HANIF
Journal of Information Systems
https://journal.ilmubersama.com/index.php/hanif

Cybersecurity

Detecting Zero-Width Characters Obfuscated in Phishing URLs using the XGBOOST Algorithm

Ahmad Asadel [1], *Andi Zulherry* [2*]

[1] *Department of Information Technology, Faculty of Computer Science and Information Technology, Universitas Muhammadiyah Sumatera Utara, Medan, 20238, North Sumatera, Indonesia*
[2] *Department of Sains Data, Faculty of Computer Science and Information Technology, Universitas Muhammadiyah Sumatera Utara, Medan, 20238, North Sumatera, Indonesia*

ARTICLE INFORMATION

Received: Aug 28, 2025
Revised: Oct 16, 2025
Available Online: Jan 31, 2026

KEYWORDS

Phishing
Zero-Width Characters
XGBoost
URL Detection
Cybersecurity
Machine Learning

CORRESPONDENCE (*)

Phone: +62 822-7314-7929
E-mail: andizulherry@umsu.ac.id

A B S T R A C T

Phishing attacks represent one of the most common and damaging cyber threats, with techniques continuously evolving to become more sophisticated and harder to detect. One of the latest evasion methods of concern is the use of Zero-Width Characters (ZWC)—invisible Unicode Characters inserted into URLs to deceive traditional detection systems and human visual perception. This research aims to develop and evaluate an effective and reliable machine learning model to detect phishing URLs that have been obfuscated using ZWC. The eXtreme Gradient Boosting (XGBoost) algorithm was chosen for its proven superiority in handling complex data and its performance optimization capabilities. This study utilized a public dataset from Kaggle consisting of 11,430 URL samples, which was then modified through a feature engineering process. Specifically, 50% of the phishing URLs were injected with one of five types of ZWC (ZWSP, ZWNJ, ZWJ, RLM, LRM), and a dedicated binary feature was created to flag the presence of these Characters. Initial training revealed signs of minor overfitting. Consequently, a hyperparameter tuning process was conducted by adjusting the max_depth and min_child_weight parameters to create a more robust model. The final model was evaluated on 20% of the test data and demonstrated exceptionally high performance, achieving an Accuracy of 97.24%, Precision of 97.03%, Recall of 97.37%, and an AUC score of 0.9972. The high Recall value is particularly crucial, proving the model's reliability in minimizing the risk of missed threats. This research successfully proves that an XGBoost-based approach with targeted feature engineering can be an effective solution against advanced phishing attacks.

INTRODUCTION

The internet has become an integral part of everyday life. It has penetrated various aspects of our lives, from communication and business to education [1,2,3]. This advancement has made access to information increasingly easy and fast. This development has also brought positive impacts, significantly increasing connectivity and efficiency globally. However, on the other hand, it has also created new threats, namely the increased risk of cybercrime, one of which is phishing.

Phishing attacks are one of the most damaging and common forms of cyberattacks [4,5]. According to the latest report from the APWG (Anti-Phishing Working Group) in 2024, phishing attacks have increased significantly annually, with financial losses continuing to rise. Since June 2023, the number of phishing attacks has been reported to be in the range of 290,000 to 370,000 attacks per month. Citing a report from Intersile Consulting, an annual increase of 50,000 attacks was recorded, bringing the total to nearly 1.9 million attacks between May 2023 and April 2024. Phishing attacks are

https://doi.org/10.56211/hanif.v3i1.56

typically carried out through psychological manipulation (social engineering) with the aim of stealing sensitive data such as passwords, credit card numbers, or users' personal information.

Currently, the phishing methods used by attackers are increasingly diverse and sophisticated. One new technique that has begun to emerge and garner attention from the security community is the use of Zero-Width Characters. Zero-Width Characters (ZWC) are Unicode characters that have no width or visual space when displayed on a user's device screen [6,7] These characters include Zero-Width Space (U+200B), Zero-Width Non-Joiner (U+200C), Zero-Width Joiner (U+200D), Left-To-Right Mark (U+200E), and Right-To-Left Mark (U+200F).

This phishing technique, implemented by exploiting Unicode, is known as a Zero-Width Attack on Security & Privacy (Z-WASP) (Kaavija, 2025). This technique works by inserting ZWC characters into sentences that appear normal to humans but will affect how machines analyze and index the text. These characters are used to evade detection by systems that inspect text, such as in the context of attacks on hate speech detection systems or translation applications [8,9]. For example, an attacker might display a link that looks like www.umsumedan.com. However, the perpetrator has actually inserted ZWC characters between several letters, for example between the letters d and a, to form www.umsumedan.com (with the symbol U+200B—Zero-Width Space). However, technically, the browser will read the modified URL and then redirect the user to a fake site.

In addition to visually deceiving users, this character is also capable of evading automated text- or URL-based security systems because it is difficult to detect using traditional text-matching checks. For example, a phishing attack using Zero-Width successfully evaded Microsoft Office 365's email security system in 2019, as well as various corporate email systems in 2024–2025 [10,11].

The problem that arises is that currently common security systems, such as signature-based antivirus and rule-based email filtering systems, still struggle to detect this type of phishing attack. Therefore, a new, smarter and more adaptive approach is needed to effectively detect Zero-Width Character-based phishing attacks.

Machine learning algorithms have proven effective in detecting various cyberattacks. This approach allows systems to learn patterns from data and make predictions or classifications without the need for explicit programming for each scenario. One machine learning algorithm that demonstrates superior performance in classification tasks is Extreme Gradient Boosting (XGBoost). XGBoost is a machine learning algorithm known for its ability to handle large and complex data sets.

In the context of phishing detection, XGBoost works by sequentially building a series of decision trees. Each new tree added focuses on correcting prediction errors made by previous trees. This means that if the first tree incorrectly classifies a URL as non-phishing when it actually contains ZWC, subsequent trees will learn from that error and attempt to classify it correctly. This iterative process allows XGBoost to gradually improve the model's overall accuracy in identifying subtle and complex patterns in URLs, including the presence of obfuscated Zero-Width Characters.

The goal of this research is to create an effective model for detecting phishing attacks using the XGBoost algorithm, which is renowned for its ability to handle large and complex data. This algorithm was chosen because of its outstanding ability to identify hidden patterns in data, which is crucial for detecting Zero-Width Characters disguised in Phishing URLs.

Furthermore, this study will evaluate the performance of the XGBoost model in detecting phishing URLs using various evaluation metrics, such as accuracy, precision, recall, and F1 score. The results of this evaluation will demonstrate the model's effectiveness in real-world situations and identify areas requiring further improvement. Based on the evaluation results, this study will provide practical recommendations for developing a better and more efficient phishing detection system.

METHOD

Workflow

This research aims to develop a phishing detection model using the XGBoost algorithm, which is designed to detect phishing URLs that have been disguised with Zero-Width characters. In this research process, stages are required so that the model built is a predictive model and can perform prediction processes accurately.

XGBoost

XGBoost, short for eXtreme Gradient Boosting, is a highly reliable machine learning technique in classification and regression tasks due to its ability to handle large and complex datasets (Mim Hanifah Permana, 2024). The XGBoost algorithm was chosen in this study because it has advantages in terms of speed and memory usage. Better utilization of processor cache, multicore processing and distributed parallel computing make the system run faster than popular algorithms commonly used. XGBoost optimizes predictive models through an ensemble of decision trees that correct each other's errors. Each new tree added focuses on the errors left by the previous tree, gradually improving the overall accuracy of the model (Ryan Afrizal et al., 2018).

AUC (Area Under the Curve)

The AUC is a score that measures the area under the ROC curve. Its value ranges from 0.0 to 1.0. 0.5 means the model is no better than random guessing, while 1.0 means the model is a simple separator. The AUC is particularly useful for evaluating models on imbalanced datasets and provides a more comprehensive picture of the model's discriminatory ability.

Confusion Matrix

A confusion matrix is a method used to measure the performance of a classification method [14,15]. This matrix presents a table summarizing the number of correct and incorrect predictions produced by a classifier or classification model for a binary classification task [12,13]. This matrix is very useful in evaluating the performance of machine learning models because it shows not only the errors made by the model but also the types of those errors.

RESULTS AND DISCUSSION

Data Collection and Preparation

This research uses a public dataset titled "Web Page Phishing Detection" obtained from the Kaggle platform. This raw dataset consists of 11,430 URL samples with 88 extracted features. Each sample is classified into two categories: legitimate for legitimate URLs and phishing for malicious URLs.

Next, a feature engineering process was carried out by modifying the original dataset to build detection capabilities for Zero-Width Characters (ZWC). The first step was to add a new feature column named “mengandung_zwc”, which was initially filled with a value of 0 for all samples. Then, 50% of the total phishing URLs were randomly selected to be inserted with one of the five types of ZWC characters that had been defined in the problem constraints, namely Zero-Width Space (U+200B), Zero-Width Non-Joiner (U+200C), Zero-Width Joiner (U+200D), Right-to-Left Mark (U+200F), and Left-to-Right Mark (U+200E). For each modified URL, the value in the mengandung_zwc column was changed to 1.

To ensure that the modification process is only applied to the correct class, data verification is performed as shown in Table 4.1. Based on the table, it can be seen that the value contains_zwc = 1 is only found in URLs with phishing status and is zero in legitimate URLs, so that the validity of the feature engineering process can be tested.

Table 1. Verifying Data Modification Results Steps

Zero-Width Character Status	Legitimate	Phishing
0	5715	2858
1	-	2857

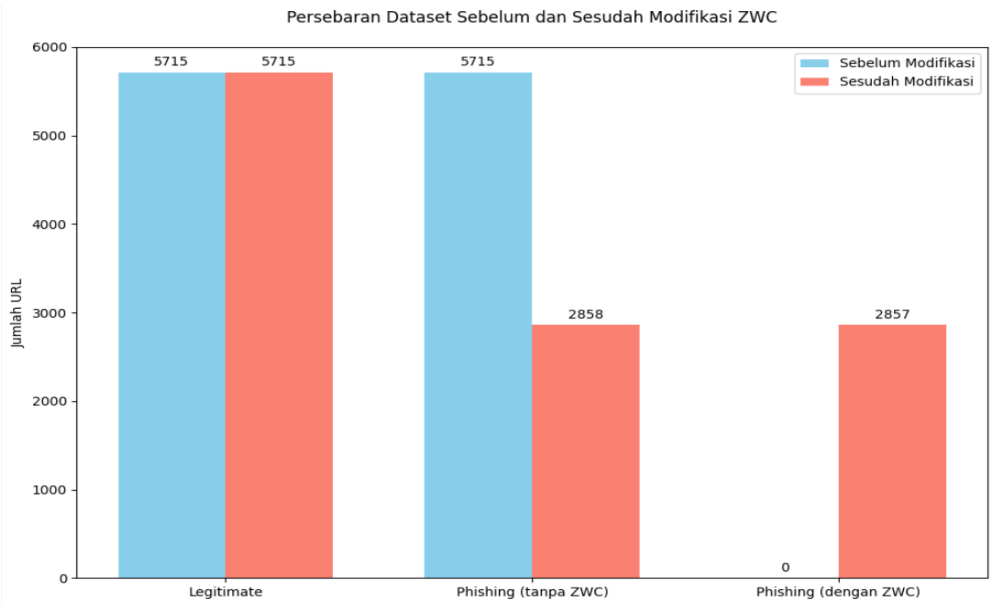

Figure 1. Dataset Distribution Before and After ZWC Modification

The final step in the preparation phase was to divide the modified final dataset into two main parts: 80% as training data and 20% as testing data. The figure below visually illustrates this division: of the total 11,430 URLs, 9,144 were allocated as training data, and the remaining 2,286 URLs were used as testing data. This division process used a stratified split method to ensure that the proportion between legitimate and phishing classes remained balanced across both datasets, thus preventing bias during the model training and testing process.

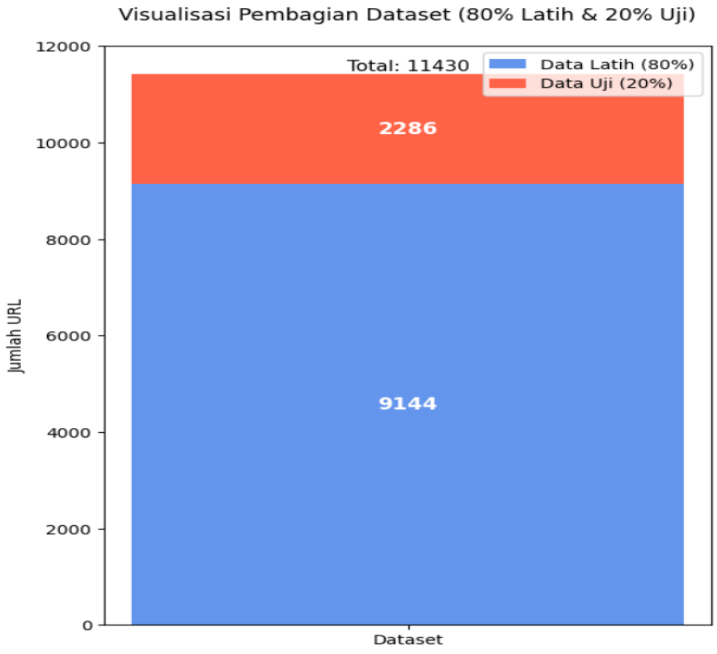

Figure 2. Dataset Distribution Visualization

Model Training and Tuning

Once the data is prepared, the next step is to train a classification model using the XGBoost algorithm. This process aims to build a model capable of recognizing patterns between URL features and their status, whether they are legitimate or phishing.

In the initial training phase, the model was trained using training data with a standard parameter configuration. However, initial evaluation results indicated mild overfitting. Overfitting occurs when a model performs too well on data it already knows (training data), but its performance declines slightly when faced with new data (test data). This is evident from the accuracy comparison, where the model achieved 100% accuracy on the training data, while the result was 98.03% on the

 https://doi.org/10.56211/hanif.v3i1.56

test data. This difference, or "gap," indicates that the model tends to "memorize" the training data rather than learn to generalize its patterns.

To address this issue, a hyperparameter tuning process was performed to reduce model complexity. Two key parameters were reset: max_depth was capped at 4 and min_child_weight was set to 3. These two key parameters were reset to limit the model's ability to 'memorize' data.

After the tuning process, the model was retrained and re-evaluated. The results, as summarized in Figure 4, showed significant improvements. The "gap" between training and test accuracy was successfully reduced to less than 1%, indicating that the final model was no longer overfitting and had better generalization capabilities. This tuned model was then used as the final model in this study.

Model Results and Evaluation

After training and tuning, the model was tested using 20% of the previously unseen test data. This evaluation phase aims to objectively measure the model's performance in classifying URLs. Performance measurements are based on four main metrics, namely Accuracy, Precision, Recall, and F1-Score.

The results of the model performance evaluation on the test data are summarized in full in Table 2. The model achieved an accuracy of 97.24%, indicating that it was able to correctly classify the majority of URLs. Furthermore, the Precision value of 97.03% indicates a low error rate in predicting safe URLs as phishing (false positives). Equally important, the high Recall value of 97.37% demonstrates the model's excellent ability to detect most of the actual phishing URLs, thus minimizing the risk of missed threats (false negatives).

Table 2. Model Evaluation Metric Results

Metrix	Value
Accuracy	97.24%
Precision	97.03%
Recall	97.37%
F1-Score	97.17%

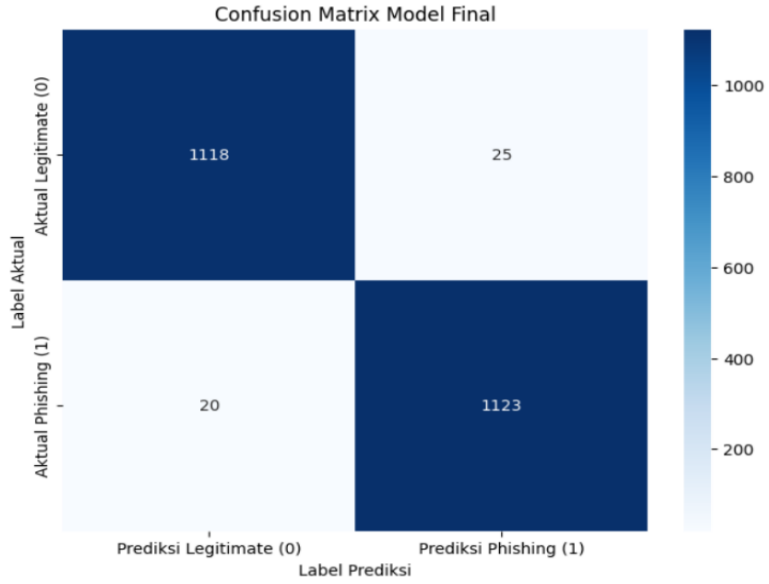

Figure 3. Confusion Matrix Visualization

Next, to measure the model's ability to distinguish between legitimate and phishing classes at various probability thresholds, a Receiver Operating Characteristic (ROC) curve was used. The area under this curve (AUC) provides a single score representing the model's discriminatory ability. As shown in Figure 4.6, the final model achieved an AUC score of

0.9972, which is close to the perfect value of 1.0. This indicates that the model has excellent capabilities in separating the two classes.

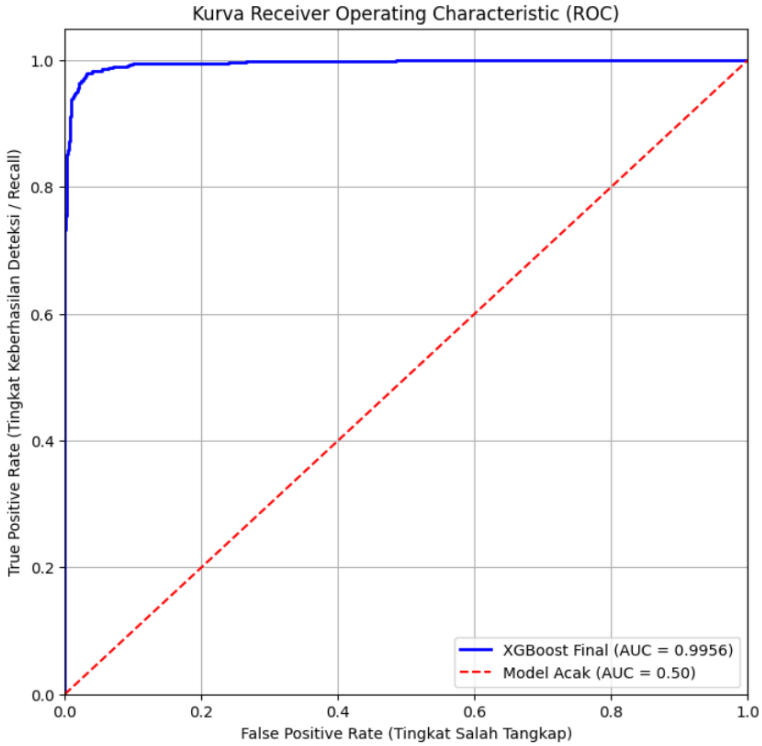

Figure 4. ROC-AUC Model Curve

Based on the evaluation results presented in the previous sub-chapter, a more in-depth discussion can be conducted to interpret the model's performance and its implications. Overall, the XGBoost model that has been optimized to show very high performance and is effective in detecting phishing URLs, including those that have been obfuscated using Zero-Width Characters (ZWC).

The 97.24% accuracy rate demonstrates the model's highly reliable ability to correctly classify URLs in the vast majority of cases. However, in the context of cybersecurity, the 97.37% Recall metric holds even greater significance. This high Recall value indicates that the model successfully identified nearly all phishing threats in the test data. This is crucial because the primary goal of a detection system is to minimize the number of undetected threats (False Negatives), which pose the greatest risk to users.

Further analysis of the Confusion Matrix (Figure 4.5) provides insight into the types of errors the model made. Only 20 cases of False Negatives were recorded, meaning only 20 of the 1,143 phishing URLs were incorrectly considered safe. Conversely, there were 25 cases of False Positives, where legitimate URLs were mistakenly identified as phishing.

An interesting example of a False Positive case was found during testing on several legitimate URLs belonging to Gunadarma University (e.g., https://praktikum.gunadarma.ac.id/login). The model incorrectly classified it as phishing, most likely because the URL had several technical characteristics that coincidentally resembled the pattern of malicious URLs, such as the use of a specific subdomain (praktikum) and the presence of sensitive keywords (login) within the URL structure. Cases like this demonstrate that, despite being highly accurate, technical feature-based models can sometimes make mistakes without the contextual understanding that humans possess.

During testing, we also discovered instances where the model showed hesitation (~50% confidence score) when confronted with a very "sophisticated" or well-crafted conventional phishing URL. This hesitation occurred because the URL had few suspicious technical characteristics, making the model's total "danger score" insufficient to make a highly confident decision.

 https://doi.org/10.56211/hanif.v3i1.56

This phenomenon further highlights the importance of the contains_zwc feature, which is the core of this research. If a subtle phishing attack is coupled with the ZWC trick, the contains_zwc feature will serve as a very strong danger signal. This feature will act as a "tie-breaker," turning the model's doubt into a very confident "Phishing" decision. Thus, it can be concluded that the ZWC detection feature not only addresses specific threats but also serves as a crucial layer of defense to strengthen the model's ability to deal with increasingly sophisticated and ambiguous cyberattacks.

User Interface (UI) Implementation

To demonstrate the model's capabilities in a practical and interactive manner, a simple user interface (UI) was developed. The goal of this implementation was to provide concrete evidence that the trained model was not only statistically valid but also applicable to real-time URL analysis.

This interface was built using the Gradio library within the Google Colaboratory environment. Gradio was chosen for its ability to quickly create and share interactive web applications directly from Python code.

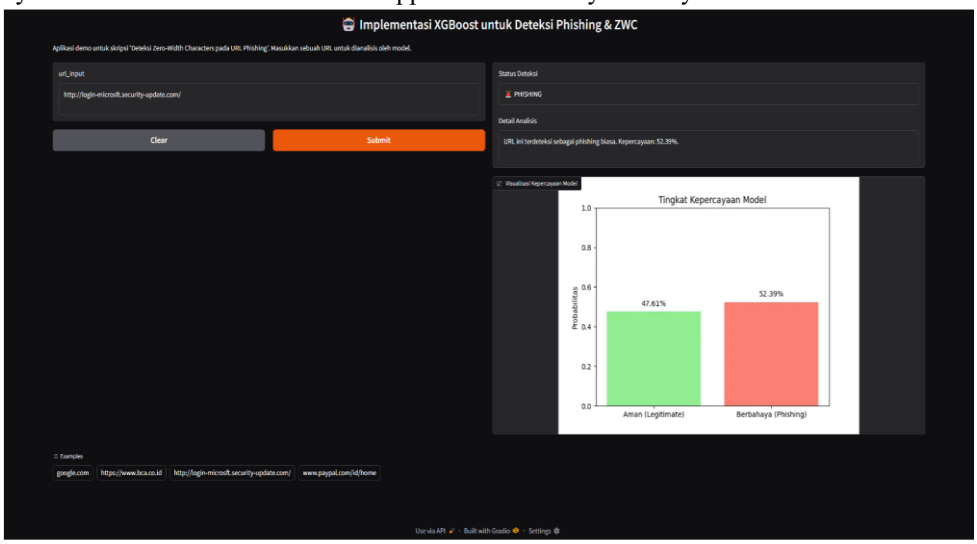

Fig 5. Phishing Detection Application UI View

The mechanics behind this interface begin when the user presses the detect button. First, the system validates the input to ensure that the entered text is in a valid URL format. If the input is valid, a function extracts features from the URL, including checking for the presence of Zero-Width Characters (ZWC). The resulting feature vector is then fed into the pre-trained and preloaded model (model_final.json) to generate predictions.

CONCLUSION

Based on the research results and discussion on "Detecting Obscured Zero-Width Characters in Phishing URLs Using the XGBoost Algorithm," the following conclusions can be drawn: A phishing detection model using the XGBoost algorithm has been successfully developed and implemented. This model is capable of identifying phishing URLs disguised with Zero-Width Characters (ZWC) through a feature engineering process, in which a special feature (contains_zwc) is created to recognize the presence of five types of ZWC (ZWSP, ZWNJ, ZWJ, RLM, and LRM). The model distinguishes between phishing and legitimate URLs by analyzing a combination of 88 technical and structural URL features—such as URL length, number of dots, and domain reputation—and the contains_zwc feature, which specifically targets ZWC attacks. The presence of the ZWC feature proved to be a very strong indicator for detecting more sophisticated attacks. The XGBoost model, optimized through hyperparameter tuning, proved highly effective in detecting phishing URLs. This is evidenced by the achievement of very high evaluation metrics on the test data, namely Accuracy of 97.24%, Precision of 97.03%, Recall of 97.37%, and an AUC score of 0.9972. The high Recall value indicates the model's highly reliable ability to minimize the risk of malicious threats escaping.

REFERENCES

Book

[1] Rolly Junius Lontaan Muhammad Fairuzabadi, Indah Purnama Sari Imam Ekowicaksono, Fatimah Nur Arifah Rahman Indra Kesuma, Nizirwan Anwar Andika Setiawan. Deep Learning untuk Pemula: Memahami Algoritma, Tools, dan Masa Depan AI. Medan: Yayasan Kita Menulis 1,2025,pp. 150

[2] Binastya Anggara Sekti, Indah Purnama Sari, Fanny Ramadhani, Andy Satria, Pasnur Pasnur, Sitti Harlina, Purwa Hasan Putra, Sitti Aisa, Nizirwan Anwar, Samuel Hasudungan Tampubolon, Siti Sundari, Muhammad Noor Hasan Siregar, Wilsen Grivin Mokodaser, Janner Simarmata, Annahl Riadi, Baso Ali, Semmy Wellem Taju, Nirsal Nirsal. Pengantar Kecerdasan Buatan untuk Pemula. Medan: Yayasan Kita Menulis, pp.350.

[3] Indah Purnama Sari. Algoritma dan Pemrograman. Medan: UMSU Press, 2023, pp. 290.

[4] Indah Purnama Sari. Buku Ajar Pemrograman Internet Dasar. Medan: UMSU Press, 2022, pp. 300.

[5] Indah Purnama Sari. Buku Ajar Rekayasa Perangkat Lunak. Medan: UMSU Press, 2021, pp. 228.

[6] Janner Simarmata Arsan Kumala Jaya, Syarifah Fitrah Ramadhani, Niel Ananto, Abdul Karim, Betrisandi, Muhammad Ilham Alhari, Cucut Susanto, Suardinata, Indah Purnama Sari, Edson Yahuda Putra. Komputer dan Masyarakat. Medan: Yayasan Kita Menulis, 2024, pp.162.

[7] Mahdianta Pandia, Indah Purnama Sari, Alexander Wirapraja Fergie Joanda Kaunang, Syarifah Fitrah Ramadhani Stenly Richard Pungus, Sudirman, Suardinata Jimmy Herawan Moedjahedy, Elly Warni, Debby Erce Sondakh. Pengantar Bahasa Pemrograman Python. Medan : Yayasan Kita Menulis, 2024, pp.180

[8] Zelvi Gustiana Arif Dwinanto, Indah Purnama Sari, Janner Simarmata Mahdianta Pandia, Supriadi Syam, Semmy Wellem Taju Fitrah Eka Susilawati, Asmah Akhriana, Rolly Junius Lontaan Fergie Joanda Kaunang. Perkembangan Teknologi Informatika. Medan: Yayasan Kita Menulis, 2024, pp.158

[9] Muhammad Hibrian Wiwi, Indah Purnama Sari, Sudirman Sudirman, Ramli Ramli, Sitti Arni, Eka Rahayu, Janner Simarmata, Boni Oktaviana, Raemon Syaljumairi, Muharman Lubis. Basis Data Terapan. Medan: Yayasan Kita Menulis, 2025.

[10] Ari Usman Muhammad Fairuzabadi, Indah Purnama Sari, Sudirman Berti Sari Br Sembiring, Cucut Susanto, Fera Damayanti Ayu Lestari Perdana, Wiranti Kusuma Hapsari. Sistem Pakar : Konsep, Model Dan Implementasi. Medan: Yayasan Kita Menulis, 2024, pp.172

Jurnal

[11] Catur Utami, D., Nur Azizah, A., & Nur Azizah, A. (2023). HUBUNGAN STATUS GIZI DENGAN PERKEMBANGAN BALITA USIA 1-5 TAHUN DI

[12] Safira, R., & Sari, I.P. (2025). Model Prediksi Perkembangan Tumbuh Kembang Anak untuk Pencegahan Risiko Gizi Buruk Menggunakan Support Vector Machine (SVM) dengan Feature Selection Backward Elimination. Blend Sains Jurnal Teknik 4 (1), 1-11

[13] Dwinanto, R. W., Sandi A, A. S., & Ardianto, R. (2024). Klasifikasi Berisiko Stunting pada Balita: Perbandingan K-Nearest Neighbor, Naïve Bayes, Support Vector Machine. METHOMIKA Jurnal Manajemen Informatika Dan Komputerisasi Akuntansi, 8(2), 264–273. https://doi.org/10.46880/jmika.Vol8No2.pp264-273

[14] F. Maulidina, Z. Rustam, S. Hartini, V. V. P. Wibowo, I. Wirasati, and W. Sadewo, “Feature optimization using Backward Elimination and Support Vector Machines (SVM) algorithm for diabetes classification,” in Journal of Physics: Conference Series, IOP Publishing Ltd, Mar. 2021. doi: 10.1088/1742-6596/1821/1/012006.

[15] Sari, I.P., Al-Khowarizmi, A.K., & Batubara, I.H. (2021). Cluster Analysis Using K-Means Algorithm and Fuzzy C-Means Clustering For Grouping Students' Abilities In Online Learning Process. Journal of Computer Science, Information Technology and Telecommunication Engineering, Vol. 2 No. 1, page 139-144

[16] Sari, I.P., Batubara, I.H., & Al-Khowarizmi, A.K. (2021). Sensitivity Of Obtaining Errors In The Combination Of Fuzzy And Neural Networks For Conducting Student Assessment On E-Learning. International Journal of Economic, Technology and Social Sciences (Injects), Vol. 2 No. 1, page 331- 338

[17] Sari, I.P., Fahroza, M.F., Mufit, M.I., & Qathrunad, I.F. (2021). Implementation of Dijkstra's Algorithm to Determine the Shortest Route in a City. Journal of Computer Science, Information Technology and Telecommunication Engineering,Vol. 2 No. 1, page 134-138

[18] Sari, I.P., Al-Khowarizmi, A.K., Ramadhani, F., & Sulaiman, O.K. (2023). Implementation of the Selection Sort Algorithm to Sort Data in PHP Programming Language. Journal of Computer Science, Information Technology and Telecommunication Engineering, Vol. 4 No. 1, page 377-381

[19] Manurung, A.A., Nasution, M.D., & Sari, I.P. (2023). Implementation of Fuzzy K-Nearest Neighbor Method in Dengue Disease Classification. 2023 11th International Conference on Cyber and IT Service Management (CITSM)

[20] Ramadhani, F., Satria, A., & Sari, I.P. (2023). Implementasi Metode Fuzzy K-Nearest Neighbor dalam Klasifikasi Penyakit Demam Berdarah. Hello World Jurnal Ilmu Komputer 2 (2), 58-62

[21] Sari, I.P., Batubara, I.H., Ramadhani, F., & Wardani, S. (2022). Perancangan Sistem Antrian pada Wahana Hiburan dengan Metode First In First Out (FIFO). Sudo Jurnal Teknik Informatika 1 (3), 116-123

[22] Az-Zahrah., A, & Sari., I.P. (2024). Perbandingan Sistem Prediksi Menggunakan Metode Monte Carlo dengan Metode K-NN pada Nilai Peserta Didik Uji Kompetensi Kejuruan. sudo Jurnal Teknik Informatika 3 (3), 127-135

[23] Ramadhani, F., Satria, A., & Sari, I.P. (2022). Aplikasi internet berbasis website sebagai E-Commerce penjualan komponen sport car. Blend Sains Jurnal Teknik 1 (2), 69-75

[24] Sari, I.P., Al-Khowarizmi, A., & Ramadhani, F. (2021). User Interface Prototype Using User Centered System Design Method in Motorvice Information System. 2021 International Conference on Computer Science and Engineering (IC2SE) 1, 1-6

[25] Ramadhani, F., Al-Khowarizmi, A.K., & Sari, I.P. (2021). Improving the Performance of Naïve Bayes Algorithm by Reducing the Attributes of Dataset Using Gain Ratio and Adaboost. 2021 International Conference on Computer Science and Engineering (IC2SE) 1, 1-5

[26] Sitompul, D.N., Rahmatika, A., & Sari, I.P. (2023). Application of The Sales and Purchase Program Using The Rapid Application Development Model. Al'adzkiya International of Computer Science and Information Technology (AIoCSIT) Journal, Vol. 4 No. 1, page 6-16

[27] Sari, I.P., Ramadhani, F., Satria, A., & Apdilah, D. (2023). Implementasi Pengolahan Citra Digital dalam Pengenalan Wajah menggunakan Algoritma PCA dan Viola Jones. Hello World Jurnal Ilmu Komputer 2 (3), 146-157

[28] Batubara, I.H., Sari, I.P., Siregar, E.F.S., & Lubis, B.S. (2021). Meningkatkan Kemampuan Penalaran Matematika Melalui Metode Penemuan Terpandu Berbantuan Software Autograph. Seminar Nasional Teknologi Edukasi Sosial dan Humaniora 1 (1), 699-705

[29] Apdilah, D., & Sari, I.P. (2021). Optimization Of The Fuzzy C-Means Cluster Center For Credit Data Grouping Using Genetic Algorithms. Al'adzkiya International of Computer Science and Information Technology (AIoCSIT) Journal, Vol. 2 No. 2, page 156-163

[30] Sulaiman, O.K., & Sari, I.P. (2021). Implementation Data Mining For Level Analysis Traffic Violation By Algorithm Association Rule. Al'adzkiya International of Computer Science and Information Technology (AIoCSIT) Journal, Vol. 2 No. 2, page 128-135

[31] Sari, I.P., Hariani, P.P ., Al-Khowarizmi, A.K., Ramadhani, F., Sulaiman, O.K., Satria, A., & Manurung, A.A. (2024). CLUSTERING HIV/AIDS DISEASE USING K-MEANS CLUSTERING ALGORITHM. Proceeding International Seminar on Islamic Studies. Vol. 5, No. 1 (2024), 1668-1676

[32] Sari, I.P., Ramadhani, F., Satria, A., & Sulaiman, O.K. (2023). Leukocoria Identification: A 5-Fold Cross Validation CNN and Adaboost Hybrid Approach. 2023 6th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI). 486-491

[33] Sari, I.P., Al-Khowarizmi, A.K., Sulaiman, O.K., & Apdilah, D. (2023). Implementation of Data Classification Using K-Means Algorithm in Clustering Stunting Cases. Journal of Computer Science, Information Technology and Telecommunication Engineering, Vol. 4 No. 2, page 402-412

[34] Sari., I.P, Batubara., I.H, Al-Khowarizmi., A, & PP Hariani. (2022). Perancangan Sistem Informasi Pengelolaan Arsip Digital Berbasis Web untuk Mengatur Sistem Kearsipan di SMK Tri Karya. Wahana Jurnal Pengabdian kepada Masyarakat 1 (1), 18-24

[35] Habibi., F, Qathrunada., I.F, & Anggraini., T. (2023). "Design and Build a Tourism Website Using Shopify Framework". Hanif Journal of Information Systems. Vol. 1 No. 1, 2023.

[36] Sari., I.P, A Syahputra, N Zaky, RU Sibuea, & Z Zakhir. (2022). Perancangan sistem aplikasi penjualan dan layanan jasa laundry sepatu berbasis website. Blend sains jurnal teknik 1 (1), 31-37

[37] Sari., I.P, A Azzahrah, FQ Isnaini, L Nurkumala, & A Thamita. (2022). Perancangan sistem absensi pegawai kantoran secara online pada website berbasis HTML dan CSS. Blend sains jurnal teknik 1 (1), 8-15

[38] Septiana., D. (2024). Forecasting Rice Prices with Holt-Winter Exponential Smoothing Model. Hanif Journal of Information Systems. Vol. 1 No. 2, 2024.

[39] Sari.,I.P, & Ramadhani., F. (2021). Pengaruh Teknologi Informasi Terhadap Kewirausahaan Pada Aplikasi Perancangan Jual Beli Jamu Berbasis WEB. Prosiding Seminar Nasional Kewirausahaan 2 (1), 874-878.

[40] Satria., A, Ramadhani., F, & Sari, I.P. (2023). Rancang Bangun Sistem Informasi Penerimaan Peserta Didik Baru (PPDB) Sekolah Menengah Kejuruan Telkom 2 Medan Menggunakan Codeigniter. Wahana Jurnal Pengabdian kepada Masyarakat 2 (1), 23-31

[41] Sari., I.P, A Jannah, AM Meuraxa, A Syahfitri, & R Omar. (2022). Perancangan Sistem Informasi Penginputan Database Mahasiswa Berbasis Web. Hello World Jurnal Ilmu Komputer 1 (2), 106-110.

[42] Mahardika., F, & Abdillah., M.L. (2024). Design of Unified Modeling Language Information System for Motorcycle Unit Selling and Buying Transactions using the Waterfall Method. Hanif Journal of Information Systems. Vol. 1 No. 2, 2024.

[43] Sari., I.P, & Batubara., I.H. (2021). Perancangan Sistem Informasi Laporan Keuangan Pada Apotek Menggunakan Algoritma K-NN. Seminar Nasional Teknologi Edukasi dan Humaniora (SiNTESa) 1 (2021 - ke 1

[44] Sari., I.P, & Batubara., I.H. (2021). User Interface Information System for Using Account Services (Joint Account) WEB-Based. International Journal of Economic, Technology and Social Sciences (Injects), 462-469

[45] Sari., I.P, Al-Khowarizmi., A, & Batubara., I.H. (2021). Implementasi Aplikasi Mobile Learning Sistem Manajemen Soal dan Ujian Berbasis Web Pada Platform Android. IHSAN: JURNAL PENGABDIAN MASYARAKAT 3 (2), 178-183

[46] Mudafri., H.A. (2024). Design of a Web-Based Coffeeshop Ordering Information System. Hanif Journal of Information Systems. Vol. 1 No. 2, 2024.

[47] Sari., I.P, Hariani., P.P, Satria., A, & Manurung., A.A. (2023). Rancang Bangun Sistem Informasi Pengelolaan Arsip Materi Ajar Berbasis Web untuk Guru MAS Darul Falah. Wahana Jurnal Pengabdian kepada Masyarakat 2 (2), 59-65

[48] Ramadhani., F, & Sari., I.P. (2021). Pemanfaatan Aplikasi Online dalam Digitalisasi Pasar Tradisional di Medan. Prosiding Seminar Nasional Kewirausahaan 2 (1), 806-811

[49] Sari., I.P, Sulaiman., O.K, & Apdillah, D. (2024). Rancang Bangun Game Zombie Menggunakan Kodular Berbasis Android. Jurnal Minfo Polgan 13 (1), 293-302

[50] Ichsan., A, Siambaton., M.Z, & Nasution., K. (2023). "Android-Based Practical Work Student Registration Form Application System Design". Hanif Journal of Information Systems. Vol. 1 No. 1, 2023.

[51] Sari., I.P, Sulaiman., O.K, Ramadhani., F, & Satria., A. (2023). Perancangan Sistem Manajemen Surat Berbasis Web Pada Kantor Camat Tano Tombangan Angkola. INCODING: Journal of Informatics and Computer Science Engineering 3 (2), 61-76.

[52] Bisono, A.T., & Zulherry, A. (2025). Analisis Sentimen Game Genshin Impact Untuk Mengetahui Reaksi Dan Harapan Pemain Menggunakan Metode Naïve Bayes. Sudo Jurnal Teknik Informatika 4(2), 183-193

[53] Zulherry, A., Gunawan, T.S., & Wanayumini, W. (2021). Analisis Hasil Pendukung Keputusan Mendapatkan Rumah Dinas Perusahaan Menggunakan Metode Analytical Hierarcy Process (AHP) Dan Teknik For Order Referenci By Similarity (Topsis). Media Informatika Budi Darma 5(2), 695-704

[54] Jannah., A, Meuraxa., A.M, & Azzahrah., A. 2023. "Web Based E-Commerce System Design at EXO Shop Using The Waterfall Method". Hanif Journal of Information Systems. Vol. 1 No. 1, 2023.

[55] Sari., I.P, Al-Khowarizmi., A, , Jannah., A, Meuraxa., A.M, & Tanjung., M.I. (2023). Web-Based Offline Game Suit Design: A Model Overview. Journal of Computer Science, Information Technology and Telecommunication Engineering 4 (2), 389-394.

[56] Guntur., S, Ichsan., A, & Sari., I.P. (2024). Designing a Web-Based Mail Management System at the Beringin Helvetia Sub-district Office. Altafani: Jurnal Pengabdian Masyarakat 1 (1)

[57] Sari., I.P, Sulaiman., O.K, Al-Khowarizmi., A, & Azhari., M. (2023). Perancangan Sistem Informasi Pelayanan Masyarakat pada Kelurahan Sipagimbar dengan Metode Prototype Berbasis Web. Blend Sains Jurnal Teknik 2 (2), 125-134.

[58] Hutasuhut., B.K, Sari., I.P, & Al-Khowarizmi, A.K. (2023). Analysis the Effect of Digitalization and Technology on Web-Based Entrepreneurship. Journal of Computer Science, Information Technology and Telecommunication Engineering

[59] Dongoran., D, & Sari., I.P. (2024). Implementasi Klasifikasi Data Tracer Study Pada Universitas Muhammadiyah Sumatera Utara Dengan Pemanfaatan Data Mining Menggunakan Kombinasi Algoritma Support Vector Machine. Hello World Jurnal Ilmu Komputer 4 (1), 12-24

[60] Sari., I.P, Azis., Z, & Hasibuan., A.R. (2025). ANALYSIS OF RON 92 OIL BASED ON MORPHOLOGY AND HISTOGRAM TECHNIQUES. BAREKENG: Jurnal Ilmu Matematika dan Terapan 19 (2), 1341-1352

[61] Aqta., D.P, & Sari., I.P. (2025). Evaluasi Keefektifan Teknik Morfologi dan Histogram Citra Digital pada Minyak RON 92 di SPBU Pertamina Medan Tembung. Hello World Jurnal Ilmu Komputer 3 (4), 161-170

[62] Yusuf., M, & Sari., I.P. (2025). Sistem Pakar Mencegah Stunting dengan Menentukan Gizi Anak Menggunakan Natural Language Processing (NLP). Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi) 9 (3), 924-934

[63] Sari., I.P, Sulaiman., O.K, & Apdilah., D. (2025). Penerapan Sistem Pakar Diagnostik Penyakit Kelapa Sawit sebagai Solusi bagi Petani dalam Meningkatkan Produktivitas Perkebunan. CivicAction: Jurnal Pengabdian dan Inovasi Masyarakat 1 (1), 8-17