



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 : Jessica Larasty

NIM : 211402116

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: Ainul Hizriadi, S.Kom., M.Sc.  
(Jika judul dari dosen maka dosen tersebut berhak menjadi pembimbing I)

Calon Dosen Pembimbing II: Niskarto Zendrato, S.Kom., M.Kom

Paraf Calon Dosen Pembimbing I  
  
Ainul Hizriadi

Medan, 07 Februari 2025

Ka. Laboratorium Penelitian,

\* Centang salah satu atau keduanya

\*\* Pilih salah satu

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

NIP. 198908172019032023



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	Identifikasi Pesan Teks Berunsur Phising Menggunakan Kombinasi Model IndoBERT Embedding dan Algoritma LSTM ( <i>Long Short-Term Memory</i> )
Latar Belakang dan Penelitian Terdahulu	<p><b>Latar Belakang</b></p> <p><i>Phishing</i> merupakan metode mencoba mendapatkan informasi seperti nama pengguna, kata sandi, dan informasi kartu kredit dengan berpura-pura menjadi perusahaan komunikasi elektronik yang sah. <i>Phising</i> juga dapat didefinisikan sebagai teknik yang digunakan <i>hacker</i> untuk dapat mengakses sebuah komputer secara tidak sah dimana menimbulkan sebuah ancaman (Felten et al., 1997). Postingan dari situs web terkenal, situs penjualan, pemroses pembayaran online, atau administrator TI sering kali digunakan untuk menarik perhatian yang tidak menaruh curiga. Informasi ini kemudian digunakan oleh penjahat untuk mengakses akun pengguna, menarik dana dari akun tersebut atau mentransfernya ke penjahat, atau melakukan pembelian online menggunakan kartu kredit orang lain. Berbagai cara digunakan untuk memenuhi keinginan penjahat, yang paling umum adalah memikat seseorang dengan hadiah, membuat email dan situs website palsu yang terlihat seperti email asli dan situs web perbankan. (Purnamasari et al., 2023)</p> <p>Faktor penyebab sehingga munculnya ancaman serangan <i>phishing</i> adalah minimnya pengetahuan pengguna, psikologis dan <i>privasi social networking services</i> pengguna. Menurut Dhamija, Tygar, &amp; Hearst (2006) mengungkapkan bahwa pengguna dianggap tidak memiliki pengetahuan yang baik mengenai sistem komputer terutama membedakan domain yang resmi dan palsu. Selain itu, banyak korban yang terjebak dikarenakan bahasa yang digunakan menyerupai komunikasi resmi.</p> <p>Menurut Ketua Pengelola Nama Domain Internet Indonesia (Pandi), Yudho Giri Sucahyo, jumlah serangan <i>phishing</i> unik yang dilaporkan pada Q3 (kuartal 3) 2022 sebanyak 7.988. Sektor bisnis adalah sektor yang paling banyak menjadi sasaran. Serangan <i>phishing</i> pada Q3 2022 adalah lembaga pemerintahan, sedangkan jumlah domain unik yang digunakan untuk serangan <i>phishing</i> pada Q3 2022 sebanyak 181. Hal ini menunjukkan bahwa ancaman <i>phising</i> sangat membahayakan, dan perlu dideteksi dengan metode dan algoritma yang tepat untuk meminimalisir terjadinya penipuan.</p> <p>Kemudian, masalah utama dalam pendeteksian pesan teks berunsur <i>phising</i> ini adalah terbatasnya metode deteksi tradisional terdahulu yang hanya menggunakan metode berbasis aturan (<i>rule-based detection</i>) dan pencocokan kata (<i>keyword matching</i>), serta masih bersifat statis dan masih berfokus pada bahasa Inggris. Akibatnya, sulit menangani pola <i>phising</i> variasi baru dan yang memahami bahasa Indonesia.</p>



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>

Untuk itu, teknologi *Machine Learning* dan *Deep Learning* dapat membantu dalam mendeteksi unsur phishing dalam pesan teks tersebut. Penelitian yang dilakukan oleh Kelvin Nathanael Lumbanraja (2024) menggunakan *Support Vector Machine (SVM)* dan *Ensembled Bagging* sebagai metode deteksi phishing dalam pesan teks dengan judul “Identifikasi Phishing pada Pesan Teks Menggunakan Algoritma Support Vector Machine dengan Ensembled Bagging”. Penelitian ini mengumpulkan 1.600 data phishing dan non-phishing dari penelitian sebelumnya serta pesan pribadi peneliti. Proses pre-processing meliputi pembersihan data, tokenisasi, eliminasi stopwords, dan stemming. Model SVM digunakan sebagai dasar klasifikasi, kemudian dioptimalkan dengan Ensembled Bagging untuk meningkatkan akurasi. Hasil yang diperoleh menunjukkan akurasi 95,2%. Akan tetapi, SVM hanya mengandalkan fitur numerik dari teks yang mengakibatkan kehilangan makna sebenarnya dari kata-kata phishing.

Penelitian lainnya oleh Gupta et al., (2024) menggunakan BERT untuk ekstraksi fitur dan CNN untuk klasifikasi dalam deteksi phishing email, dengan judul “*Advanced BERT and CNN-Based Computational Model for Phishing Detection in Enterprise Systems*”. Model ini menghasilkan akurasi mencapai 97,5%.

Kemudian, penelitian oleh Tawil et al., (2024) menggunakan TF-IDF, Word2Vec, dan BERT untuk mengevaluasi efektivitas berbagai algoritma dalam mendeteksi phishing, dengan judul “*Comparative Analysis of Machine Learning Algorithms for Email Phishing Detection Using TF-IDF, Word2Vec, and BERT*”. Model terbaik adalah BERT, yang mencapai akurasi 99%, mengungguli Multilayer Perceptron dengan TF-IDF dan Word2Vec (98%). Studi ini menunjukkan bahwa model pre-trained seperti BERT dapat secara signifikan meningkatkan akurasi sistem deteksi phishing.

Penelitian lain yang menggunakan BERT dalam mendeteksi phishing dilakukan oleh Elsadig et al., (2022) yang berjudul “*Intelligent Deep Machine Learning Cyber Phishing URL Detection Based on BERT Features Extraction*”. Penelitian ini menggunakan data yang dikumpulkan dari Phishing Site Predict. Selanjutnya, dilakukan preprocessing data menggunakan teknik Natural Language Processing (NLP) seperti tokenization, stopwords removal, stemming, dan lemmatization agar data dapat digunakan untuk training pada model BERT. Akurasi yang diperoleh mencapai 96,66%, menunjukkan efektivitas tinggi dalam deteksi phishing.

SMS phishing (smishing) juga diteliti oleh Gunikhan Sonowal (2020) dengan judul “*Detecting Phishing SMS Based on Multiple Correlation Algorithms*” dimana penelitian ini menggunakan empat algoritma ranking korelasi yaitu Pearson, Spearman, Kendall, dan Point Biserial untuk menganalisis fitur terbaik dalam mendeteksi pesan smishing, dengan menggunakan algoritma AdaBoost. Hasil eksperimen menunjukkan bahwa Kendall rank correlation memberikan akurasi terbaik sebesar 98,40%, dan berhasil mengurangi dimensi fitur sebanyak 61,53%.

Penelitian lainnya tentang deteksi phishing dilakukan oleh Ariyadasa et al., (2020) dengan judul “*Detecting phishing attacks using a combined model of LSTM and CNN*”. Penelitian ini menggunakan model gabungan LSTM dan CNN untuk menganalisis fitur dari URL dan halaman HTML dalam mendeteksi serangan phishing. Data URL dipelajari menggunakan LSTM dengan 1D convolutional, sementara fitur HTML dipelajari menggunakan jaringan 1D



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

convolutional. Kedua jaringan ini dilatih secara terpisah dan digabungkan melalui sigmoid layer untuk membentuk model yang diusulkan. Hasil eksperimen menunjukkan bahwa model yang diusulkan mencapai akurasi sebesar 98,34%, melebihi rekor sebelumnya sebesar 97,3%. Model ini hanya memerlukan ekstraksi fitur HTML dan pemrosesan minimal pada URL, sehingga dapat digunakan untuk aplikasi deteksi phishing secara real-time tanpa bergantung pada layanan pihak ketiga.

Data Mining dengan algoritma C4.5 juga pernah diteliti oleh Salim et al., (2017), dengan judul “Data Mining Identifikasi Website Phising Menggunakan Algoritma C4.5”. Penelitian ini menggunakan metode data mining Decision Tree untuk menemukan pola yang mengandung informasi dari sejumlah besar data sampel website. Dataset yang digunakan berasal dari UCI Dataset, yang menyediakan berbagai kumpulan data untuk keperluan penelitian dan akademik. Hasil analisis menunjukkan bahwa beberapa faktor dapat dijadikan referensi atau tanda phishing website.

Berdasarkan penelitian yang telah dilakukan, penulis mengusulkan penerapan kombinasi model IndoBERT Embedding dan algoritma *Long Short-Term Memory* untuk mengidentifikasi apakah pesan teks tersebut memiliki unsur phising atau tidak. Penulis memberikan penelitian ini judul “Identifikasi Pesan Teks Berunsur Phising Menggunakan Kombinasi Model IndoBERT Embedding dan Algoritma LSTM (*Long Short-Term Memory*)”.

**Penelitian Terdahulu**

No.	Penulis	Judul	Tahun
1.	Kelvin Nathanael Lumbanraja	Identifikasi Phising pada Pesan Teks Menggunakan Algoritma Support Vector Machine dengan Ensembled Bagging	2024
2.	Brij B. Gupta, Akshat Gaurav, Varsha Arya, RazazWaheeb Attar, Shavi Bansal, Ahmed Alhomoud and Kwok Tai Chui	Advanced BERT and CNN-Based Computational Model for Phishing Detection in Enterprise Systems	2024
3.	Arar Al Tawil, Laiali Almazaydeh, Doaa Qawasmeh, Baraah Qawasmeh, Mohammad Alshinwan, and Khaled Elleithy	Comparative Analysis of Machine Learning Algorithms for Email Phishing Detection Using TF-IDF, Word2Vec, and BERT	2024



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>

	4.	Yunita Renta Hutagaol, Yulyani Arifin	Semantic-Based Email Spam Classification Using Bert Method	2024
	5.	Muna Elsadig, Ashraf Osman Ibrahim, Shakila Basheer, Manal Abdullah Alohal, Sara Alshunaifi, Haya Alqahtani, Nihal Alharbi and Wamda Nagmeldin	Intelligent Deep Machine Learning Cyber Phishing URL Detection Based on BERT Features Extraction	2022
	6.	Gunikhan Sonowal	Detecting Phishing SMS Based on Multiple Correlation Algorithms	2020
	7.	Subhash N. Ariyadasa, Shantha Fernando, Subha Fernando	Detecting phishing attacks using a combined model of LSTM and CNN	2020
	8.	Tomy Salim, Yo Ceng Giap	Data Mining Identifikasi Website Phising Menggunakan Algoritma C4.5	2017
<b>Rumusan Masalah</b>	<p>Phishing merupakan salah satu ancaman siber yang berbahaya karena dapat menipu pengguna untuk memberikan informasi pribadi. Serangan phishing yang semakin canggih sering kali sulit dikenali oleh pengguna, terutama karena teknik manipulasi sosial yang digunakan menyerupai komunikasi resmi. Oleh karena itu, diperlukan sistem deteksi yang efektif untuk mengidentifikasi pesan phishing secara akurat dan mencegah dampak negatifnya. Penelitian ini berfokus pada pengembangan sistem deteksi phishing dalam bahasa Indonesia dengan menggabungkan model IndoBERT dan algoritma LSTM dan membangun model deteksi phishing yang efektif serta lebih unggul dibandingkan metode deteksi tradisional. Selain itu, penelitian ini juga meneliti bagaimana kombinasi algoritma ini dapat menangani berbagai variasi dan pola pesan phishing yang semakin kompleks. Optimalisasi model juga menjadi perhatian utama agar akurasi deteksi phishing dapat ditingkatkan dan mampu memberikan perlindungan lebih baik bagi pengguna dari ancaman phishing yang terus berkembang.</p>			





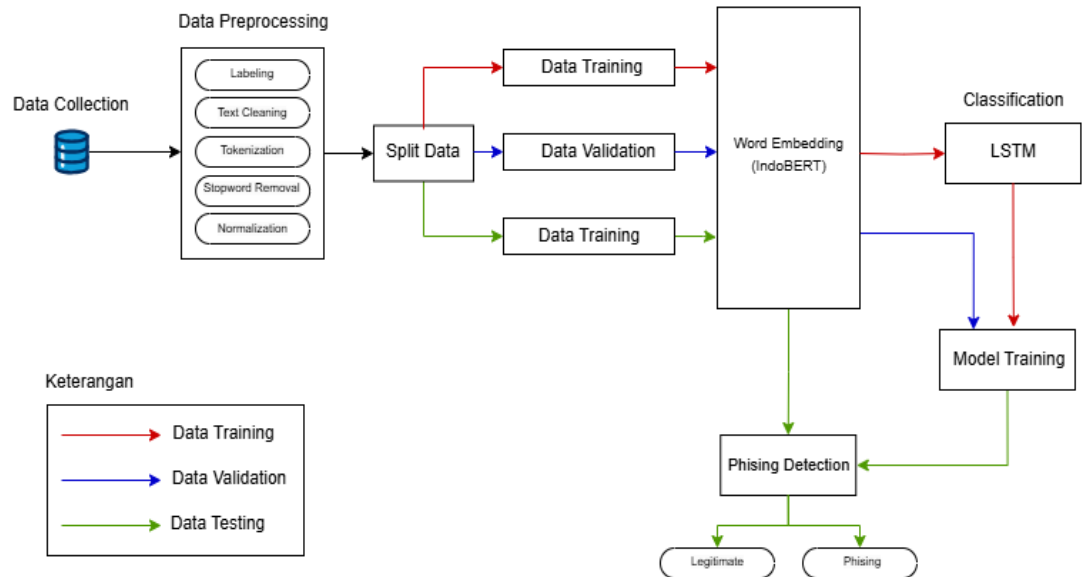
# KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI

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

### Metodologi



### Penjelasan Tahapan Penelitian

#### 1. Data Collection

Tahapan ini adalah tahapan awal, yaitu mengumpulkan data pesan *phishing* dan pesan sah dari berbagai sumber dataset (misalnya, SMS, email, atau data online). Dataset juga harus seimbang untuk menghindari bias dalam klasifikasi.

#### 2. Data Pre-processing

Pada tahap ini, dilakukan serangkaian proses yang bertujuan untuk menghasilkan data yang baik sehingga data dapat lebih mudah dimengerti oleh model. Tahapan ini dibagi menjadi beberapa bagian atau proses, yaitu:

- *Labeling*

Tahapan ini melakukan proses pemberian label berupa 0 dan 1 untuk setiap data testing. Dimana, kelas 0 untuk menyatakan pesan teks non-phishing dan kelas 1 untuk pesan teks *phishing*.

- *Text Cleaning*

Menghapus simbol, tautan, dan tag HTML yang tidak diperlukan.

- *Tokenization*

Memecah teks menjadi token kata.

- *Stopword Removal*

Menghapus kata-kata umum yang tidak memiliki makna signifikan.

- *Normalization*

Mengubah kata tidak baku atau slang menjadi bentuk standar.

#### 3. Data Splitting

Setelah melalui tahapan preprocessing, maka dataset akan dibagi menjadi:

- *Training Set (60%)*

Digunakan untuk melatih model.



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>

	<ul style="list-style-type: none"><li>• <i>Validation Set (15%)</i> Digunakan untuk menyesuaikan parameter model dan mencegah <i>overfitting</i>.</li><li>• <i>Testing Set (25%)</i> Digunakan untuk menguji kinerja model setelah pelatihan.</li></ul> <p><b>4. Word Embedding</b> Pada tahap ini, teks yang telah diproses akan dikonversi menjadi representasi vektor menggunakan model <i>IndoBERT</i>. Kemudian proses ini akan menghasilkan <i>embedding</i> kata untuk menangkap makna kata berdasarkan konteks kata sekitarnya.</p> <p><b>5. Modelling</b> Tahapan ini akan melakukan pelatihan model agar dapat mengidentifikasi data train yang telah diubah menjadi vektor menggunakan algoritma LSTM. Algoritma ini bekerja secara dua arah, dimana lapisan pertama (<i>forward layer</i>) akan berfokus untuk urutan setiap kata yang ada pada teks dan lapisan kedua (<i>backward layer</i>) akan merepresentasikan konteks dari kata-kata tersebut. Hasil pelatihan model (<i>learned model</i>) akan digunakan untuk mengidentifikasi pesan teks.</p> <p><b>6. Output</b> <i>Learned model</i> yang dihasilkan akan digunakan untuk mendeteksi pesan teks, apakah <i>legitimate</i> atau <i>phishing</i></p>
Referensi	<p>Gupta, B. B., Gaurav, A., Arya, V., Attar, R. W., Bansal, S., Alhomoud, A., &amp; Chui, K. T. (2024). Advanced BERT and CNN-based computational model for phishing detection in enterprise systems. <i>Computer Modeling in Engineering &amp; Sciences</i>, 141(3), <a href="https://doi.org/10.32604/cmes.2024.056473">https://doi.org/10.32604/cmes.2024.056473</a></p> <p>Al Tawil, A., Almazaydeh, L., Qawasmeh, D., Qawasmeh, B., Alshinwan, M., &amp; Elleithy, K. (2024). Comparative analysis of machine learning algorithms for email phishing detection using TF-IDF, Word2Vec, and BERT. <i>CMC</i>, 2024, <a href="https://doi.org/10.32604/cmc.2024.057279">https://doi.org/10.32604/cmc.2024.057279</a></p> <p>Hutagaol, Y. R., &amp; Arifin, Y. (2024). Semantic-Based Email Spam Classification Using BERT Method. <i>Journal of Information Technology and Computer Science (INTECOMS)</i>, 7(5), 1823–1836</p> <p>Elsadig, M., Ibrahim, A. O., Basheer, S., Alohal, M. A., Alshunaifi, S., Alqahtani, H., Alharbi, N., &amp; Nagmeldin, W. (2022). Intelligent Deep Machine Learning Cyber Phishing URL Detection Based on BERT Features Extraction. <i>Electronics</i>, 11(3647). <a href="https://doi.org/10.3390/electronics11223647">https://doi.org/10.3390/electronics11223647</a></p> <p>Sonowal, G. (2020). Detecting Phishing SMS Based on Multiple Correlation Algorithms. <i>SN Computer Science</i>, 1(361). <a href="https://doi.org/10.1007/s42979-020-00377-8">https://doi.org/10.1007/s42979-020-00377-8</a></p> <p>Ariyadasa, S., Fernando, S., &amp; Fernando, S. (2020). Detecting phishing attacks using a combined model of LSTM and CNN. <i>International Journal of Advanced and Applied Sciences</i>, 7(7), 56-67. <a href="https://doi.org/10.21833/ijaas.2020.07.007">https://doi.org/10.21833/ijaas.2020.07.007</a></p>



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>

- |  |   |
|--|---|
|  | <p>Salim, T., &amp; Giap, Y. C. (2017). Data mining identifikasi website phishing menggunakan algoritma C4.5. <i>Jurnal TAM (Technology Acceptance Model)</i>, 8, 130-135.</p> <p>Putra, V. F. (2021). Modus operandi tindak pidana phishing menurut UU ITE. <i>Jurist-Diction</i>, 4(6), 2525-2548.<br/><a href="https://doi.org/10.20473/jd.v4i6.31857">https://doi.org/10.20473/jd.v4i6.31857</a></p> <p>Purnamasari, S., &amp; Sutabri, T. (2023). Analisis kejahatan online phishing pada institusi pemerintah/pendidikan sehari-hari. <i>Jurnal Digital Teknologi Informasi</i>, 6(1), 29-34.<br/><a href="https://doi.org/10.1234/jdti.v6i1.1234">https://doi.org/10.1234/jdti.v6i1.1234</a></p> <p>Ginting, E., Sinaga, M. P., Nurdin, M. R., &amp; Putra, M. D. (2023). Analisis ancaman phishing terhadap layanan online perbankan (studi kasus pada Bank BRI). <i>UNES Journal of Scienteck Research</i>, 8(1), 041-047.<br/><a href="https://ojs.ekasakti.org/index.php/UJSR/">https://ojs.ekasakti.org/index.php/UJSR/</a></p> |
|--|---|

Medan, 07 Februari 2025  
Mahasiswa yang mengajukan,

(Jessica Larasty)

NIM. 211402116