

**SISTEM PEMENDEK URL BERBASIS PRETRAINED LANGUAGE
MODEL-GUIDED MULTI-LEVEL ATTENTION NETWORK
UNTUK PENDETEKSIAN URL BERBAHAYA**

SKRIPSI



Disusun Oleh:

NOVIN ARDIAN YULIANTO

09020621039

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA**

2025

PERNYATAAN KEASLIAN

PERNYATAAN KEASLIAN

Saya yang bertanda tangan dibawah ini :

Nama : Novin Ardian Yulianto
NIM : 09020621039
Program Studi : Sistem Infomasi
Angkatan : 2021

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul “SISTEM PEMENDEK URL BERBASIS *PRETRAINED LANGUAGE MODEL-GUIDED MULTI-LEVEL ATTENTION NETWORK* UNTUK PENDETEKSIAN URL BERBAHAYA”. Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya

Surabaya, 24 Februari 2025

Yang menyatakan,



Novin Ardian Yulianto

NIM 09020621039

LEMBAR PERSETUJUAN PEMBIMBING

LEMBAR PERSETUJUAN PEMBIMBING

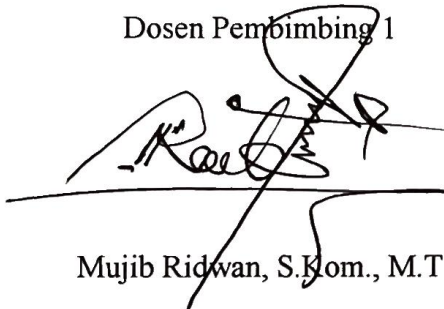
Skripsi oleh :

NAMA : Novin Ardian Yulianto
NIM : 09020621039
JUDUL : PENGEMBANGAN SISTEM PEMENDEK UNIFORM
RESOURCE LOCATOR DENGAN INTEGRASI
PENDETEKSIAN URL BERBAHAYA

Ini telah diperiksa dan disetujui untuk diujikan.

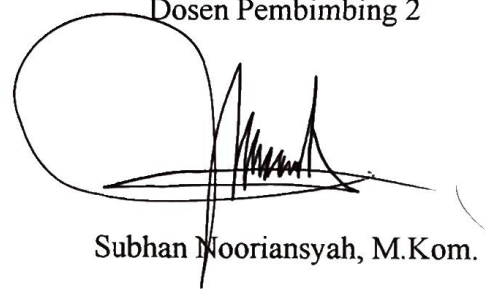
Surabaya, 18 Oktober 2024

Dosen Pembimbing 1



Mujib Ridwan, S.Kom., M.T
NIP. 198604272014031004

Dosen Pembimbing 2



Subhan Nooriansyah, M.Kom.
NIP. 199012282020121010

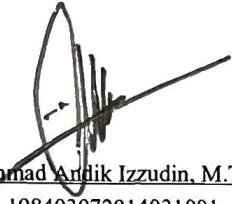
PENGESAHAN TIM PENGUJI SKRIPSI

PENGESAHAN TIM PENGUJI SKRIPSI

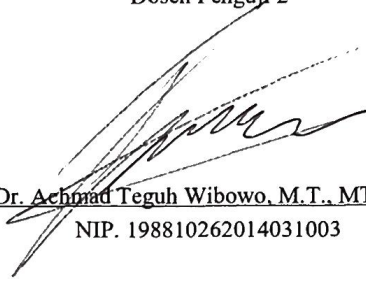
Skripsi Novin Ardian Yulianto ini telah
dipertahankan di depan Tim Penguji Skripsi
di Surabaya, 11 Maret 2025

Mengesahkan,
Dewan Penguji

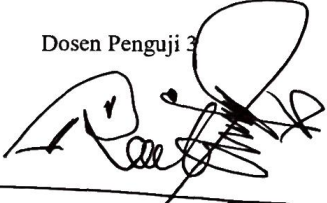
Dosen Penguji 1


Muhammad Ardik Izzudin, M.T
NIP. 198403072014031001

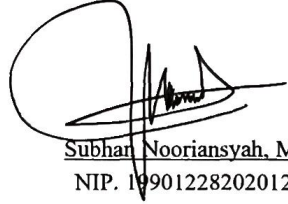
Dosen Penguji 2


Dr. Achmad Teguh Wibowo, M.T., MTCNA
NIP. 198810262014031003

Dosen Penguji 3


Mujib Ridwan, S.Kom., M.T
NIP. 198604272014031004

Dosen Penguji 4


Subhan Nooriansyah, M.Kom
NIP. 199012282020121010

Mengetahui,


Dekan Fakultas Sains dan Teknologi
UIN Sunan Ampel Surabaya
Supul Hamdani, M. Pd
NIP. 196507312000031002



**KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN**

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpus@uinsby.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : Novin Ardian Yulianto

NIM : 09020621039

Fakultas / Jurusan : SAINS DAN TEKNOLOGI / SISTEM INFORMASI

E-mail Address : work.novinnn@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

☒ Sekripsi ☐ Tesis ☐ Desertasi ☐ Lain-lain (.....)

yang berjudul :

SISTEM PEMENDEK URL BERBASIS PRETRAINED LANGUAGE MODEL-GUIDED MULTI-LEVEL
ATTENTION NETWORK UNTUK PENDETEKSIAN URL BERBAHAYA

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara fulltext untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 10 April 2025
Penulis

Novin Ardian Yulianto

ABSTRAK

SISTEM PEMENDEK URL BERBASIS PRETRAINED LANGUAGE MODEL-GUIDED MULTI-LEVEL ATTENTION NETWORK UNTUK PENDETEKSIAN URL BERBAHAYA

Oleh :

Novin Ardian Yulianto

Pada era digital yang semakin pesat, internet menjadi alat esensial dalam berbagai aspek kehidupan, seperti komunikasi, bisnis, dan pendidikan. Meskipun internet menawarkan kemudahan dalam pertukaran informasi, tetapi dapat mengakibatkan risiko keamanan siber. Salah satu contoh serangan yang memanfaatkan penggunaan pemendek URL yang menghasilkan penyebaran *malware* dan *phishing*. Hal tersebut berdampak kerusakan dan kehilangan data bagi pengunjung atau pengguna pemendek URL karena berbahayanya URL. Penelitian ini bertujuan mengembangkan sistem pemendek URL yang berintegrasi dengan fitur deteksi keamanan berbasis *deep learning*. Sistem ini akan menganalisis atribut URL seperti struktur domain, pola karakter, dan reputasi situs untuk memprediksi potensi ancaman. Dengan integrasi sistem ini, pengguna dapat memperoleh manfaat dari URL yang lebih ringkas sekaligus rasa aman dalam berbagi dan mengakses tautan. Sistem ini diharapkan mampu mengurangi resiko serangan siber berbasis URL pendek dan memberikan proteksi yang lebih baik bagi pengguna individu maupun organisasi.

Kata Kunci: Pemendek URL, Deteksi URL, *Deep Learning*, *SDLC*, *Waterfall*

ABSTRACT

URL SHORTENING SYSTEM BASED ON PRETRAINED LANGUAGE MODEL-GUIDED MULTI-LEVEL ATTENTION NETWORK FOR MALICIOUS URL DETECTION.

By :

Novin Ardian Yulianto

In the fast-paced digital era, the internet has become an essential tool in various aspects of life, such as communication, business, and education. While the internet offers convenience in information exchange, it can also result in cybersecurity risks. One example of an attack that utilizes the use of URL shorteners results in the spread of malware and phishing. This results in damage and data loss for visitors or users of the URL shortener due to the danger of the URL. This research aims to develop a URL shortener system that integrates with deep learning-based security detection features. The system will analyze URL attributes such as domain structure, character patterns, and site reputation to predict potential threats. With this system integration, users can benefit from more concise URLs as well as a sense of security in sharing and accessing links. The system is expected to reduce the risk of short URL-based cyberattacks and provide better protection for individual users and organizations.

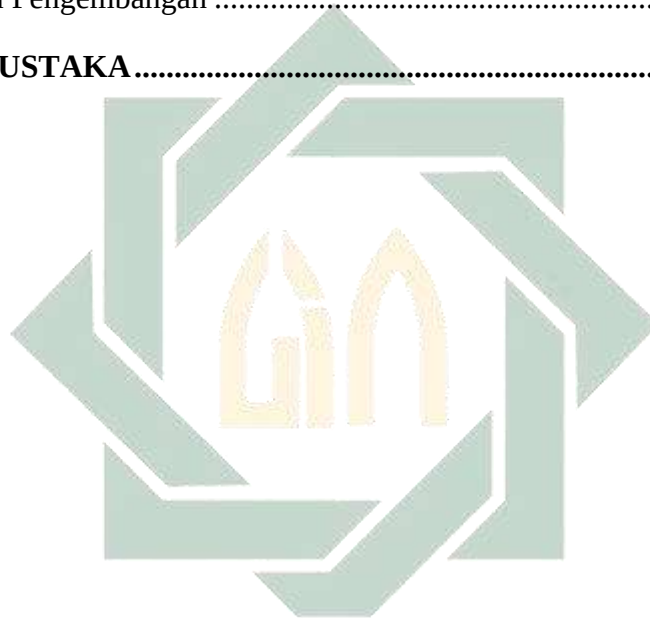
Keywords: URL Shortener, URL Detection, Deep Learning, SDLC, Waterfall

DAFTAR ISI

LEMBAR PERSETUJUAN PEMBIMBING.....	ii
PENGESAHAN TIM PENGUJI SKRIPSI	iii
PERNYATAAN KEASLIAN	iv
MOTO	v
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR.....	viii
ABSTRAK	x
ABSTRACT	xi
DAFTAR ISI.....	xii
DAFTAR GAMBAR.....	xv
DAFTAR TABEL	xvii
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian.....	4
1.5. Manfaat Penelitian.....	4
1.6. Sistematika Penulisan.....	5
BAB 2 TINJAUAN PUSTAKA.....	7
2.1. Penelitian Terdahulu	7
2.2. Dasar Teori	10
2.2.1. Sistem Informasi	10
2.2.2. Uniform Resource Locator.....	11
2.2.3. URL Pendek	11
2.2.4. Pemendekan URL	12
2.2.5. Deteksi Keamanan URL	13
2.2.6. Deep Learning untuk Deteksi Keamanan URL	14
2.2.7. Website	18
2.2.8. Waterfall.....	19
2.2.9. Unified Modelling Language	20
2.2.10. Node JS	24

2.2.11. Express JS	25
2.2.12. PostgreSQL	25
2.2.13. Prisma ORM	25
2.2.14. React JS	26
2.2.15. Next JS	26
2.2.16. Tailwind CSS	26
2.3. Integrasi Keilmuan	27
BAB 3 Metodologi penelitian	29
3.1. Pendekatan Penelitian	29
3.2. Objek Penelitian	30
3.3. Teknik Pengumpulan Data	30
3.3.1. Studi Literatur	30
3.3.2. Wawancara	30
3.4. Teknik Pengembangan Sistem	31
3.4.1. Pengumpulan Kebutuhan	31
3.4.2. Desain Sistem	31
3.4.3. Implementasi Sistem	31
3.4.4. Pengujian Sistem	32
3.4.5. Evaluasi Sistem	34
3.4.6. Maintenance Sistem	34
3.5. Peralatan Pendukung	35
3.5.1. Perangkat Lunak	35
3.5.2. Perangkat Keras	36
3.6. Teknik Analisa Data	36
BAB 4 Hasil dan Pembahasan	38
4.1. Requirement Analysis	38
4.1.1. Analisis kebutuhan fungsional	38
4.1.2. Analisis kebutuhan non fungsional	43
4.2. System and Software Design	43
4.2.1. Struktur Navigasi	43
4.2.2. Use Case Diagram	45
4.2.3. Activity Diagram	46
4.2.4. Sequence Diagram	52
4.2.5. Class Diagram	56
4.2.6. C4 Model	58
4.3. Implementasi	64
4.3.1. Desain Interface	64
4.3.2. Implementasi Deteksi URL	70
4.3.3. Implementasi Pemendek URL	77
4.4. Testing	79
4.4.1. Testing Deteksi URL	79

4.4.2. Testing Website Pemendek URL	82
4.5. Deployment	89
4.5.1. Proses CI/CD.....	90
4.5.2. Deployment Frontend	90
4.5.3. Deployment Backend.....	91
4.5.4. Deployment Komponen Deep Learning	91
4.5.5. Integrasi Keseluruhan.....	91
4.6. Maintenance	92
BAB 5 Penutup	94
5.1. Kesimpulan.....	94
5.2. Saran Pengembangan	94
DAFTAR PUSTAKA.....	95



UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR GAMBAR

Gambar 2. 1 Komponen Sistem Informasi.....	10
Gambar 2. 2 Struktur URL.....	11
Gambar 2. 3 Konversi URL Panjang ke Pendek.....	12
Gambar 2. 4 Website di berbagai Perangkat.....	18
Gambar 2. 5 Metode Waterfall pada Pengembangan Aplikasi	19
Gambar 2. 6 Use Case Diagram.....	21
Gambar 2. 7 Activity Diagram.....	22
Gambar 2. 8 Sequence Diagram.....	23
Gambar 2. 9 Class Diagram	24
Gambar 3. 1 Metode Penelitian.....	29
Gambar 4. 1 Struktur Navigasi	45
Gambar 4. 2 Use Case Diagram.....	46
Gambar 4. 3 Activity Diagram Autentikasi dengan Email & Password.....	47
Gambar 4. 4 Activity Diagram Autentikasi dengan Google OAuth	48
Gambar 4. 5 Activity Diagram Membuat URL Pendek.....	49
Gambar 4. 6 Activity Diagram Melihat Data Analitik URL.....	50
Gambar 4. 7 Activity Diagram Mengupdate Profil Pengguna	51
Gambar 4. 8 Sequence Diagram Autentikasi Email dan Password	52
Gambar 4. 9 Sequence Diagram Autentikasi dengan Google OAuth.....	53
Gambar 4. 10 Sequence Diagram Pembuatan URL Pendek.....	54
Gambar 4. 11 Sequence Diagram Melihat Data Analitik URL	55
Gambar 4. 12 Sequence Melihat dan Mengupdate Profil Pengguna	56
Gambar 4. 13 Class Diagram	57
Gambar 4. 14 Context Diagram	58
Gambar 4. 15 Container Diagram	59
Gambar 4. 16 Component Diagram	61
Gambar 4. 17 Deployment Diagram	62
Gambar 4.18 Desain Halaman Utama.....	64
Gambar 4. 19 Design Halaman <i>SignIn</i>	65

Gambar 4. 20 Design Halaman <i>SignUp</i>	66
Gambar 4. 21 Design Halaman Dashboard.....	67
Gambar 4. 22 Design Halaman untuk Memendekkan URL	68
Gambar 4. 23 Design Halaman untuk Mengisi UTM.....	69
Gambar 4. 24 Design Halaman untuk Mengubah <i>Profile</i>	69
Gambar 4. 25 Hasil Pelatihan Deep Learning.....	76



UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	7
Tabel 3. 1 Skenario Pengujian Black Box	33
Tabel 3. 2 Perangkat Lunak Pengembangan	35
Tabel 3. 3 Perangkat Lunak Pengembangan	36
Tabel 4. 1 Kebutuhan Fungsional User.....	39
Tabel 4. 2 Kebutuhan Non Fungsional	43
Tabel 4. 3 Sampel Dataset untuk Training.....	71
Tabel 4. 4 Sampel Dataset untuk Pemblokiran	71
Tabel 4. 5 Implementasi Training Deep Learning	74
Tabel 4. 6 Implementasi Validasi Deep Learning.....	75
Tabel 4. 7 Implementasi Pemendakan URL	77
Tabel 4. 8 Implementasi Otomatis Pembuatan Short Code	78
Tabel 4. 9 Implementasi Pengecekan Unik untuk Short Code	78
Tabel 4. 10 Hasil Batch Test Model Deteksi URL	80
Tabel 4. 11 Hasil Pengujian White Box.....	83
Tabel 4. 12 Hasil Pengujian Black Box	84
Tabel 4. 13 Hasil Pengujian Non Fungsional.....	88

UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR PUSTAKA

- Achmad, A.K., 2020, 'REAKTUALISASI PENDIDIKAN AL ISLAM DAN KEMUHAMMADIYAHAN (AIK) SEBAGAI PENGUAT PENDIDIKAN KARAKTER', *Muhammadiyah University Press*, 21(2), 167–178.
- Al-Haija, Q.A. & Al-Fayoumi, M., 2023, 'An intelligent identification and classification system for malicious uniform resource locators (URLs)', *Springer Science+Business Media*, 35(23), 16995–17011.
- Al-Milli, N. & Hammo, B., 2020a, 'A Convolutional Neural Network Model to Detect Illegitimate URLs'.
- Anugrah, I.Y., Akbar, R.R. El & Rahmatulloh, A., 2022, 'Monitoring Elementary School Students Achievement using Android-based RESTful Web Service', 11(1), 84–84.
- Arifwidodo, S.D. & Chandrasiri, O., 2015, 'Urban Heat Island and Household Energy Consumption in Bangkok, Thailand', *Elsevier BV*, 79, 189–194.
- Asnawi, N., Pamungkas, R. & Prasetyo, D.G., 2023, 'Analisis Usability Website Program Studi Sistem Informasi Unipma Menggunakan Metode System Usability Scale', *University of Darussalam Gontor*, 8(1), 21–25.
- Bashir, N., Bilal, M., Liaqat, M., Marjani, M., Malik, N. & Ali, M., 2021, 'Modeling Class Diagram using NLP in Object-Oriented Designing'.
- Budiarto, D.S., Meylina, A.P. & Diansari, R.E., 2023, 'Pentingnya E-Commerce Dan Sistem Informasi Dalam Pengambilan Keputusan Untuk Berwirausaha', 26(1), 110–122.
- Cahyadi, D., Tiara, K. & Mayosi, D.R., 2020, 'Website Sebagai Alat Promosi Online Dalam Upaya Merekrut CMB Pada Universitas Raharja', 6(1), 80–89.

- Hartono, R. & Ramadhan, T.I., 2022, 'Implementasi Metode User Centered Design (UCD) dengan Framework Kanban dalam Membangun Desain Interaksi', 19(2), 823–831.
- Hendriana, D., 2023, 'PERAN ILMU PENGETAHUAN DAN PENGARUH KEMAJUAN TEKNOLOGI DIGITAL DALAM PELAKSANAAN TUGAS KEKHALIFAHAN MANUSIA', *Universitas Muhammdiyah Magelang*, 19(1).
- Herdiana, Y., Munawar, Z. & Putri, N.I., 2021, 'Mitigasi Ancaman Resiko Keamanan Siber Di Masa Pandemi Covid-19', 20(1), 42–52.
- Hunt, P., O'Shannessy, P., Smith, D.T. & Coatta, T., 2016, 'React: Facebook's Functional Turn on Writing JavaScript', *Association for Computing Machinery*, 14(4), 96–112.
- Iqbal, L.M., Septiana, Y. & Setiawan, R., 2020, 'Rancang Bangun Aplikasi Jasa Service Peralatan Elektronik Berbasis Android', 17(1), 122–129.
- Islami, M.J., 2018, 'TANTANGAN DALAM IMPLEMENTASI STRATEGI KEAMANAN SIBER NASIONAL INDONESIA DITINJAU DARI PENILAIAN GLOBAL CYBERSECURITY INDEX', 8(2), 137–137.
- Jasmir, J., Nurmaini, S., Malik, R.F. & Tutuko, B., 2021, 'Bigram feature extraction and conditional random fields model to improve text classification clinical trial document', *Ahmad Dahlan University*, 19(3), 886–886.
- Kurniawan, W.M., Fauziah, F. & Gunaryati, A., 2020, 'Rancang Bangun MyAKAD Apps Berbasis Android Menggunakan Algoritma Sequential Searching', 4(3), 582–582.
- Lee, O.V., Heryanto, A., Razak, M.F.A., Raffei, A.F.M., Phon, D.N.E., Kasim, S. & Sutikno, T., 2020, 'A malicious URLs detection system using optimization and machine learning classifiers', *Institute of Advanced Engineering and Science (IAES)*, 17(3), 1210–1210.

- Liu, R., Wang, Y., Xu, H., Qin, Z., Liu, Y. & Cao, Z., 2023, 'Malicious URL Detection via Pretrained Language Model Guided Multi-Level Feature Attention Network', *ArXiv*, abs/2311.12372.
- Maneriker, P., Stokes, J.W., Lazo, E.G., Carutasu, D., Tajaddodianfar, F. & Gururajan, A., 2021, 'URLTran: Improving Phishing URL Detection Using Transformers', *MILCOM 2021 - 2021 IEEE Military Communications Conference (MILCOM)*, 197–204.
- Novianti, D. & Amin, S., 2021, 'Rancangbangun Sistem Informasi Surat Perintah Perjalanan Dinas pada Lembaga Penjaminan Mutu Pendidikan Papua Barat Berbasis Web', 6(6), 2716–2716.
- P., A., Wahyudi, I., Styawati, M. & Agustin, T., 2023, 'Web-Based Altamart Store Sales System Design', 2(2), 181–194.
- Pandowo, H., Suhasto, R.B.I.N. & Kirowati, E., 2020, 'IMPLEMENTASI SISTEM INFORMASI PERPUSTAKAAN SEKOLAH BERBASIS WEB DENGAN MENGGUNAKAN FRAMEWORK LARAVEL PADA SMK BP SUBULUL HUDA KEMBANGSAWIT KEBONSARI MADIUN', *Politeknik Harapan Bersama Tegal*, 9(1), 26–30.
- Patil, S., Daware, S., Bhagat, A. & Sawarkar, P.J., 2021, 'College ERP Using MERN Stack', 190–196.
- Permana, G.G.S., Mulyani, A. & Tresnawati, D., 2020, 'Rancang Bangun Aplikasi Penyedia Informasi Jasa Terapi Syar'i Berbasis Web', 17(2), 199–203.
- Prasetyadi, G.C., Tri, U. & Mutiara, A.B., 2018, 'Singkat: A Keyword-Based URL Shortener and Click Tracker Package for Django Web Application', *Science and Information Organization*, 9(9).
- Pribadi, N., Sarno, R., Ahmadiyah, A.S. & Sungkono, K.R., 2021, 'Semantic Recommender System Based on Semantic Similarity Using FastText and Word Mover's Distance', *Intelligent Networks and Systems Society*, 14(2), 377–385.

- Putra, M.G.L. & Octantia, H., 2021, 'Analisis dan Perancangan Aplikasi E-Learning Berbasis Gamification (Studi Kasus Program Studi Sistem Informasi Institut Teknologi Kalimantan)', *Brawijaya University*, 8(3), 571–578.
- Rakhmadani, D.P., Adhinata, F.D. & Wardhana, A.C., 2021, 'PERANCANGAN SISTEM INVENTORY RUANG KELAS DENGAN PENDEKATAN METODE QUALITY CONTROL STATISTICAL SAMPLING BERBASIS WEB STUDI KASUS : INSITUT TEKNOLOGI TELKOM PURWOKERTO', 6(1), 68–76.
- Rasyid, H.A.R.H.A., 2022, 'Sistem Informasi Geografis Pemetaan Lokasi Usaha Mikro dan Kecil (UMK) di Kecamatan Percut Sei Tuan', 1(2).
- Reynolds, J.K., Kumar, D., Ma, Z., Subramanian, R., Wu, M., Shelton, M., Mason, J., Stark, E. & Bailey, M., 2020, 'Measuring Identity Confusion with Uniform Resource Locators'.
- Sama, H., Licen, L., Saragi, J.S.D., Erline, M., Kelvin, K., Hartanto, Y., Winata, J. & Devalia, M., 2021, 'STUDI KOMPARASI FRAMEWORK NIST DAN ISO 27001 SEBAGAI STANDAR AUDIT DENGAN METODE DESKRIPTIF STUDI PUSTAKA', 6(2), 116–121.
- Samah Malibari, 2022, *Benign and Malicious URLs*, Kaggle.
- Sergio, G.C. & Lee, M., 2020, 'Stacked DeBERT: All attention in incomplete data for text classification', *Elsevier BV*, 136, 87–96.
- Setiaji, S. & Sastra, R., 2021, 'Implementasi Diagram UML (Unified Modelling Language) Pada Perancangan Sistem Informasi Penggajian', 7(1), 106–111.
- Sukmasetya, P., 2022, 'Education Management Information System: Kebijakan, Perencanaan dan Implementasi (Exploratory Study)', *Muhammadiyah University of Metro*, 1(1), 19–29.

Uda, R., 2020, 'Vulnerable Web Server Protection by Hash Based URL Transformation'.

Wayahdi, M.R. & Ruziq, F., 2023, 'Pemodelan Sistem Penerimaan Anggota Baru dengan Unified Modeling Language (UML) (Studi Kasus: Programmer Association of Battuta)', 12(1), 1514–1521.

Widjaja, V. & Widodo, N.M., 2021, 'Pengaruh Teknologi Internet terhadap Pengetahuan Masyarakat Jakarta Seputar Informasi Vaksinasi Covid-19', 8(1), 1–13.



UIN SUNAN AMPEL
S U R A B A Y A