KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
PROGRAM STUDI S1 TEKNOLOGI INFORMASI
Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: http://it.usu.ac.id

FORM PENGAJUAN JUDUL

Nama : Beatrics Sahcayani Purba

NIM : 211402115

Judul diajukan oleh* : ☐ Dosen
☑ Mahasiswa

Bidang Ilmu (tulis dua bidang) :
1. Data Science
2. Intelligent System

Uji Kelayakan Judul** : ⊘ Diterima ○ Ditolak

Hasil Uji Kelayakan Judul :

Calon Dosen Pembimbing I: Umaya Ramadhani Putri Nasution S.TI., M.Kom
(Jika judul dari dosen maka dosen tersebut berhak menjadi pembimbing I)

Calon Dosen Pembimbing II: Annisa Fadhillah Pulungan S.Kom., M.Kom

Paraf Calon Dosen Pembimbing I

Medan, 28 Februari 2025

Ka. Laboratorium Penelitian,

** Centang salah satu atau keduanya*

*** Pilih salah satu*

(Fanindia Purnamasari, S.TI., M.IT)

NIP.198908172019032023

Dipindai dengan CamScanner

KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
PROGRAM STUDI S1 TEKNOLOGI INFORMASI
Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: http://it.usu.ac.id

RINGKASAN JUDUL YANG DIAJUKAN

*Semua kolom di bawah ini diisi oleh mahasiswa yang sudah mendapat judul

Judul / Topik Skripsi	**IMPLEMENTASI ALGORITMA FP-GROWTH DAN RANDOM FOREST PADA SISTEM PENDETEKSIAN URL TERINDIKASI PHISHING**
Latar Belakang dan Penelitian Terdahulu	**Latar Belakang** Seiring dengan semakin berkembangnya teknologi pada era digital ini, banyak ancaman yang membuat masyarakat resah berkaitan dengan keamanan siber. Keamanan siber merupakan salah satu hal yang sangat krusial terhadap perlindungan data dan privasi pengguna. Dari banyaknya jenis keamanan siber yang mengancam perlindungan data dan privasi pengguna, salah satu bentuk ancaman yang sering ditemukan yakni *phishing*. Secara umum *phishing* dapat diartikan sebagai bentuk kejahatan digital yang bertujuan untuk mendapatkan data maupun informasi sensitif dari seseorang melalui email, pesan singkat atau situs website palsu yang dibuat mirip dengan sumber resminya (Putra, et al. 2022). Berdasarkan laporan yang disampaikan oleh Anti-Phishing Working Group (APWG) pada periode kuartal ketiga tahun 2024 tercatat sebanyak 932.923 kasus serangan *phishing* yang terjadi di seluruh dunia, nilai yang tercatat ini 6% lebih tinggi dari serangan yang terjadi pada kuartal sebelumnya. Ancaman *phishing* saat ini memiliki banyak bentuk dan tersebar secara luas, beberapa bentuknya berupa email, pesan singkat, telepon, bahkan dokumen yang secara sengaja disebar untuk mengelabui calon korban. Dari bentuk-bentuk yang telah disebutkan, salah satu jenis yang paling sering diterima yakni situs website (URL). Masyarakat secara umum masih sering melakukan klik pada sembarang tautan yang mengarah kepada situs web (URL) terindikasi phishing dan terkecoh dengan situs web palsu yang dibuat sangat mirip dengan situs web asli untuk mencuri data banyak orang. Berbagai upaya telah dilakukan untuk mencegah pencurian data melalui situs web yang terindikasi phishing. Beberapa upaya yang telah dilakukan diantaranya, analisis sintaksis URL dan penggunaan daftar hitam *(blacklist)*, namun kedua metode ini masih kurang efektif untuk mendeteksi *phishing* dikarenakan situs website yang sekarang beredar dibuat semirip mungkin dengan website resminya dan pendekatan *blacklist* tidak mampu mengenali tautan *phishing* baru yang belum terdaftar pada database (Mahalakshmi, et al. 2024). Selain itu, saat ini banyak dari antara situs website terindikasi *phishing* yang sudah menggunakan sertifikat keamanan yang sah sehingga menyebabkan semakin sulitnya untuk mendeteksi URL *phishing* dengan efektif dan secara *real-time*. Penerapan metode *machine learning* dapat menjadi solusi yang lebih efektif untuk mendeteksi situs website terindikasi *phishing*. Metode *machine learning* dinilai lebih mampu untuk menganalisis pola yang kompleks, mudah beradaptasi dengan ancaman baru, memungkinkan otomatisasi dan pemanfaatan berbagai sumber data. Beberapa penelitian telah berhasil dilakukan untuk menerapkan *machine learning* sebagai metode untuk pendeteksian URL *phishing*. Salah satu metode *machine learning* yang digunakan untuk pendeteksian *phishing* adalah metode klasifikasi yang bertujuan untuk mengkategorikan terhadap suatu URL terindikasi *phishing* atau tidak. Berbagai algoritma telah digunakan untuk proses pendeteksian URL *phishing*, seperti Support Vector Machine (SVM), Logistic Regression, XGBoost dan Random Forest. Penelitian yang dilakukan oleh Sindhu *et al.* (2021), telah berhasil menunjukkan

penerapan Random Forest, SVM dan Neural Network dengan Backpropagation mampu mendeteksi URL *phishing* dengan efektif.

Dari antara berbagai algoritma klasifikasi yang sudah diterapkan, algoritma Random Forest merupakan salah satu algoritma yang telah banyak diterapkan dan terbukti efektif untuk melakukan pendeteksian URL *phishing*. Berdasarkan penelitian yang dilakukan oleh Mahmud & Wirawan (2024) yang melakukan studi untuk membandingkan berbagai algoritma klasifikasi terhadap pendeteksian situs web *phishing*, diperoleh hasil bahwa Random Forest memberikan hasil yang paling baik dalam hal akurasi, presisi, dan recall diantara algoritma lainnya yang digunakan. Hasil yang sama juga diperoleh dari penelitian yang dilakukan oleh Putri & Wijayanto (2022) yang menunjukkan bahwa Random Forest memberikan performa terbaik untuk mengklasifikasikan situs website terindikasi *phishing* atau tidak dibandingkan dengan algoritma lainnya. Penelitian terhadap penggunaan Random Forest juga telah diimplementasikan dalam sistem yang nyata. Sistem website dibuat untuk memungkinkan pengguna memasukkan situs web dan mendapatkan hasil deteksi secara otomatis (Harahap, et al. 2024)

Untuk dapat meningkatkan performa pendeteksian situs web *phishing* dengan penerapan *machine learning*, beberapa penelitian bahkan sudah menggabungkan metode klasifikasi dengan metode lainnya. Penerapan metode aturan asosiasi (*association rule mining*) juga telah terbukti efektif digunakan untuk pendeteksian situs website terindikasi *phishing*. Teknik aturan asosiasi digunakan untuk menemukan hubungan atau pola tersembunyi antar fitur-fitur dalam dataset. Shawkat *et al.* (2021) mengembangkan algoritma *Modified FP-Growth (MFP-Growth)* yang lebih efisien dalam menemukan aturan asosiasi tanpa memerlukan pembentukan ulang *conditional subtrees*, sehingga lebih hemat waktu dan memori. Sementara itu, Tiwari & Arjariya (2021) membandingkan algoritma *FP-Tree* dan *Apriori* dalam mendeteksi *phishing*, menemukan bahwa *FP-Tree* lebih unggul dalam hal efisiensi dan akurasi.

Kombinasi yang dilakukan antara aturan asosiasi dengan klasifikasi merupakan pendekatan yang mampu mengoptimalkan akurasi dan efektivitas model untuk menemukan pola atau hubungan tersembunyi dari data karena aturan asosiasi dapat menemukan keterkaitan antar fitur dalam data, sementara itu klasifikasi akan menggunakan pola yang ditemukan tersebut untuk memberikan hasil prediksi yang lebih akurat. Penelitian yang dilakukan oleh Lin & Gao (2021) memanfaatkan aturan asosiasi untuk menyeleksi fitur dan interaksinya sebelum digunakan dalam algoritma klasifikasi. Hasilnya menunjukkan terjadinya peningkatan efisiensi serta akurasi model terutama untuk menangani dataset berukuran besar dan beragam jenis tugas klasifikasi. Penelitian lainnya yang mengembangkan metode Intelligent Association Classification (IAC) menggantikan ukuran *support* dan *confidence* dengan *Harmonic Mean*, sehingga mampu meningkatkan akurasi deteksi *phishing* (Al-Fayoumi, et al. 2020).

Berdasarkan berbagai penelitian yang telah dilakukan, penulis mengusulkan penerapan kombinasi metode asosiasi dan klasifikasi dengan algoritma FP-Growth dan Random Forest untuk mendeteksi situs web terindikasi *phishing*. Penulis memberikan judul untuk penelitian ini "Implementasi Algoritma FP-Growth dan Random Forest pada Sistem Pendeteksian URL Phishing".

Penelitian Terdahulu

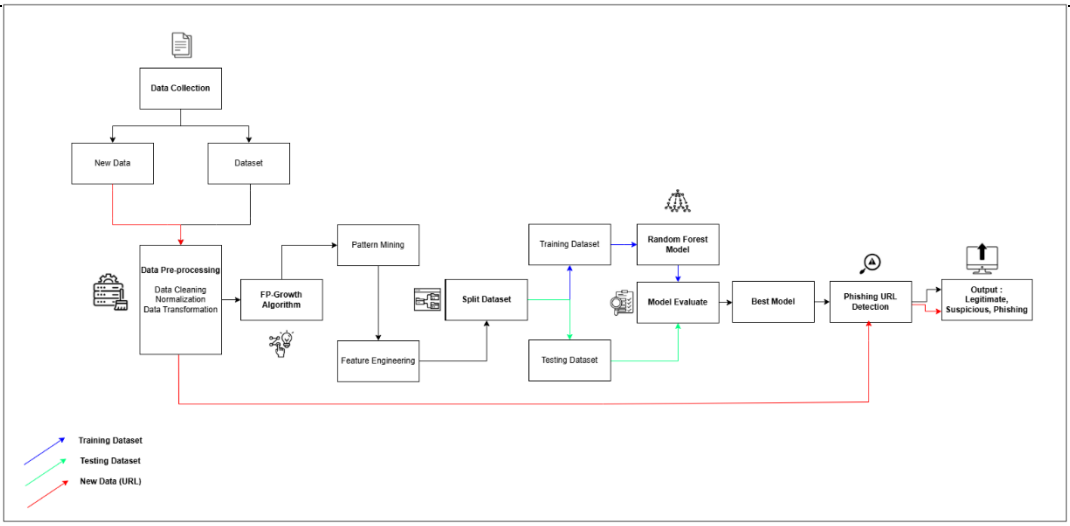
No.	Penulis	Judul	Tahun
1.	Mustafa Al-Fayoumi, Jaber Alwadian dan Mohammad Abusail	Intelligent Association Classification Technique for Phishing Website Detection	2020
2.	Mai Shawkat, Mahmoud Badawi, Sally El-Ghamrawy, Reham Arnous dan Ali El-Desoky	An optimized FP growth algorithm for discovery of association rules	2021
3.	Manish Tiwari dan Tripti Arjariya	A Phishing URL Classification Technique using Machine Learning Approach	2021
4.	Qiuqiang Lin dan Chuanou Gao	Discovering Categorical Main and Interaction Effects Based on Association Rule Mining	2021
5.	Simitha Sindhu, Arya Sreevalsan, Sunil Parameshwar Patil, Faiz Rahman dan Saritha A. N	Phishing Detection using Random Forest, SVM and Neural Network with Backpropagation	2021
6	Nabila Bianca Putri dan Arie Wahyu Wijayanto	Analisis Komparasi Algoritma Klasifikasi Data Mining dalam Klasifikasi Website Phishing	2022

	7	Ali Dongan Harahap, Didi Juardi dan Agung Susilo Yuda Irawan	Rancang Bangun Sistem Pendeteksi Link Phishing Menggunakan Algoritma Random Forest Berbasis Web	2024
	8	Azzam Fawwaz Mahmud dan Setia Wirawan	Deteksi Phishing Website menggunakan Machine Learning Metode Klasifikasi	2024

Kebaruan Penelitian:
Penelitian mengenai pendeteksian URL *phishing* dengan metode *machine learning* telah banyak dilakukan, tetapi sebagian besar hanya menggunakan satu pendekatan, yaitu klasifikasi. Keterbatasan ini menyebabkan tidak adanya proses optimasi dalam pemilihan fitur atau analisis hubungan antar fitur dalam dataset. Di sisi lain, penelitian yang menerapkan aturan asosiasi telah membuktikan kemampuannya dalam mengidentifikasi pola tersembunyi dalam URL *phishing*. Aturan asosiasi bermanfaat karena dapat mengungkap keterkaitan antar fitur yang mungkin tidak terdeteksi oleh metode klasifikasi murni, sehingga memungkinkan pemodelan yang lebih mendalam terhadap karakteristik URL *phishing*. Namun, masih sedikit penelitian yang menggabungkan aturan asosiasi dan klasifikasi, terutama dengan algoritma yang secara terpisah menangani proses pencarian pola dan klasifikasi. Oleh karena itu, kombinasi kedua metode ini diharapkan dapat menghasilkan sistem deteksi URL *phishing* yang lebih efektif dan akurat.

Rumusan Masalah	Serangan phishing yang semakin canggih menjadi ancaman serius bagi keamanan siber, terutama melalui situs web (URL) yang menipu pengguna. Mengidentifikasi URL *phishing* merupakan tantangan bagi para peneliti dan praktisi keamanan dalam mengembangkan sistem deteksi yang lebih akurat dan andal. Berbagai metode berbasis *machine learning* telah diterapkan, tetapi akurasi dan efektivitasnya masih perlu ditingkatkan. Oleh karena itu, diperlukan sistem yang mampu menemukan pola tersembunyi dalam data serta menghasilkan klasifikasi yang lebih tepat untuk mendeteksi *phishing* secara lebih optimal.

Metodologi



Tahapan Penelitian:

1. *Data Collection*
   Tahapan awal yang dilakukan yakni pengumpulan data yang akan digunakan untuk proses pendeteksian URL *phishing*. Data yang akan digunakan terdiri dari dua jenis data yakni *dataset* dan juga *new data*. Dataset merupakan data yang sudah diekstrak beserta dengan fitur-fitur yang akan digunakan untuk proses pembuatan model, sementara *new data* merupakan data berupa masukan berbentuk tautan atau URL dari *user* yang nantinya akan dideteksi. *Dataset* yang digunakan merupakan data tabular dengan nilai fiturnya berupa bilangan bulat *(integer)*. Fitur-fitur yang terdapat dalam *dataset* merupakan fitur-fitur yang mencakup karakteristik suatu URL *phishing*, seperti panjang URL, usia domain, IP address dan fitur pendukung lainnya.

2. *Data Pre-processing*
   Setelah data yang digunakan terkumpul, tahapan berikutnya adalah pre-processing data untuk memastikan data-data yang sudah terkumpul siap untuk diproses lebih lanjut dalam analisis dan pelatihan model. Hal-hal yang dilakukan dalam preprocessing data yakni pengecekan nilai yang hilang (missing value), menghilangkan data yang duplikat (duplicate), dan juga normalisasi. Selain itu, nantinya untuk data baru berupa tautan atau URL yang diberikan user akan dilakukan transformasi data untuk mengubah dan menyesuaikan URL tersebut sesuai dengan bentuk dan fitur dari dataset yang digunakan untuk pembuatan dan pelatihan model.

3. *FP-Growth Algorithm*
   Algoritma Frequent Pattern-Growth (FP Growth) adalah salah satu jenis algoritma yang merupakan penerapan dari aturan asosiasi. Tujuan dari penggunaan FP-Growth yakni untuk menemukan pola atau hubungan antar item yang ada di dalam dataset. FP-Growth membangun struktur seperti pohon yang akan mencari pola yang sering muncul tanpa melakukan iterasi berulang terhadap seluruh dataset. Proses penerapan FP-Growth ini juga akan memberikan *insight* baru dari dataset yang digunakan sehingga memungkinkan untuk adanya penyesuaian fitur yang relevan untuk digunakan atau penambahan fitur baru berdasarkan pola atau hubungan yang diperoleh melalui proses FP-Growth.

4. *Split Dataset*
   Setelah dataset secara keseluruhan sudah disesuaikan dengan hasil dari algoritma FP-Growth, dataset tersebut akan dibagi menjadi dua subset utama, yakni training dataset dan juga testing dataset. Proses pembagian ini bertujuan agar model nantinya dapat belajar dari data tanpa adanya *overfitting*. Secara umum, beberapa rasio yang sering digunakan seperti 80:20 atau 70:30, di mana training dataset bernilai lebih besar dikarenakan akan digunakan untuk melatih model sementara testing dataset digunakan untuk mengukur dan mengevaluasi kinerja model dari terhadap data yang belum digunakan sebelumnya.

5. *Random Forest Model*
   Model Random Forest merupakan salah satu model klasifikasi dengan basis pohon keputusan. Dengan training dataset yang sudah ditentukan sebelumnya, model ini dilatih untuk memberikan hasil deteksi terhadap tautan atau URL terindikasi *phishing*.

6. *Model Evaluate*
   Setelah model dilatih dengan training dataset, kemudian model akan diuji dan di evaluasi dengan menggunakan testing dataset. Metrik evaluasi yang akan digunakan diantaranya confusion matrix, akurasi, presisi, recall dan F1 score. Selain itu akan dilakukan juga pencarian terhadap Feature Importance dari keseluruhan fitur dalam dataset untuk mengetahui seberapa besar pengaruh tiap fitur terhadap hasil yang dikeluarkan. Dari proses Model Evaluate yang dilakukan akan menghasilkan satu model yang merupakan model terbaik untuk mendeteksi URL *phishing*.

7. *Output*
   Setelah model dievaluasi dan ditemukan model terbaik untuk mendeteksi, selanjutnya dilakukan proses *deploy* agar dapat digunakan untuk mendeteksi URL *phishing* secara *real-time*. Pada tahap ini data baru berupa tautan atau URL secara langsung sudah dapat dijadikan sebagai inputan yang akan diproses untuk menghasilkan deteksi. Hasil akhir yang ditampilkan yakni kategori deteksi berupa *legitimate*, *suspicious* dan juga *phishing*.

Referensi

Al-Fayoumi, M., Alwidian, J., & Abusaif, M. (2020). Intelligent association classification technique for phising website detection. *The International Arab Journal of Information Technology*, *17*(4), 488-496.

Anti-Phising Working Group. (2024). *Phishing activity trends* (3Q 2024). https://apwg.org/trendsreports/

Harahap, A. D., Juardi, D., & Irawan, A. S. (2024). Rancang bangun sistem pendeteksi link phishing menggunakan algoritma random forest berbasis web. *Jurnal Informatika dan Teknik Terapan Elektro*, *12*(3), 2677-2686.

Lin, Q., & Gao, C. (2023). Discovering categorical main and interaction effects based on association rule mining. *IEEE Transaction on Knowledge and Data Engineering*, *35*(2), 1379-1390.

KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
PROGRAM STUDI S1 TEKNOLOGI INFORMASI
Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: http://it.usu.ac.id

Mahalakshmi, S., Meena, P., & Gopinath, R. (2024). Detection of phishing website using machine learning. *International Journal of Research Publication and Reviews*, *5*(2), 1817-821.

Mahmud, A. F., & Wirawan, S. (2024). Deteksi phishing website menggunakan machine learning metode klasifikasi. *Sistemasi: Jurnal Sistem Informasi*, *13*(4), 1368-1380.

Putra, I. K., Darmawan, I. M., Juliana, I. P., & Indriyani. (2022). Tindak kejahatan pada dunia digital dalam bentuk phishing. *CyberSecurity dan Forensik Digital*, *5*(2), 77-82.

Putri, N. B., & Wijayanto, A. W. (2022). Analisis komparasi algoritma klasifikasi data mining dalam klasifikasi website phishing. *Komputika: Jurnal Sistem Komputer*, *11*(1), 59-66.

Shawkat, M., Badawi, M., El-ghamrawy, S., Arnous, R., & El-desoky, A. (2021). An optimized fp-growth algorithm for discovery of association rules. *The Journal of Supercomputing*, *78*(3), 5479-5506.

Sindhu, S., Patil, S. P., Sreevalsan, A., Rahman, F., & N, S. A. (2020). Phising detection using random forest svm and neural network with backpropagation. *International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE 2020)*, pp 391-394.

Tiwari, M., & Arjaria, T. (2021). A phishing url classification technique using machine learning approach. *International Journal of Innovative Technology and Exploring Engineering*, *10*(3), 73-79.

Medan, 28 Februari 2025
Mahasiswa yang mengajukan,

(Beatrics Sahcayani Purba)
NIM. 211402115