sebanyak mungkin transaksi tidak sah. Dari hasil tuning, kombinasi parameter terbaik yang diperoleh adalah: {'svm____C': 10, 'svm____class_weight': 'balanced', 'svm__kernel': 'rbf'}.



Gambar 12. Tuning Model SVM dengan SMOTE dan GridSearchCV

Akurasi keseluruhan tercatat sebesar 45%, namun yang menjadi perhatian utama adalah recall untuk kelas minoritas mencapai 57%, lebih tinggi dibanding model Random Forest. Hal ini menunjukkan bahwa meskipun model masih memiliki kelemahan dalam precision, kemampuannya dalam mengenali transaksi bermasalah meningkat secara signifikan. Kinerja model ini menunjukkan bahwa pendekatan resampling dan tuning terhadap SVM pada subset data dapat memberikan indikasi awal terhadap efektivitas pendeteksian anomali. Namun, untuk memperoleh generalisasi yang lebih kuat, diperlukan pengujian ulang terhadap keseluruhan dataset.

Analisis Hasil Pengujian

Evaluasi dan Perbandingan Akhir

Tahap akhir dari eksperimen ini adalah melakukan evaluasi komprehensif terhadap seluruh model yang telah diuji, baik sebelum maupun setelah proses optimasi dan tuning. Penilaian dilakukan dengan mempertimbangkan metrik evaluasi utama, yaitu precision, recall, f1-score, dan akurasi, dengan fokus khusus pada kemampuan mendeteksi transaksi tidak valid (kelas minoritas). Berikut ini adalah ringkasan hasil evaluasi seluruh model:

Tabel 4. Evaluasi Performa

Model	Dataset	Accuracy	Precision (kelas 1)	Recall (kelas 1)	F1-Score (kelas 1)
Random Forest (Awal)	Full	95%	0.00	0.00	0.00
SVM (Awal)	Full	95%	0.00	0.00	0.00
Random Forest (Tuned + SMOTE)	Full	57%	0.05	0.42	0.09
SVM (Tuned + SMOTE)	Sample (50k)	45%	0.05	0.57	0.10

Tabel di atas menunjukkan bahwa baik model Random Forest maupun SVM gagal total dalam kondisi awal, dengan recall nol terhadap transaksi bermasalah. Hal ini menunjukkan bahwa model sangat bias terhadap kelas mayoritas karena distribusi yang sangat tidak seimbang. Setelah diterapkan teknik resampling menggunakan SMOTE, diikuti dengan tuning parameter melalui GridSearchCV, performa kedua model mengalami peningkatan signifikan pada aspek recall. SVM yang dituning pada sampel data 50.000 transaksi menunjukkan recall tertinggi sebesar 57%, sedangkan Random Forest yang dituning pada keseluruhan dataset menghasilkan recall sebesar 42%. Namun, peningkatan recall ini diiringi penurunan akurasi keseluruhan dan precision, khususnya pada kelas minoritas. Ini merupakan trade-off yang umum terjadi dalam skenario deteksi anomali, di mana recall lebih diutamakan karena tujuannya adalah menangkap sebanyak mungkin potensi transaksi bermasalah, meskipun dengan risiko kesalahan deteksi (false

positive) yang lebih tinggi. Dengan mempertimbangkan fokus penelitian dalam konteks keamanan sistem voting berbasis blockchain, yaitu kemampuan mendeteksi ancaman atau transaksi tidak sah, maka model SVM hasil tuning dapat dikatakan paling efektif dalam mengidentifikasi anomali meskipun akurasinya rendah. Model ini lebih sensitif terhadap transaksi abnormal dan dapat digunakan sebagai mekanisme awal untuk seleksi transaksi yang perlu diverifikasi lebih lanjut oleh sistem keamanan tambahan.

Pembahasan

Implikasi terhadap Sistem Voting Berbasis Blockchain

Hasil eksperimen menunjukkan bahwa model pembelajaran mesin memiliki potensi yang signifikan dalam mendeteksi transaksi tidak valid pada sistem voting elektronik berbasis blockchain. Namun, pada kondisi awal, baik model Random Forest maupun SVM tidak mampu mengenali transaksi bermasalah akibat dominasi kelas mayoritas yang sangat tinggi.Dengan penerapan teknik resampling menggunakan SMOTE, disertai tuning parameter secara sistematis, performa model meningkat secara drastis dalam aspek recall, terutama untuk kelas minoritas. Ini penting dalam konteks keamanan, karena sistem voting yang tidak mampu mengenali ancaman atau anomali dapat membuka celah terhadap manipulasi hasil pemungutan suara.

Model SVM yang dituning berhasil mencapai recall sebesar 57%, sedangkan Random Forest sebesar 42%, menunjukkan bahwa metode pembelajaran mesin mampu menjadi bagian dari lapisan pertahanan keamanan sistem dengan menandai transaksi-transaksi mencurigakan yang dapat ditindaklanjuti oleh sistem validasi berikutnya.

Analisis Keandalan dan Deteksi Ancaman

Dari sisi keandalan, model yang telah dituning cenderung lebih sensitif dalam mendeteksi ancaman, meskipun dengan akurasi total yang lebih rendah. Dalam sistem keamanan informasi, khususnya dalam e-voting, sensitivitas tinggi terhadap anomali seringkali lebih diutamakan daripada akurasi absolut, karena transaksi yang dicurigai masih dapat diverifikasi lebih lanjut secara manual atau melalui smart contract verifikasi. Namun demikian, precision yang rendah juga menunjukkan bahwa masih terdapat banyak transaksi valid yang salah ditandai. Hal ini perlu diperhatikan pada implementasi sistem riil agar tidak menimbulkan gangguan terhadap suara sah. Oleh karena itu, sistem deteksi ini sebaiknya diintegrasikan dengan mekanisme verifikasi tambahan, bukan sebagai penentu akhir

SIMPULAN

Berdasarkan hasil penelitian dan eksperimen yang telah dilakukan, diperoleh beberapa kesimpulan sebagai berikut:

1. Proses pengolahan data telah dilakukan secara sistematis, meliputi pembersihan data, transformasi fitur, dan pembagian dataset menjadi data pelatihan dan pengujian. Algoritma Random Forest dan Support Vector Machine (SVM) berhasil diterapkan untuk mendeteksi transaksi tidak valid pada sistem voting berbasis blockchain.
2. Ketidakseimbangan kelas menyebabkan model awal gagal mengenali kelas minoritas, dengan recall yang sangat rendah meskipun akurasi tinggi.
3. Penerapan teknik SMOTE dan tuning parameter melalui GridSearchCV mampu meningkatkan performa model, terutama pada metrik recall kelas minoritas. Model SVM hasil tuning memberikan recall tertinggi sebesar 57%, menunjukkan sensitivitas tinggi

terhadap transaksi anomali, meskipun precision dan akurasi menurun.
4. Evaluasi menggunakan metrik precision, recall, f1-score, dan confusion matrix penting untuk menilai performa model secara menyeluruh, khususnya dalam konteks keamanan sistem.
5. Ukuran dataset yang besar menimbulkan tantangan dalam efisiensi komputasi, sehingga perlu pendekatan bertahap dalam pengujian dan tuning model.

Saran

Untuk pengembangan selanjutnya, disarankan beberapa hal berikut:

1. Terapkan teknik penyeimbangan data seperti SMOTE, ADASYN, atau undersampling untuk menangani distribusi kelas yang tidak seimbang.
2. Tambahkan fitur berbasis domain, seperti pola waktu, frekuensi aktivitas, atau hubungan antar alamat.
3. Uji algoritma alternatif seperti XGBoost, LightGBM, atau metode deep learning untuk meningkatkan kinerja model.
4. Pertimbangkan sistem deteksi real-time yang terintegrasi dengan blockchain guna respons cepat terhadap potensi ancaman

DAFTAR PUSTAKA

Ningrum, M. S. (2023). Pengembangan sistem keamanan data berbasis blockchain untuk aplikasi e-voting. ResearchGate. https://www.researchgate.net/publication/374033995_Pengembangan_Sistem_Keamanan_Data_Berbasis_Blockchain_untuk_Aplikasi_EVoting_Muthia_Sari_Ningrum_158700005

Ajib Susanto, (2020). Implementation of Smart Contracts Ethereum Blockchain in Web-Based Electronic Voting (e-voting). Transformatika: Jurnal Pengembangan Teknologi Informasi dan Komputer, 18(1), 13–21. https://doi.org/10.26623/transformatika.v18i1.1779

Setiawan, R. A., & Putri, W. T. H. (2023). Implementasi Teknologi Blockchain dalam Aplikasi E-Voting Berbasis Mobile. Digital Zone: Jurnal Teknologi Informasi dan Komunikasi, 14(2), 219–231. https://doi.org/10.31849/digitalzone.v14i2.16682

Pratama, Y. W., & Kurniadi, D. (2021). *Implementasi Blockchain dalam Aplikasi Pemilu*. Indonesian Journal of Social and Humanities Studies (INCARE), 2(2), 152–158. https://ejournal.ijshs.org/index.php/incare/article/view/256

Khan, A. I., Alghamdi, A., & Rehman, A. (2025). AI and Blockchain in Cybersecurity: A Sustainable Approach to Protecting Digital Assets. ResearchGate. https://www.researchgate.net/publication/392401066_AI_and_Blockchain_in_Cybersecurity_A_Sustainable_Approach_to_Protecting_Digital_Assets/fulltext/68407cb1c33afe388aca2387/AI-and-Blockchain-in-Cybersecurity-A-Sustainable-Approach-to-Protecting-Digital-Assets.pdf

Eko Yanuarso Budi, Cahyo Prihantoro, & Nicolaus Euclides Wahyu Nugroho. (2023). Perancangan Website E-Voting Menggunakan Smart Contract Pada Blockchain Polygon. The Indonesian Journal of Computer Science (IJCS), 12(3), 202–209. http://ijcs.net/ijcs/index.php/ijcs/article/view/3234/172

Potalangi, Joshua Felix, Kartikasari, Dany Primanita, & Shaffan, Nur Hazbiy. (2025). Implementasi Jaringan Permissioned Blockchain pada Sistem E- Voting Pemilwa untuk Menjamin Autentikasi Pemilih dan Integritas Data. J- PTIIK (Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer), 9(4), 1680–1686. https://j-

ptiik.ub.ac.id/index.php/j-ptiik/article/view/14708

Septiana, I., & Alita, D. (2024). Perbandingan Random Forest dan SVM dalam Analisis Sentimen Quick Count Pemilu 2024. Jurnal Informatika: Jurnal Pengembangan IT, 9(3), 224–231. https://ejournal.poltekharber.ac.id/index.php/informatika/article/view/6640

Naura Fayza I, Nicholas Svensons, Sri Asni Fatmawati, Pricillia Rotua S, & Khanaya Erviona. (2025). Analisis Perbandingan Algoritma Klasifikasi Decision Tree, K-Nearest Neighbors, Naive Bayes, dan Random Forest pada Data Pemilihan Legislatif KPU Menggunakan Kurva ROC. Journal of Data Management and Applications (JoDMApps), 1(1), 10–17. https://journal.darmajaya.ac.id/index.php/JoDMApps/article/view/911

Hasan, M., Aung, M. T., & Islam, S. R. (2024). *Detecting Anomalies in Blockchain Transactions using Machine Learning Classifiers and Explainability Analysis*. arXiv. https://arxiv.org/abs/2401.03530

Cholevas, C., Zarpalas, D., & Daras, P. (2024). *Anomaly Detection in Blockchain Networks Using Unsupervised Learning: A Survey*. Algorithms, 17(5),201.https://doi.org/10.3390/a17050201

Prabakar, R., & Kanchana, P. (2023). *E-Voting Based Blockchain Mechanism Using Feature Selection Based Machine Learning*. IJISAE: International Journal of Intelligent Systems and Applications in Engineering. https://www.ijisae.org/index.php/IJISAE/article/view/4866

Fitriani Dewi Rizki. (2023). Analisis Rekayasa Fitur untuk Mendeteksi Komentar Spam YouTube pada Pilpres 2019 di Indonesia Menggunakan Random Forest (Skripsi Sarjana, Fakultas Sains dan Teknologi, UIN Syarif Hidayatullah Jakarta) https://repository.uinjkt.ac.id/dspace/handl /123456789/71764

Mohammad Sani, Suryani, & Pahlevi. (2022). Deteksi Serangan Botnet Pada Jaringan Internet of Things Menggunakan Algoritma Random Forest (RF). Jurnal Open Library Engineering, 9(3), 297– 303. https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/17988