KLASIFIKASI MALICIOUS *URL* PADA *FILE* MENGGUNAKAN METODE *K-NEAREST NEIGHBOR* BERDASARKAN *LEXICAL FEATURE EXTRACTION*

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah Satu Syarat

Memperoleh Gelar Sarjana Komputer

OLEH :

RIZKI VALEN MAFAZA

09011281823134

JURUSAN SISTEM KOMPUTER

FAKULTAS ILMU KOMPUTER

UNIVERSITAS SRIWIJAYA

2023

LEMBAR PENGESAHAN

Klasifikasi Malicious *URL* pada *File* menggunakan Metode *K-Nearest Neighbor* berdasarkan *Lexical Feature Extraction*

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh

RIZKI VALEN MAFAZA
09011281823134

Indralaya, September 2023

Mengetahui,

Ketua Jurusan Sistem Komputer	**Pembimbing Tugas Akhir**
Dr. Ir. Sukemi, M.T.	**Ahmad Heryanto, S. Kom, M.T**
NIP. 196612032006041001	**NIP. 198701222015041002**

AUTHENTICATION PAGE

Classification of Malicious URLs In Files using the K-Nearest Neighbor Method based on Lexical Feature Extraction

FINAL TASK

Submitted To Fulfill One Of The Requirements
To Obtain A Bachelor's Degree in Computer Science

By

RIZKI VALEN MAFAZA
09011281823134

Indralaya, September 2023

Acknowledge,

Head of Computer System Deparment	*Supervisor*
Dr. Ir. Sukemi, M.T.	**Ahmad Heryanto, S. Kom, M.T.**
NIP. 196612032006041001	**NIP. 198701222015041002**

HALAMAN PERSETUJUAN

Telah diuji dan lulus pada :

Hari : Senin

Tanggal : 18 September 2023

Tim Penguji :

1. **Ketua** : **Dr. Ahmad Zarkasi, M.T.**
2. **Sekretaris** : **Nurul Afifah, M.Kom.**
3. **Pembimbing I** : **Ahmad Heryanto S.Kom., M.T.**
4. **Penguji** : **Ahmad Fali Oklilas, M.T.**

Mengetahui,

Ketua Jurusan Sistem Komputer

Dr.Ir. Sukemi, M.T.

NIP. 196612032006041001

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Rizki Valen Mafaza

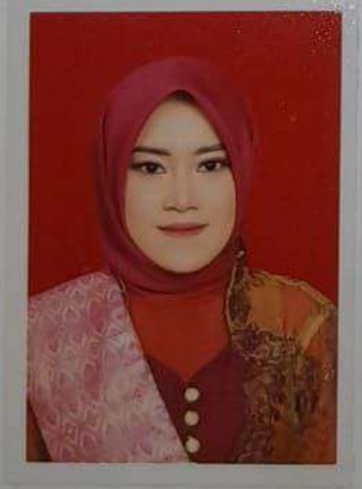
NIM : 09011281823134

Judul : Klasifikasi Malicious URL pada File menggunakan Metode K-Nearest Neighbor berdasarkan Lexical Feature Extraction

Hasil Pengecekan Software *iThenticate/Turnitin* : 7%

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan atau plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam laporan tugas akhir ini, maka saya bersedia menerima sanksi akademik dari universitas sriwijaya.

Demikian, pernyataan ini saya buat dalam keadaan sadar dan tanpa paksaan dari siapapun.



Indralaya, September 2023

Rizki Valen Mafaza

NIM.09011281823134

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Puji dan syukur atas kehadirat Allah SWT yang telah melimpahkan rahmat, kasih sayang dan karunia-Nya, sehingga penulis dapat menyelesaikan Proposal Tugas Akhir dengan judul "Klasifikasi *Malicious URL* pada *File* menggunakan Metode *K-Nearest Neighbor* berdasarkan *Lexical Feature Extraction*".

Selama penyusunan Tugas Akhir ini, penulis banyak mendapatkan ide, bantuan, serta saran dari semua pihak, baik secara langsung maupun tak langsung. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih dan rasa syukur yang sebesar – besarnya kepada:

1. Allah Subhanahu Wa Ta'ala yang telah melimpahkan berkah serta nikmat kesehatan dan kesempatan yang tak terhingga.
2. Kedua Orang Tua sebagai support system terbesar yang selalu mendoakan dan memberi dukungan.
3. Bapak Prof. Dr. Erwin., S.Si., M.Si., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Dr. Ir. Sukemi, M.T., selaku Ketua Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Bapak Ahmad Heryanto, S.Kom., M.T., selaku Dosen Pembimbing Tugas Akhir dan Pembimbing Akademik yang telah berkenan meluangkan waktunya untuk membimbing, memberikan saran dan motivasi serta bimbingan terbaik untuk penulis dalam menyelesaikan Tugas Akhir ini.
6. Kak Tri Wanda Septian,S.Kom.,M.Sc yang selalu memberikan masukan dan saran.
7. Mbak Renny Virgasari selaku admin Jurusan Sistem Komputer yang telah membantu mengurus seluruh berkas administrasi.

8. Teman – teman seperjuangan riset malicious URL, Rachmawati Dwinanti Putri, Muhammad Imam Rafi, dan Muhammad Andiko Putra yang telah membantu dan berjuang Bersama penulis dalam pengerjaan tugas ahkir ini.
9. Dimas Aditya Kristianto yang telah banyak membantu penulis dalam permasalahan coding selama perkuliahan dan tugas akhir.
10. Teman seperjuangan Indah Cahya Resti dan Alifah Fidela yang telah banyak membantu penulis selama proses perkuliahan dan proses skripsi.
11. Selvia Oktiriani, Ainil Qalbin, kak Farhan Furqan, kak Moh. Ardiansyah Ramadhan, dan Kak Bayu Catur sebagai tempat penulis berkeluh kesah dalam berbagai suka duka dan memberi semangat.
12. Teman – teman seperjuangan Sistem Komputer Angkatan 2018.
13. Seluruh pihak yang tidak dapat disebutkan satu persatu yang telah membantu dalam proses tugas akhir ini.
14. Serta seluruh civitas akademika Universitas Sriwijaya dan nama almamater Universitas Sriwijaya.

Penulis menyadari bahwa Tugas Akhir ini masih sangat jauh dari kata sempurna. Untuk itu penulis memohon maaf serta menerima kritik maupun saran sebagai evaluasi untuk di masa mendatang Akhir kata penulis berharap, semoga Tugas Akhir ini bermanfaat dan berguna bagi setiap pembaca Tugas Akhir ini.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Indralaya, September 2023

Penulis,

Rizki Valen Mafaza
09011281823134

KLASIFIKASI MALICIOUS URL PADA FILE MENGGUNAKAN METODE K-NEAREST NEIGHBOR BERDASARKAN LEXICAL FEATURE EXTRACTION

Rizki Valen Mafaza (09011281823134)
Department of Computer System, Faculty of Computer Science,
University of Sriwijaya
Palembang, Indonesia
Email : valenmafaza@gmail.com

ABSTRAK

Berbagai bentuk model penyerangan mulai dari hosting, penyebaran malware dan situs web phishing, tindakan tersebut dapat berawal dari mengakses *Uniform Resource Locator* (URL) atau file yang mengandung link berbahaya di dalamnya. *Uniform Resource Locator* (URL) adalah pengidentifikasi khusus yang digunakan untuk menemukan sumber melalui internet. Malicious URL dapat menjadi ancaman menggunakan berbagai jenis serangan, biasanya malicious URL disamarkan sehingga mudah untuk dilewatkan. Hal yang perlu dilakukan untuk membedakan malicious URL dan URLyang normal adalah menggunakan ektraksi fitur untuk mengidentifikasi berbagai karakteristik penting dari malicious URL. Fitur ekstraksi yang digunakan adalah lexical feature yang terdiri dari 18 fitur. Setelah dilakukan ekstraksi, hasil dari dataset tidak seimbang akan diproses resampling menggunakan oversampling dengan SMOTE. Penelitian ini menggunakan algoritma machine learning k-nearest neighbor untuk melakukan klasifikasi pada dataset. K-Nearest Neighbor adalah algoritma klasifikasi karakteristik berbeda yang menentukan kelas dimana milik data yang tidak berlabel menggunakan jarak untuk menghitung tetangga terdekat. Algoritma ini mampu mencapai akurasi klasifikasi yang tinggi serta memberikan hasil yang terbaik. Penelitian ini, memperoleh hasil evaluasi dengan nilai akurasi tertinggi sebesar 98,78%, presisi 98,785%, recall 98,795% dan f1-score 98,79%.

Kata Kunci : URL, Malicious URL, Lexical Feature, Synthetic Minority Over-sampling Technique (SMOTE), K-Nearest Neighbor, Machine Learning.

CLASSIFICATION OF MALICIOUS URLS IN FILES USING THE K-NEAREST NEIGHBOR METHOD BASED ON LEXICAL FEATURE EXTRACTION

Rizki Valen Mafaza (09011281823134)
Department of Computer System, Faculty of Computer Science,
University of Sriwijaya
Palembang, Indonesia
Email : valenmafaza@gmail.com

ABSTRACT

Various forms of attack models ranging from hosting, spreading malware and phishing websites, these actions can start from accessing the Uniform Resource Locator (URL) or files that contain malicious links in them. A Uniform Resource Locator (URL) is a special identifier used to find resources over the internet. Malicious URLs can be a threat using various types of attacks, usually malicious URLs are disguised so they are easy to miss. What needs to be done to differentiate between malicious URLs and normal URLs is to use feature extraction to identify various important characteristics of malicious URLs. The extraction feature used is a lexical feature which consists of 18 features. After extraction, the results of the unbalanced dataset will be resampled using oversampling with SMOTE. This research uses the k-nearest neighbor machine learning algorithm to classify the dataset. K-Nearest Neighbor is a different characteristic classification algorithm that determines the class to which unlabeled data belongs using distance to calculate the nearest neighbors. This algorithm is able to achieve high classification accuracy and provide the best results. This research obtained evaluation results with the highest accuracy value of 98.78%, precision 98.785%, recall 98.795% and f1-score 98.79%.

Keyword : *Uniform Resource Locator, Malicious URL, Lexical Features, Synthetic Minority Over-sampling Technique (SMOTE), K-Nearest Neighbor, Machine Learning.*

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB I

PENDAHULUAN

1.1. Latar Belakang

Saat ini website merupakan platform yang banyak digunakan untuk mendukung aktivitas sehari-hari yang semakin meningkat. Website memudahkan kegiatan sehari-hari dalam mencari informasi, sehingga semakin meningkatkan tingginya ketergantungan setiap orang terhadap situs website [1]. Hal ini memberikan peluang bagi para penjahat *cyber* untuk menyerang pengguna melalui URL *(Uniform Resource Locator)* atau halaman web yang dirancang khusus, *short-term web*, dan berbahaya. Biasanya *malicious* URL *(Uniform Resource Locator)* disebarluaskan melalui berbagai platform media sosial, pesan, *email*, *pop up*, iklan situs web, dan lainnya. URL *(Uniform Resource Locator)* ini berisi beberapa file berbahaya seperti virus (*ransomware*), *malware*, atau *keylogger*, untuk melakukan pencurian data pada computer pengguna[2].

Web telah digunakan sebagai pusat berbagai aktivitas jahat mulai dari *hosting* dan penyebaran *malware* hingga situs web phishing yang menipu pengguna untuk memberikan informasi pengguna pribadi mereka. *Malicious* URL dimaksudkan untuk tujuan jahat. Pengunjung URL tersebut berada di bawah ancaman menjadi korban serangan tertentu [3].

Pada penelitian [3] menunjukkan bahwa analisis lexical efektif dan efisien untuk mendeteksi malicious URL. Menurut laporan penelusuran Google tahun 2017, pencarian Google memasukkan lebih dari 50.000 situr malware dan lebih dari 90.0000 situs phishing setiap bulan ke dalam daftar hitam atau *blacklist*. Teknik *blacklist* merupakan pendekatan untuk menangani website berbahaya yang sederhana dengan memberikan akurasi yang baik. Dalam pendekatan teknik *blacklist* hanya efektif ketika daftar diperbarui tepat waktu

dan situs website dikunjungi secara ekstensif untuk menemukan website berbahaya, namun teknik ini gagal ketika memberikan perlindungan online yang tepat waktu pada pengguna. Pada penelitian [4] setiap bulan mencatat total serangan *phishing* sekitar 33.000 website yang menghasilkan kerugian sekitar $687 juta. Untuk mencegah serangan jenis ini, penulis menerapkan metode *Largest Common Substring* (LCS), dengan menggunakan metode LCS penulis menyebutkan mengenai cara suatu URL dapat digunakan untuk menggunakan sumber daya computer korban untuk melakukan berbagai jenis serangan seperti *phishing* dan *denial of service*. Penulis menggunakan 17 fitur untuk diekstraksi dari URL berdasarkan URL mana yang dinyatakan sebagai *phishing* atau tidak.

K-Nearest Neighbor (K-NN) merupakan algoritma *unsupervised machine learning* untuk klasifikasi yang digunakan di berbagai bidang, seperti *data mining, speech recognition, computer vision, text categorization, data compression, computational genomics,* dan analisis prediktif [5]. KNN mengklasifikasikan titik data baru berdasarkan kemiripan.

Pada [2] dengan mempertimbangkan waktu eksekusi dan akurasi, maka K-NN memberikan hasil yang lebih baik dibandingkan *Support Vector Classifier*. Ekstraksi fitur menjadi fase utama untuk teknik *machine learning.* Data dari 20 fitur yang dipilih dimasukkan ke dalam pengklasifikasi untuk pelatihan dan pengujian dengan rasio 70 : 30 menggunakan fitur lesikal. Pada [6] dengan menggunakan *confusion matrix* untuk mengukur dan mengevaluasi efektivitas metode *machine learning*, digunakan parameter tingkat akurasi, tingkat presisi, tingkat false negative, dan tingkat false positif. Hasilnya menunjukkan tingkat akurasi pendeteksian 94,16% dan tingkat presisi 93,33% dengan tingkat false positif dan tingkat false negative sedikit menurun.

Pada tugas akhir ini, penulis akan melakukan deteksi dari *malicious* URL yang diperoleh dari hasil ekstrak file menggunakan *pdf parser* untuk melihat URL dan digunakan sebagai dataset untuk masuk ke *machine learning* yaitu *K-Nearest Neighbor* (K-NN) agar dapat dideteksi dan menjadi bahan analisa yang dapat digunakan sebagai referensi. Untuk dapat mengenali

ciri dari *malicious* URL maka dilakukan ekstraksi informasi website berbahaya pada URL dengan *Lexical Feature Extraction*. Dan dilakukan klasifikasi dengan bantuan *machine learning* menggunakan algoritma *K-Nearest Neighbor*. Adapun pemberian judul dari tugas akhir ini adalah "Klasifikasi *Malicious URL* pada *File* menggunakan Metode *K-Nearest Neighbor* berdasarkan *Lexical Feature Extraction*".

1.2. Rumusan Masalah

Rumusan masalah dari penyusunan Tugas Akhir adalah sebagai berikut :

1. Bagaimana perbedaan karakteristik yang terdapat pada *benign* URL atau *malicious* URL berdasarkan fitur-fitur yang digunakan untuk mengekstraksi URL *Lexical Feature Extraction*?
2. Bagaimana cara penerapan klasifikasi URL berbahaya berupa *benign* dan *malicious* URL menggunakan algoritma *K-Nearest Neighbor*?
3. Bagaimana pengaruh hasil dari *benign* URL dan *malicious* URL berdasarkan fitur-fitur dari *Lexical Feature Extraction* menggunakan *K-Nearest Neighbor Classifier*?

1.3. Tujuan

Tujuan dari penyusunan Tugas Akhir adalah sebagai berikut :

1. Menerapkan algoritma *classifier* terhadap URL yang dikategorikan berdasarkan jenis *benign URL* dan *malicious URL*.
2. Mengimplementasikan *Lexical Feature Extraction* untuk mendeteksi *URL* yang tergolong *malicious* atau *benign*.
3. Mengimplementasikan metode *K-Nearest Neighbor* untuk klasifikasi antara *malicious* dan *benign URL*.

1.4. Manfaat

Manfaat dari penyusunan Tugas Akhir adalah sebagai berikut :

1. Memahami ciri-ciri dari *malicious URL* dan *benign URL*.
2. Memahami proses dari *Lexical Feature Extraction* untuk mengekstraksi *URL*.
3. Memberikan informasi mengenai kemampuan deteksi *malicious URL* dari algoritma *K-Nearest Neighbor.*

1.5. Batasan Masalah

Batasan masalah dari penyusunan Tugas Akhir adalah sebagai berikut :

1. Penelitian dilakukan untuk mengklasifikasikan *Malicious* URL dan *Benign* URL.
2. Metode yang diterapkan menggunakan algoritma *K-Nearest Neighbor Classifier*.
3. Penelitian ini membahas mengenai bagaimana cara mendeteksi *Malicious* URL menggunakan *Lexical Feature Extraction*.

1.6. Metodologi Penelitian

1. Metode Studi Pustaka dan Literature

   Tahap pertama dilakukan untuk mencari dan mengumpulkan referensi berupa studi literatur yang terdapat pada jurnal, buku, dan internet yang berkaitan dalam pembuatan tugas akhir.

2. Metode Pengolahan Data

   Metode ini melakukan proses pembuatan data mentah menjadi data siap olah.

3. Metode Pengujian

   Metode ini melakukan pengujian terhadap rancangan model agar mendapatkan hasil uji yang akurat dengan algoritma *K-Nearest Neighbor*.

4. Metode Hasil dan Analisa

   Hasil dari pengujian pada tugas akhir ini akan dianalisis kekurangannya agar dapat menghasilkan nilai yang objektif dari proses pengujian data yang telah diperoleh.

5. Metode Kesimpulan dan Saran

   Metode ini memberi kesimpulan dari rumusan masalah, metodologi, dan Analisa hasil pengujian. Terdapat juga saran untuk penelitian selanjutnya.

1.7. Sistematika Penulisan

Adapun sistematika penulisan dalam Tugas Akhir adalah sebagai berikut :

BAB I.PENDAHULUAN

Pada Bab ini berisi tentang landasan topik penelitian yang meliputi Latar Belakang, Rumusan Masalah, Batasan Masalah, Tujuan, dan Manfaat, serta termasuk Metodologi Penelitian dan Sistematika Penulisan.

BAB II. TINJAUAN PUSTAKA

Pada Bab ini menjelaskan dasar teori dari penelitian tugas akhir tentang *Malicious URL, K-Nearest Neighbor Classifier*, serta teori yang berhubungan dengan penelitian.

BAB III. METODOLOGI PENELITIAN

Pada Bab III akan melakukan rancangan ataupun rincian sistematis terhadap penelitian yang akan dilakukan. Rincian mengenai kerangka kerja penelitian, tahapan pemrosesan data, hingga penerapan algoritma *K-Nearest Neighbor*.

BAB IV. ANALISA DAN PEMBAHASAN

Pada Bab ini akan membahas dan menganalisa hasil yang telah diuji dan diperoleh dari tahap sebelumnya serta validasi hasil agar mendapatkan data yang akurat.

BAB V. KESIMPULAN DAN SARAN

Pada Bab terakhir ini akan menuliskan kesimpulan yang didapatkan selama proses penelitian sebagai jawaban dari target yang akan dicapai, serta saran yang diharapkan dapat dikembangkan lebih baik lagi.

DAFTAR PUSTAKA

[1] W. Yang, W. Zuo, and B. Cui, "Detecting Malicious URLs via a Keyword-Based Convolutional Gated-Recurrent-Unit Neural Network," *IEEE Access*, vol. 7, pp. 29891–29900, 2019, doi: 10.1109/ACCESS.2019.2895751.

[2] A. Saleem Raja, R. Vinodini, and A. Kavitha, "Lexical features based malicious URL detection using machine learning techniques," *Mater. Today Proc.*, vol. 47, no. xxxx, pp. 163–166, 2021, doi: 10.1016/j.matpr.2021.04.041.

[3] G. S. B, C. Tang, W. Gao, and Y. Yin, "MD- VC M atrix : An Efficient Scheme for Publicly Verifiable Computation," vol. 1, pp. 349–362, 2016, doi: 10.1007/978-3-319-46298-1.

[4] A. Desai, J. Jatakia, R. Naik, and N. Raul, "Malicious web content detection using machine leaning," *RTEICT 2017 - 2nd IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol. Proc.*, vol. 2018-Janua, pp. 1432–1436, 2017, doi: 10.1109/RTEICT.2017.8256834.

[5] J. Vieira, R. P. Duarte, and H. C. Neto, "Knn-stuff: Knn streaming unit for fpgas," *IEEE Access*, vol. 7, pp. 170864–170877, 2019, doi: 10.1109/ACCESS.2019.2955864.

[6] H. Zhao, Z. Chang, W. Wang, and X. Zeng, "Malicious Domain Names Detection Algorithm Based on Lexical Analysis and Feature Quantification," *IEEE Access*, vol. 7, pp. 128990–128999, 2019, doi: 10.1109/ACCESS.2019.2940554.

[7] T. Li, G. Kou, and Y. Peng, "Improving malicious URLs detection via feature engineering: Linear and nonlinear space transformation methods," *Inf. Syst.*, vol. 91, p. 101494, 2020, doi: 10.1016/j.is.2020.101494.

[8] G. Palaniappan, S. Sangeetha, B. Rajendran, S. Goyal, and B. S. Bindhumadhava, "ScienceDirect ScienceDirect Malicious Domain Detection Using Machine Learning On Domain Name Features , Features

and Web-Based Features Malicious Domain Host-Based Detection Using Machine Learning On Domain Name Features , and a Web-Based Features," *Procedia Comput. Sci.*, vol. 171, no. 2019, pp. 654–661, 2020, doi: 10.1016/j.procs.2020.04.071.

[9] G. A. Sandag, J. Leopold, and V. F. Ong, "Klasifikasi Malicious Websites Menggunakan Algoritma K-NN Berdasarkan Application Layers dan Network Characteristics," *CogITo Smart J.*, vol. 4, no. 1, p. 37, 2018, doi: 10.31154/cogito.v4i1.100.37-45.

[10] M. Yunus, D. Widiastuti, H. Rasjid, and Y. Chalri, "Metode Klasifikasi Untuk Deteksi Uniform Resource Locator (URL) Berdasarkan Jenis Serangan Menggunakan Algoritma Naive Bayes, C4.5 dan K-Nearest Neighbor," *Semin. Nas. Teknol. Inf. dan Komun. STI&K*, vol. 3, no. 1, 2019.

[11] T. A. Assegie, "K-Nearest Neighbor Based URL Identification Model for Phishing Attack Detection," no. April 2021, 2022, doi: 10.35940/ijainn.B1019.041221.

[12] H. Choi, B. B. Zhu, and H. Lee, "Detecting Malicious Web Links and Identifying Their Attack Types."

[13] S. Arabia, S. Arabia, S. Arabia, and S. Arabia, "Detecting Malicious URL," pp. 0–4, 2020.

[14] B. Cui, S. He, P. Shi, and X. Yao, "Malicious URL detection with feature extraction based on machine learning," *Int. J. High Perform. Comput. Netw.*, vol. 12, no. 2, pp. 166–178, 2018, doi: 10.1504/ijhpcn.2018.094367.

[15] F. Vanhoenshoven, N. Gonzalo, R. Falcon, K. Vanhoof, and K. Mario, "Detecting Malicious URLs using Machine Learning Techniques," 2016.

[16] C. Do Xuan, H. D. Nguyen, and T. V. Nikolaevich, "Malicious URL detection based on machine learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 1, pp. 148–153, 2020, doi: 10.14569/ijacsa.2020.0110119.

[17] H. Kumar, P. Gupta, and R. P. Mahapatra, "Protocol based ensemble classifier for malicious URL detection," *Proc. 3rd Int. Conf. Contemp.*

Comput. Informatics, IC3I 2018, pp. 331–336, 2018, doi: 10.1109/IC3I44769.2018.9007255.

[18] R. Siringoringo, "KLASIFIKASI DATA TIDAK SEIMBANG MENGGUNAKAN ALGORITMA SMOTE DAN k-NEAREST NEIGHBOR," *J. ISD*, vol. 3, no. 1, pp. 44–49, 2018.

[19] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A brief review of nearest neighbor algorithm for learning and classification," *2019 Int. Conf. Intell. Comput. Control Syst. ICCS 2019*, no. January, pp. 1255–1260, 2019, doi: 10.1109/ICCS45141.2019.9065747.

[20] E. M. F. El Houby, N. I. R. Yassin, and S. Omran, "A hybrid approach from ant colony optimization and K-nearest neighbor for classifying datasets using selected features," *Inform.*, vol. 41, no. 4, pp. 495–506, 2017.

[21] A. Altaher, "Phishing Websites Classification using Hybrid SVM and KNN Approach," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 6, pp. 90–95, 2017, doi: 10.14569/ijacsa.2017.080611.

[22] "Detecting Malicious URLs," 2019.

[23] R. Verma and A. Das, "What's in a URL: Fast feature extraction and malicious URL detection," *IWSPA 2017 - Proc. 3rd ACM Int. Work. Secur. Priv. Anal. co-located with CODASPY 2017*, pp. 55–63, 2017, doi: 10.1145/3041008.3041016.

[24] A. Altaher, "Klasifikasi Website Phishing menggunakan Pendekatan Hybrid SVM dan KNN," vol. 8, no. 6, pp. 90–95, 2017.

[25] D. Bajpai and L. He, "Evaluating KNN Performance on WESAD Dataset," *Proc. - 2020 12th Int. Conf. Comput. Intell. Commun. Networks, CICN 2020*, pp. 60–62, 2020, doi: 10.1109/CICN49253.2020.9242568.

[26] J. Ispahany and R. Islam, "Detecting malicious COVID-19 URLs using machine learning techniques," *2021 IEEE Int. Conf. Pervasive Comput. Commun. Work. other Affil. Events, PerCom Work. 2021*, pp. 718–723, 2021, doi: 10.1109/PerComWorkshops51409.2021.9431064.