Informatech: Jurnal Ilmiah Informatika dan Komputer
Volume. 1, Nomor. 1, Juni 2024, hlm 23-29 **E-ISSN: 3047-4752**

Implementasi Algoritma Random Forest Untuk Optimasi Keamanan Autentikasi *One-Time Password* (OTP)

Andri Nova Riswanto[1], Derman Janner Lubis[2*]

[1] Sistem Informasi/Universitas Binaniaga Indonesia
Email: anvarisy@gmail.com
[2] Sistem Informasi/Universitas Binaniaga Indonesia
Email: derman_janner@yahoo.com

(Naskah diterima: *23 Maret 2024*; Naskah direvisi: *8 April 2024*; Naskah diterbitkan: *1 Juni 2024*)

ABSTRAK *– Penelitian ini berfokus pada optimasi keamanan autentikasi One-Time Password (OTP) melalui implementasi algoritma Random Forest. Tujuan utamanya adalah mengembangkan sistem deteksi dan pencegahan penipuan OTP yang efisien dan efektif. Dengan menggunakan model 4D (Define, Design, Development, and Dissemination) dan pendekatan throwaway prototyping, penelitian ini menghasilkan sistem yang sesuai dengan kebutuhan pengguna dan berfungsi optimal dalam penggunaan nyata. Hasilnya menunjukkan bahwa integrasi algoritma Random Forest dalam mekanisme OTP meningkatkan keamanan secara signifikan.Uji akurasi klasifikasi sistem mencapai skor akurasi sebesar 98.5%, menegaskan efektivitas metode yang digunakan dalam mendeteksi dan mencegah penipuan OTP.*

Kata Kunci *– One-Time Password (OTP), Random Forest, Klasifikasi, Throwaway Prototyping, Kejahatan Digital.*

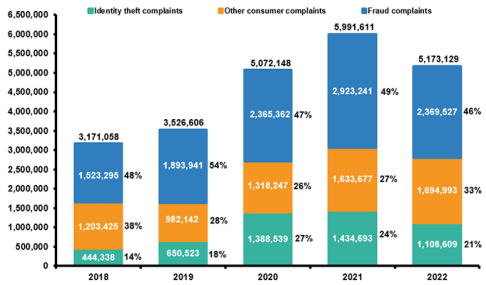
Implementation of Random Forest Algorithm for One-Time Password (OTP) Authentication Security Optimisation

ABSTRACT *– This research focuses on optimizing the security of One-Time Password (OTP) authentication through the implementation of the Random Forest algorithm. The primary goal is to develop an efficient and effective system for detecting and preventing OTP fraud. Utilizing the 4D model (Define, Design, Development, and Dissemination) and a throwaway prototyping approach, this study produced a system that meets user needs and functions optimally in real-world applications. The results show that the integration of the Random Forest algorithm into the OTP mechanism significantly enhances security. The system's classification accuracy testing reached an accuracy score of 98.5%, confirming the effectiveness of the method used in detecting and preventing OTP fraud.*

Keywords *- One-Time Password (OTP), Random Forest, Classification, Throwaway Prototyping, Digital Crime.*

1. PENDAHULUAN

Penipuan melalui OTP adalah fenomena baru yang muncul seiring dengan peningkatan transaksi digital. Kenaikan kasus penipuan OTP menjadi masalah serius dalam sektor keuangan digital.



Gambar 1. Laporan pencurian identitas dan penipuan
Sumber: Federal Trade Commission (https://www.iii.org/graph-archive/96074)

Sebagaimana yang ditunjukkan oleh Gambar 1. Data menunjukkan bahwa setiap tahun, keluhan akibat penipuan selalu mendominasi dibandingkan dengan keluhan akibat pencurian identitas.

Gambar 2. Laporan keluhan dan kerugian
Sumber: Federal Bureau Of Investigation (https://www.ic3.gov/)

Sebagaimana yang ditunjukkan oleh Gambar 2. Data menunjukkan peningkatan signifikan dalam jumlah keluhan dan kerugian finansial akibat penipuan OTP selama lima tahun terakhir. Data ini membuktikan bahwa penipuan OTP telah menjadi masalah yang serius dan memerlukan solusi yang efektif dan segera. Beberapa penelitian telah dilakukan untuk melakukan optimasi terhadap autentikasi OTP seperti menambahkan bentuk enkripsi tambahan dan penggunaan machine learning. Random Forest merupakan algoritma machine learning yang termasuk dalam kategori pembelajaran ensemble. Dalam konteks ini berarti metode ini menggabungkan sejumlah model prediktif untuk menghasilkan yang lebih baik daripada masing-masing model secara individu. Manfaat utama dari pendekatan ini adalah bahwa random forest cenderung lebih tahan terhadap overfitting dibandingkan dengan pohon keputusan tunggal. Berdasarkan pertimbangan tersebut, penelitian ini akan melakukan implementasi algoritma random forest untuk optimasi keamanan autentikasi one-time password (OTP).

2. TINJAUAN PUSTAKA

a. Machine Learning

Machine learning merupakan cabang dari ilmu komputer yang dikhususkan untuk pengembangan algoritma yang mampu belajar dari dan membuat prediksi berdasarkan data. Teknik ini melibatkan pemahaman dan penerapan algoritma yang belajar dari kumpulan contoh fenomena (Andriy Burkov, 2019, p. 3). Contoh-contoh ini bisa berasal dari alam, dibuat secara manual oleh manusia, atau dihasilkan oleh algoritma lain.

Machine learning dalam implementasinya memiliki berbagai jenis pendekatan. Menurut (Murphy, 2012, p. 2) machine learning memiliki tiga jenis pendekatan, yaitu:

a) supervised learning adalah pendekatan untuk mempelajari pemetaan dari input (x) ke output (y);
b) unsupervised learning adalah pendekatan untuk menemukan pola menarik dalam data;
c) reinforcement learning adalah pendekatan untuk belajar bagaimana bertindak atau berperilaku ketika diberikan sinyal penghargaan atau hukuman yang sesekali.

Pemahaman terhadap ketiga pendekatan dalam machine learning ini sangat penting dalam menentukan strategi yang paling efektif dalam menangani masalah.

b. Klasifikasi

Menurut (Prasetyo, 2013:45) menyatakan bahwa klasifikasi adalah aktivitas memperkirakan data untuk menggolongkan ke dalam golongan yang telah ditetapkan asal sejumlah golongan yang sudah ada. Pola yang sudah dibuat ketika pelatihan kemudian dapat digunakan untuk memprediksi label kelas data baru yag belum diketahui. Dalam pembangunan pola selama proses pelatihan memerlukan suatu algoritma untuk membangunnya yaitu Algoritma pelatihan. Ada berbagai macam algoritma pelatihan yang sudah dikembangkan oleh para peneliti yaitu C4.5, Nearest Neighbor, Bayesian Classification, Neural Network, dan sebagainya (Prasetyo, 2013:46).

3. METODE PENELITIAN

a. Algoritma Random Forest

Random Forest adalah metode dalam machine learning yang merupakan bentuk lanjutan dari pohon keputusan (decision trees). Menurut (Andriy Burkov, 2019, p. 9) Random Forest menerapkan dua paradigma utama dari ensemble learning, yaitu:

a) bagging : proses membuat beberapa sub set dari data set asli dengan prosedur bootstrap, yaitu dengan mengambil sampel secara acak dengan penggantian;
b) boosting : proses melatih model secara berurutan yang setiap model berikutnya mencoba untuk memperbaiki kesalahan yang dibuat oleh model sebelumnya, dengan memberikan bobot lebih kepada contoh-contoh yang diprediksi salah oleh

model sebelumnya.

b. One-Hot Encoding

One-Hot Encoding adalah teknik yang digunakan untuk mengubah variabel kategori atau label menjadi bentuk numerik. Pada dasarnya, setiap nilai kategori diubah menjadi vektor biner di mana setiap nilai kategori memiliki kolom sendiri dan diwakili oleh nilai 0 atau 1. Oleh karena itu, penerapan One-Hot Encoding menjadi langkah krusial dalam pengolahan data non-numerik untuk memastikan keefektifan model Random Forest. Dengan demikian, transformasi data ini memungkinkan Random Forest untuk menangani beragam jenis data, meningkatkan fleksibilitas dan kegunaannya dalam berbagai aplikasi machine learning.

c. Teknik Analisis Data

Uji hasil akurasi pada penelitian ini menggunakan confussion matrix.

Actual Values	Predicted Values	
	Class = Yes	Class = No
Class = Yes	a (true positive-TP)	b (false negative-FN)
Class = No	c (false positive-TP)	d (true negative-TN)

Akurasi adalah perbandingan kasus yang diidentifikasi benar dengan jumlah semua kasus. Rumus untuk menghitung tingkat akurasi pada matrik adalah.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} = \frac{A + D}{A + B + C + D} \dots\dots\dots$$

Keterangan:
A=jika hasil prediksi positif dan data sebenarnya positif
B=jika hasil prediksi negatif dan data sebenarnya positif
C=jika hasil prediksi positif dan data sebenarnya negatif
D=jika hasil prediksi negatif dan data sebenarnya negative

d. Definisi Variable Penelitian

Tabel 1. Variabel penelitian

No	Nama Variabel	Definisi
1	PIC ID	Merujuk pada identifikasi unik pengguna.
2	Info Perangkat	Memberikan informasi mengenai perangkat yang digunakan. Informasi ini mencakup manufacturer perangkat, nama perangkat, versi OS dan jenis platform.
3	Info Lokasi	Berisi informasi tentang lokasi pengguna (lat, lng)
4	Alamat IP	Mencakup alamat IP perangkat yang digunakan oleh pengguna.
5	Jenis Transaksi	Menyatakan jenis transaksi yang sedang dilakukan oleh pengguna.

4. HASIL DAN PEMBAHASAN

a. Hasil

Tabel 2. Uji Coba Penelitian

Jumlah Pohon	Kriteria	Random State	Data Test (%)	Jumlah Data	Jumlah Data Test
7	Gini	42	20	330	66

a) Sample Data

Tabel 3. Perilaku transaksi OTP pada satu end user

no	1	2	3
pic_id	085811751000	085811751000	085811751000
purpose	LOGIN	LOGIN	LOGIN
latitude	-6.6502314	-6.6502314	28.61318
longitude	106.7560309	106.7560309	77.209153
device_name	SM-A235F	SM-A235F	Xiaomi Redmi Note 10
os_version	14	14	13
manufacturer	samsung	samsung	Xiaomi
cpu_info	Qualcomm Technolog	Qualcomm Technolo	Snapdragon 865

	ies, Inc KHAJE	gies, Inc KHAJE	
platform	ANDROID	ANDROID	ANDROID
ip	103.28.116.75	103.28.116.75	203.123.123.123

no	4	5	6
pic_id	085811751000	085811751000	085811751000
purpose	LOGIN	TRANSACTION	TRANSACTION
latitude	-6.6502314	-6.6502764	28.61318
longitude	106.7560309	106.7556889	77.209153
device_name	Infinix X670	SM-A235F	Infinix X670
os_version	14	14	13
manufacturer	Infinix	samsung	Infinix
cpu_info	MT6781V/CDZAMB-H	Qualcomm Technologies, Inc KHAJE	MT6781V/CDZAMB-H
platform	ANDROID	ANDROID	ANDROID
ip	103.28.116.75	103.28.116.75	103.28.116.75

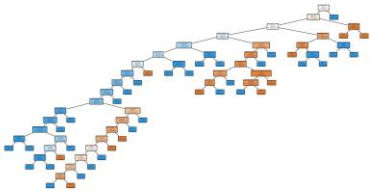
b) Hasil One Hot Encoding

Tabel 4. Hasil OneHot Encoding

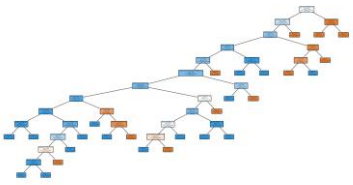
Atribute	Index					
	1	2	3	4	5	6
purpose_TRANSACTION	0	0	0	0	1	1
latitude_-6.6502764	0	0	0	0	1	0
latitude_-6.6502314	1	1	0	1	0	0
latitude_28.61318	0	0	1	0	0	1
longitude_77.209153	0	0	1	0	0	1
longitude_106.7556889	0	0	0	0	1	0
longitude_106.7560309	1	1	0	1	0	0
device_name_Infinix X670	0	0	0	1	0	1
device_name_SM-A235F	1	1	0	1	0	0
device_name_Xiaomi Redmi Note 10	0	0	1	0	0	0
os_version_13	0	0	1	0	0	1
os_version_14	1	1	0	1	1	0
manufacturer_Infinix	0	0	0	1	0	1
manufacturer_Xiaomi	0	0	1	0	0	0
manufacturer_samsung	1	1	0	0	1	0
cpu_info_MT6781V/CDZAMB-H	0	0	0	1	0	1

Atribute	Index					
	1	2	3	4	5	6
cpu_info_Qualcomm Technologies, Inc KHAJE	1	1	0	0	1	0
cpu_info_Snapdragon 865	0	0	1	0	0	0
platform_ANDROID	1	1	1	1	1	1
ip_103.28.116.75	1	1	0	1	1	1
ip_203.123.123.123	0	0	1	0	0	0

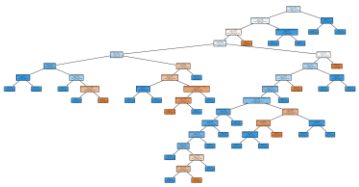
c) Pohon Keputusan



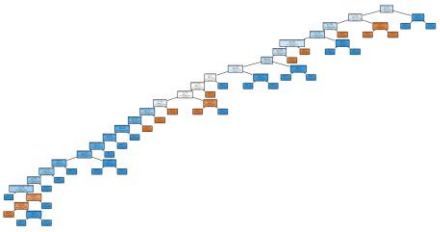
Gambar 1. Pohon keputusan n-1



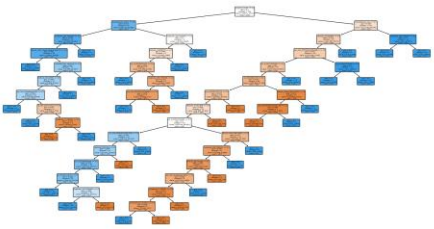
Gambar 2. Pohon keputusan n-2



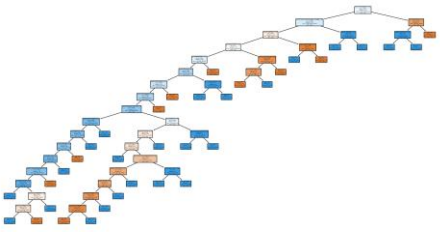
Gambar 3. Pohon keputusan n-3



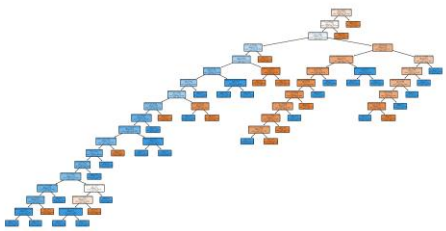
Gambar 4. Pohon keputusan n-4



Gambar 5. Pohon keputusan n-5



Gambar 6. Pohon keputusan n-6



Gambar 7. Pohon keputusan n-7

d) Hasil Voting

No	Sampel	Voting		Predicted	Actual
		anomaly	normal		
1	Sampel 1	0	7	normal	normal
2	Sampel 3	0	7	normal	normal
3	Sampel 2	4	3	anomaly	anomaly
4	Sampel 4	3	4	normal	normal
5	Sampel 5	0	7	normal	normal
6	Sampel 6	2	5	normal	normal

b. Pembahasan

Tabel 5. Hasil Confusion Matrix

Actual Values	Predicted Values	
	Anomaly	Normal
Anomaly	6	1
Normal	0	59

Nilai akurasi model machine learning ini dapat dihitung menggunakan rumus

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} = \frac{6 + 59}{6 + 0 + 59 + 1} = 98.5\%$$

a) Precision

Mengukur proporsi prediktif positif yang benar

$$Precision\ (Anomaly) = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

$$Precision\ (Anomaly) = \frac{6}{6 + 1}$$

$$Precision(Anomaly) = 0.86$$

$$Precision\ (Normal) = \frac{True\ Negatives}{True\ Negatives + False\ Negatives}$$

$$Precision\ (Normal) = \frac{59}{59 + 0}$$

$$Precision(Normal) = 1.0$$

b) Recall

Mengukur proporsi observasi yang benar benar dikenali.

$$Recall\ (Anomaly) = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

$$Recall\ (Anomaly) = \frac{6}{6+0}$$

$$Recall\ (Anomaly) = 1.0$$

$$Recall\ (Normal) = \frac{True\ Negatives}{True\ Negatives + False\ Positives}$$

$$Recall\ (Normal) = \frac{59}{59+1}$$

$$Recall\ (Normal) = 0.98$$

c) F1-Score

F1-score adalah rata-rata harmonik dari precision dan recall. F1-score berkisar antara 0 (paling buruk) hingga 1 (paling baik) dan memberikan pembobotan yang sama antara precision dan recall.

$$F1 - Score(Anomaly) = 2 \times \frac{Precision(Anomaly) \times Recall(Anomaly)}{Precision(Anomaly) + Recall(Anomaly)}$$

$$F1 - Score(Anomaly) = 2 \times \frac{0.86 \times 1.0}{0.86 + 1.0}$$

$$F1 - Score(Anomaly) = 0.92$$

$$F1 - Score(Normal) = 2 \times \frac{Precision(Normal) \times Recall(Normal)}{Precision(Normal) + Recall(Normal)}$$

$$F1 - Score(Normal) = 2 \times \frac{1.0 \times 0.98}{1.0 + 0.98}$$

$$F1 - Score(Normal) = 0.99$$

d) Support

Support adalah jumlah kasus aktual untuk masing-masing kelas di set data uji. Dalam konteks ini, support untuk 'anomaly' adalah 7, yang berarti ada 7 sampel 'anomaly' dalam data uji. Sedangkan untuk 'normal', support-nya adalah 59, yang berarti ada 59 sampel 'normal' dalam data uji.

5. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, kesimpulan yang bisa diuraikan antara lain:

a. Penggunaan algoritma Random Forest dalam sistem autentikasi OTP terbukti efektif. Dengan implementasi algoritma ini, efektivitas sistem mengalami peningkatan yang signifikan.
b. Dari segi akurasi, algoritma Random Forest menunjukkan kinerja yang sangat baik dengan tingkat akurasi mencapai 98.5% dalam mengklasifikasi transaksi OTP.

6. DAFTAR PUSTAKA

[1] Andriy Burkov, B. (2019). *The Hundred-Page Machine Learning*.

[2] Arikunto, S. (2019). Prosedur penelitian : suatu pendekatan praktik / Suharsimi Arikunto | OPAC Perpustakaan Nasional RI. In *Jakarta: Rineka Cipta*.

[3] Belciug, S., & Gorunescu, F. (2020). Intelligent Decision Support Systems - A Journey to Smarter Healthcare. In *Intelligent Decision Support Systems - A Journey to Smarter Healthcare*

[4] Bahrawi, N. (2019). Sentiment Analysis Using Random Forest Algorithm-Online Social Media Based. *Journal of Information Technology and Its Utilization*, *2*(2), 29. https://doi.org/10.30818/jitu.2.2.2695

[5] Benedictus Simarmata, K., & Dwi, K. (2022). *Analisa Rekomendasi Fitur Persetujuan Pinjaman Perusahaan Financial Technology Menggunakan Metode Random Forest*. *9*(3). http://jurnal.mdp.ac.id

[6] Cattani, G. (2023). Combining data envelopment analysis and Random Forest for selecting optimal locations of solar PV plants. *Energy and AI*, *11*. https://doi.org/10.1016/j.egyai.2022.100222

[7] Dios Kurniawan, M. S. (2020). *Pengenalan Machine Learning dengan Python*.

[8] Kurniawan, A., & Yulianingsih, Y. (2021). Pendugaan Fraud Detection pada kartu kredit dengan Machine Learning. *KILAT*, *10*(2), 320–325. https://doi.org/10.33322/kilat.v10i2.1482

[9] M. Shalahuddin, & Rosa A.S. (2015). *Rekayasa Perangkat Lunak: Terstruktur dan Berorientasi Objek*.

[10] Lu, Y., Yu, K., & Lv, X. (2021). Image encryption with one-time password mechanism and pseudo-features. *Multimedia Tools and Applications*, *80*(10), 15041–15055. https://doi.org/10.1007/s11042-021-10522-x

[11] Murphy, K. P. (2012). *Machine Learning A Probabilistic Perspective*.

[12] Prasetyo, Eko. (2013). Data Mining Konsep dan Aplikasi Menggunakan Matlab.

[13] Sarna, S., & Czerwinski, R. (2022). Small prime divisors attack and countermeasure against the rsa-otp algorithm. *Electronics*

(Switzerland), *11*(1). https://doi.org/10.3390/electronics11010095

[14] Soogun, A. O., Kharsany, A. B. M., Zewotir, T., North, D., & Ogunsakin, R. E. (2022). Identifying Potential Factors Associated with High HIV viral load in KwaZulu-Natal, South Africa using Multiple Correspondence Analysis and Random Forest Analysis. *BMC Medical Research Methodology*, *22*(1). https://doi.org/10.1186/s12874-022-01625-6

[15] Sumarni, S., & Rustam, S. (2023). Analisa Bonus Demografi Dengan Algoritma Machine Learning Di Kabupaten Gorontalo Utara. *JTKSI (Jurnal Teknologi Komputer Dan Sistem Informasi)*, *6*(1), 45. https://doi.org/10.56327/jtksi.v6i1.1391