



DETEKSI *WEB PAGE PHISHING* BERDASARKAN FITUR URL DENGAN METODE *RANDOM FOREST* DAN *LONG SHORT TERM MEMORY (LSTM)*

By Kelompok 11



Anggota

Ihsan Kamil
(1217070031)

Ilham Muzhafar Azhar
(1217070032)

Zidan Aditya
(1217070089)

Latar Belakang

Seiring berkembangnya teknologi internet, ancaman keamanan siber seperti phishing semakin meningkat. Phishing adalah metode serangan dengan membuat situs web palsu yang menyerupai situs web resmi untuk mencuri informasi sensitif. Tantangan utama dalam mendeteksi phishing adalah keragaman karakteristik situs web dan kemampuan penyerang menciptakan URL yang meyakinkan. Deteksi phishing berbasis fitur URL seperti panjang URL, karakter khusus, keabsahan domain, dan pola struktur URL menjadi solusi yang menjanjikan.

Metode pembelajaran mesin seperti Random Forest dan Long Short Term Memory (LSTM) digunakan untuk mengembangkan model deteksi phishing yang andal, cepat, dan akurat.

Kombinasi kedua metode ini diharapkan mampu memberikan solusi efektif dalam melindungi pengguna dari ancaman phishing.

Tujuan

Mendeteksi web page phishing berdasarkan karakteristik fitur URL yang relevan dalam membedakan antara URL phishing dan URL yang valid.

Mengintegrasikan metode LSTM dan Random Forest untuk menciptakan sistem deteksi phishing URL yang lebih efisien dan dapat diandalkan.

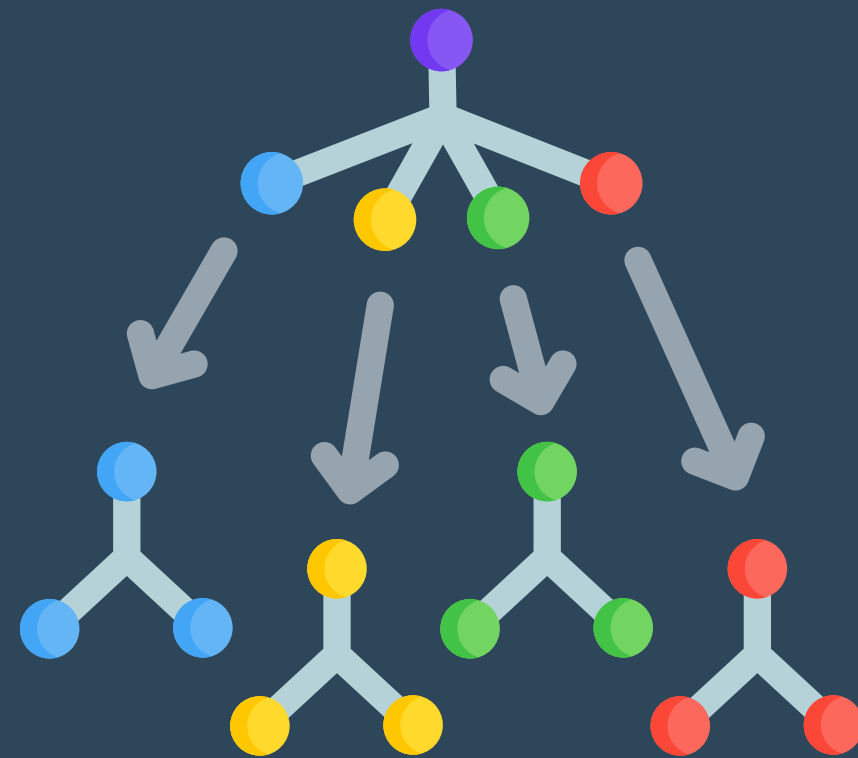
DATASET

DETEKSI URL PHISHING

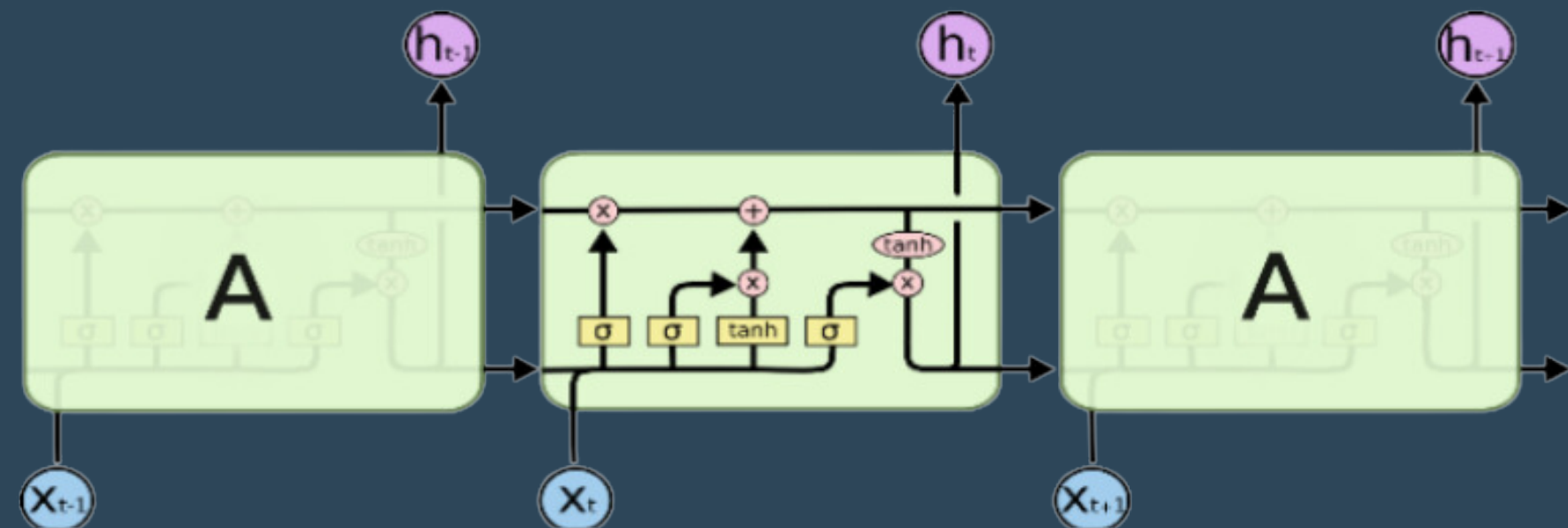
url	length_url	length_host_name	ip	nb_dots	nb_hyphens	nb_at	nb_qm	nb_and	nb_or	...	domain_in_title	domain_with_copyright	whois_registered_domain	domain_registration_length	domain_age	web_traffic	dns_record	google_index	page_rank	status	
0	http://www.crestonwood.com/router.php	37	19	0	3	0	0	0	0	0	...	0	1	0	45	-1	0	1	1	4	legitimate
1	http://shadetechnology.com/V4/validation/a...	77	23	1	1	0	0	0	0	0	...	1	0	0	77	5767	0	0	1	2	phishing
2	https://support-appleid.com.secureupdate.duila...	126	50	1	4	1	0	1	2	0	...	1	0	0	14	4004	5828815	0	1	0	phishing
3	http://rgipt.ac.in	18	11	0	2	0	0	0	0	0	...	1	0	0	62	-1	107721	0	0	3	legitimate
4	http://www.iracing.com/tracks/gateway-motorspo...	55	15	0	2	2	0	0	0	0	...	0	1	0	224	8175	8725	0	0	6	legitimate

Metode

Random Forest



Long Short Term Memory (LSTM)



Kelebihan & Kekurangan

Kelebihan

- Efisiensi dalam seleksi fitur
- Kemampuan analisis pola kompleks
- Ketahanan terhadap noise
- Akusari lebih tinggi
- Fleksibilitas

Kekurangan

- Biaya komputasi
- Kompleksitas implementasi
- Resiko overfitting
- Ketergantungan pada data berkualitas
- Kesulitan interpretasi

Pengujian

URL: **<http://www.cerostwood.com/router.php>**

Status: **legitimate**

Meskipun panjang URL cukup pendek (37 karakter), jumlah trafik web rendah (-1), tetapi domain memiliki indikator positif seperti Google Index (1) dan terdaftar di Whois (1).

Pengujian

URL: **http://shadetree...**

Status: **phishing**

Domain ini memiliki trafik lebih tinggi (5767) tetapi beberapa parameter seperti indikator domain lainnya menunjukkan potensi mencurigakan.

Pengujian

URL: **https://support-...**

Status: **phishing**

Meskipun URL panjang (128 karakter) dan memiliki trafik web tinggi (5829815), beberapa parameter lainnya menunjukkan indikasi phishing, seperti angka pada karakteristik hostname.

Pengujian

URL: **<http://rgipt.ac.in>**

Status: **legitimate**

URL ini relatif pendek (18 karakter) dan memiliki beberapa karakteristik positif seperti usia domain dan Google Index.

Pengujian

URL: **http://www.iracing...**

Status: **legitimate**

URL ini sah, dengan indikator positif seperti trafik web dan Google Index.

Hasil Metrik Evaluasi

1. Akurasi

2. Presisi

Random Forest

Accuracy: 0.9694

Precision: 0.9749

LSTM

Accuracy: 0.9379

Precision: 0.9371

Hasil Metrik Evaluasi

3. Recall

4. F1-Score

Random Forest

Recall: 0.9628

F1-Score: 0.9688

LSTM

Recall: 0.9371

F1-Score: 0.9371

Hasil Metrik Evaluasi

5. AUC (Area Under Curve)

AUC: 0.9945

AUC: 0.9820

6. Waktu Pelatihan dan Infrensi

Random Forest

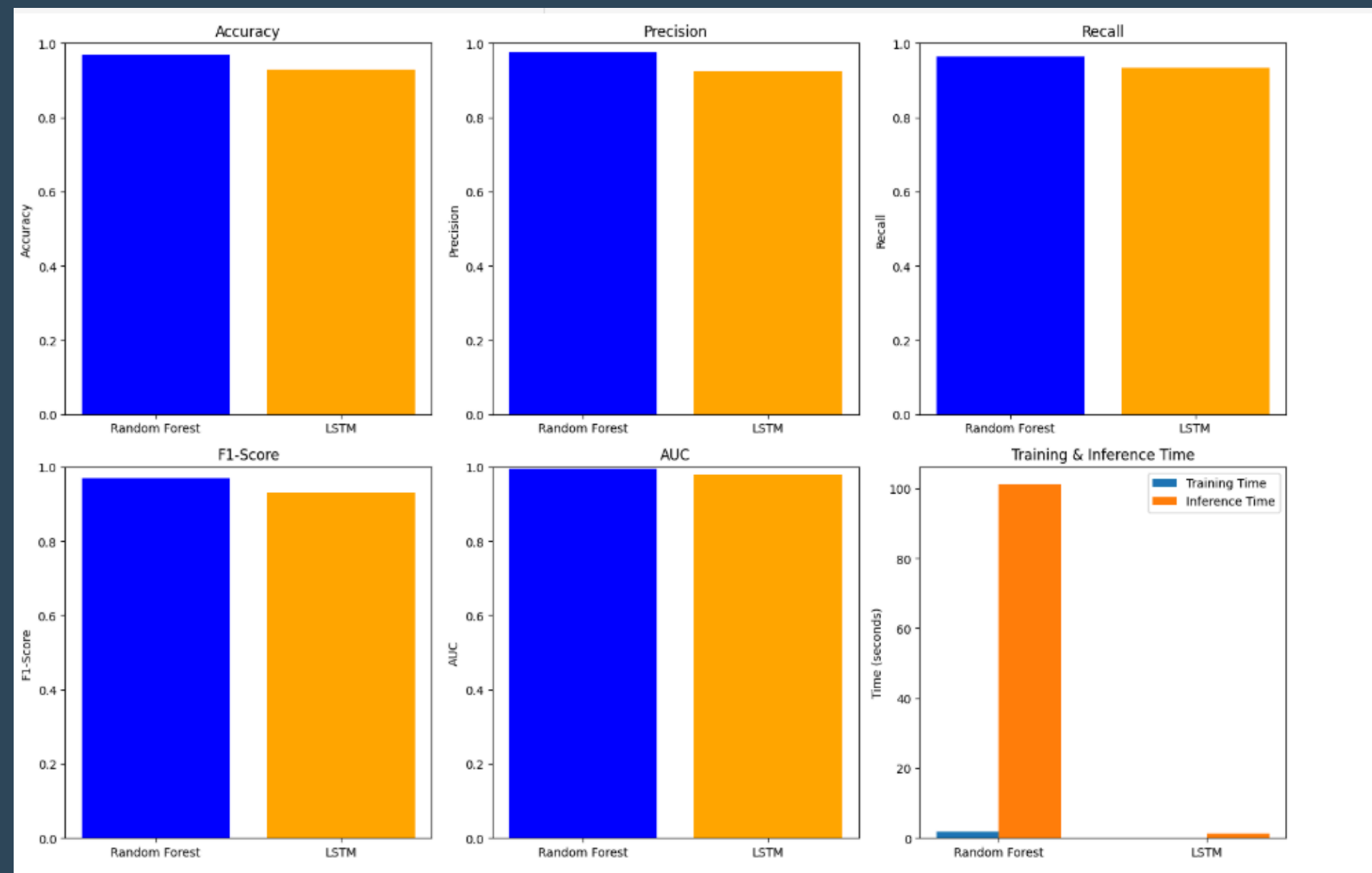
Training Time: 4.4499 seconds
Inference Time: 0.0323 seconds

LSTM

Training Time: 97.1390 seconds
Inference Time: 1.3317 seconds

Hasil Metrik Evaluasi

Grafik



Future Works

1. Ekspansi Dataset dan Variasi Fitur

- Memperluas dataset dengan menambahkan lebih banyak sampel URL dari berbagai sumber dan jenis.
- Menambahkan fitur lain seperti informasi tentang konten halaman, metadata, atau pola perilaku pengguna untuk meningkatkan akurasi model.

Future Works

2. Pengembangan Sistem Deteksi Realtime

- Mengembangkan sistem deteksi phishing secara realtime dengan memanfaatkan model LSTM dan Random Forest.
- Mengoptimalkan waktu inferensi agar cocok untuk aplikasi real-world dengan kebutuhan latensi rendah.

Future Works

3. Evaluasi di Berbagai Lingkungan

- Melakukan evaluasi model pada lingkungan berbeda, seperti aplikasi berbasis cloud, untuk mengetahui performa dan skalabilitasnya dalam skenario nyata.

Kesimpulan

Deteksi web page phishing menggunakan Random Forest dan LSTM adalah pendekatan efektif yang menggabungkan keunggulan seleksi fitur penting oleh Random Forest dan analisis pola kompleks oleh LSTM. Kombinasi ini meningkatkan akurasi dan adaptabilitas model dalam mengenali berbagai jenis phishing. Meskipun memiliki tantangan berupa biaya komputasi dan kompleksitas implementasi, pendekatan ini berpotensi menjadi solusi andal dalam melindungi pengguna dari ancaman phishing yang berkembang.



THANK YOU