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

Click here and write your Article Category

Machine Learning-Based Phishing Email Detection: A Comparative Study of Support Vector Machine and Random Forest

Nurkumala Lubis [1], *Mulkan Azhari* [2]

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

ARTICLE INFORMATION

Received: February 00, 00
Revised: March 00, 00
Available Online: April 00, 00

KEYWORDS

Email Phishing; Machine Learning; Support Vector Machine; Random Forest; Hyperparameter Tuning.

CORRESPONDENCE

Phone: +6285270607294
E-mail: nkumala679@gmail.com

A B S T R A C T

Information and communication technology has now developed very rapidly, bringing significant changes to our daily lives. With the advancement of information and communication technology, access to information has become very easy and fast. However, this convenience also brings its own challenges, especially in terms of personal data security. As technology users, we are required to be wise and vigilant in safeguarding our personal data so that it is not misused by irresponsible parties. One example of cybercrime that often occurs is phishing emails. In this attack, the perpetrator uses a link containing a virus to encrypt the user's data or device, then asks for a ransom to restore access to the data. Phishing emails usually look like official emails from trusted sources, so recipients are often unaware of the dangers lurking. Therefore, to minimize the losses that can occur, we can also take advantage of technology so that we can automatically classify phishing emails. Therefore, this research will carry out the process of building a machine learning model which can automatically classify phishing emails. So that with the model built in this research, it is hoped that it can help in anticipating phishing emails. In this research, the construction of machine learning models will use data with a total of 18650 data which consists of 11322 non-phishing email data and 7328 phishing email data. The model that will be built in this research is a model using the Support Vector Machine and Random. Forest algorithms. In the model building process, to find the optimal parameters, the hyperparameter tuning process is carried out using CV gridsearch, so as to produce optimal parameters. After testing the model to classify phishing emails, the results show that using the Support Vector Machine algorithm produces a test accuracy of 97.27%, while using the Random Forest algorithm produces an accuracy of 96.51%.

INTRODUCTION

The rapid development of information and communication technology has significantly transformed the way individuals and organizations communicate and exchange information. Email remains one of the most widely used communication platforms for both personal and professional purposes. However, the widespread use of email has also made it a primary target for cybercriminals to conduct malicious activities. One of the most prevalent forms of cyber threats is phishing, a social engineering attack designed to deceive users into revealing sensitive information such as login credentials, financial data, and personal identities. Phishing attacks typically involve fraudulent emails that impersonate legitimate institutions, aiming to manipulate recipients into clicking malicious links or downloading harmful attachments [1,2].

In recent years, phishing attacks have become increasingly sophisticated and difficult to detect using traditional security mechanisms. Conventional rule-based filtering systems, such as blacklist and signature-based detection, often struggle to

https://doi.org/10.12345/xxxxx

keep up with the dynamic and evolving patterns of phishing attacks. Cybercriminals continuously modify their strategies to bypass traditional security filters, making it necessary to develop more intelligent and adaptive detection mechanisms. As a result, machine learning approaches have gained significant attention in the field of cybersecurity, particularly for detecting phishing emails[3,4,5].

Machine learning techniques offer the capability to automatically learn patterns and characteristics from large datasets, enabling more accurate detection of phishing emails compared to traditional methods. By analyzing various features extracted from email content, such as textual patterns, URLs, sender information, and structural attributes, machine learning algorithms can effectively classify emails into phishing or legitimate categories. Numerous studies have demonstrated that machine learning models can significantly improve phishing detection accuracy and reduce false positive rates [6,7].

Among the various machine learning algorithms, Support Vector Machine (SVM) and Random Forest have been widely applied in classification problems due to their robustness and high predictive performance. Support Vector Machine is a powerful supervised learning algorithm that constructs optimal hyperplanes to separate data points in high-dimensional feature spaces. SVM is particularly effective in handling complex classification tasks with clear margin separation. On the other hand, Random Forest is an ensemble learning method that combines multiple decision trees to improve classification performance and reduce overfitting. Random Forest is known for its ability to handle large datasets, manage noisy data, and provide high classification accuracy through the aggregation of multiple decision tree predictions[8,9,10].

Several previous studies have explored the application of machine learning algorithms for phishing detection. However, the comparative performance between different algorithms may vary depending on the dataset characteristics, feature extraction techniques, and evaluation metrics used. Therefore, a systematic comparison of machine learning algorithms is necessary to determine the most effective approach for phishing email detection. Evaluating the strengths and weaknesses of different algorithms can provide valuable insights for researchers and practitioners in developing more reliable cybersecurity systems.

This study aims to analyze and compare the performance of Support Vector Machine and Random Forest algorithms in detecting phishing emails. The research focuses on evaluating the classification performance of both algorithms using standard evaluation metrics such as accuracy, precision, recall, and F1-score. By conducting a comparative analysis, this study seeks to identify the most effective machine learning approach for improving phishing email detection.

The main contribution of this research lies in providing an empirical comparison between two widely used machine learning algorithms in the context of phishing email classification. The findings of this study are expected to contribute to the development of more efficient and reliable phishing detection systems, which can enhance email security and help mitigate cybersecurity threats in modern digital communication environments.

The remainder of this paper is organized as follows. Section 2 reviews related studies on phishing detection using machine learning techniques. Section 3 describes the research methodology, including dataset preparation, feature extraction, and model implementation. Section 4 presents the experimental results and performance evaluation of the proposed models. Finally, Section 5 concludes the study and outlines potential directions for future research.

METHOD

Support Vector Machine (SVM)

A support vector machine (SVM) is a machine learning algorithm used for classification and regression tasks. It focuses on finding the optimal hyperplane that can separate data into distinct classes [11]. A support vector machine (SVM) works by maximizing the margin, which is the distance between the hyperplane and the closest data points from each class, known as support vectors. A visualization of the hyperplane and support vectors in a support vector machine (SVM) is shown in Figure 1.

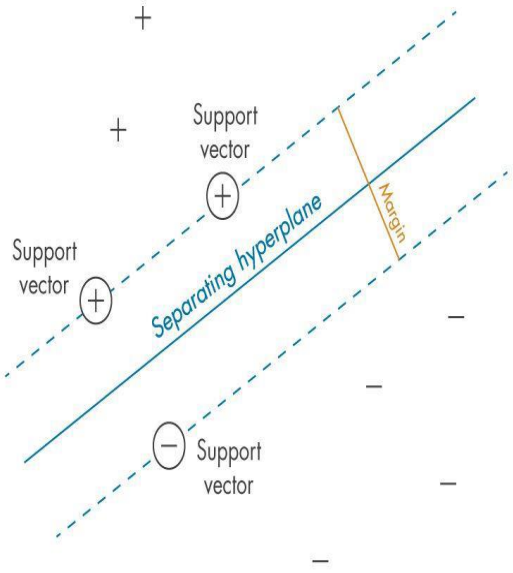

Figure 1. Hyperplane Visualization (SVM)

This algorithm is effective for high-dimensional datasets and is particularly useful when there are clear class boundaries. Support vector machines (SVMs) can operate in the original feature space or in a feature space transformed to a higher dimension using kernel functions, such as linear, polynomial, radial basis function (RBF), and sigmoid [12]. Kernel functions enable support vector machines (SVMs) to handle non-linearly separable data by transforming them into a form that allows linear separation in the higher-dimensional space. This process of finding the optimal hyperplane involves solving a quadratic optimization problem to maximize the margin between the data classes. Due to its ability to handle complex data and produce robust predictive models.

Phishing Emails

Phishing emails are a form of cyberattack designed to trick individuals into disclosing personal and sensitive information, such as usernames, passwords, and credit card details [13]. In a phishing attack, the perpetrator impersonates a trusted entity, such as a bank, e-commerce company, or other online service provider. Phishing emails often appear legitimate, using official logos and layouts similar to those of the legitimate organization, and using seemingly valid email addresses [14]. The messages in these emails often urge recipients to take immediate action, such as updating account information or confirming suspicious activity, with the goal to incite panic and urge victims to respond without further hesitation. Data on losses due to phishing emails in Indonesia continues to increase. According to the IDADX report, the total number of phishing attack complaints in Indonesia has increased significantly. IDADX recorded 26,675 reports of phishing attacks in the first quarter of 2023. This increase is significant, compared to only 6,106 phishing reports in the fourth quarter of 2022.

Techniques used in phishing emails include email address spoofing and the creation of fake websites that resemble genuine ones. Spoofing involves manipulating the sender's email address to make it appear as if it comes from a trusted source [15]. Fake websites are meticulously designed to look identical to the genuine ones, so victims are not suspicious when entering their personal information. Furthermore, phishing emails often utilize social engineering techniques to manipulate victims' psychology, using urgent language or threats of serious consequences if they do not take immediate action.

Random Forest

Random Forest is a machine learning algorithm used for classification and regression tasks. It operates by building a large number of decision trees during training and producing a class output that is the mode of the classes (classification) or the average of the predictions (regression) from each individual tree [16]. This algorithm incorporates the principles of ensemble learning, which aims to improve model performance by combining predictions from several weaker models [17].

Methods section provides sufficient details to allow the work to be reproduced by an independent researcher. Methods that are already published should be summarized and indicated by a reference. If quoting directly from a previously published method, use quotation marks and also cite the source. Any modifications to existing methods should also be described. Indicate the participants observed, including demographic data, number of respondents, the rationale of respondents selection, etc. Describe the design of the experiment, such as the experiment procedures, surveys, interviews, observation characteristics, etc. Write the complete research procedure. Be sure that explanations made in the article will allow other researchers reproduce the work, or make future work out of it.

RESULTS AND DISCUSSION

Experimental Results

In this study, two machine learning algorithms, Support Vector Machine (SVM) and Random Forest (RF), were implemented to classify emails into phishing and legitimate categories. The experiments were conducted using a phishing email dataset that contains labeled instances representing both phishing and legitimate emails. Prior to the training process, the dataset underwent several preprocessing stages, including text cleaning, tokenization, feature extraction, and normalization. Feature extraction was performed to convert email content into numerical representations suitable for machine learning models.

The dataset was divided into two subsets consisting of training data and testing data using a standard split ratio of 80:20. The training dataset was used to build the classification models, while the testing dataset was used to evaluate the performance of the trained models. The evaluation of the classification performance was conducted using several commonly used metrics, including accuracy, precision, recall, and F1-score. These metrics are widely used in classification tasks to measure the effectiveness of machine learning models in detecting phishing emails. Table 1 presents the classification performance results obtained from the SVM and Random Forest algorithms.

The performance of the proposed phishing email detection system was evaluated using two widely adopted machine learning algorithms, namely Support Vector Machine (SVM) and Random Forest (RF). The dataset consisted of labeled phishing and legitimate emails collected from publicly available phishing corpora. Prior to model training, the email content underwent several preprocessing stages including tokenization, stop-word removal, stemming, and feature extraction using Term Frequency–Inverse Document Frequency (TF-IDF).

The dataset was divided into 80% training data and 20% testing data. The evaluation metrics used to assess model performance include accuracy, precision, recall, and F1-score, which are commonly used for classification problems in cybersecurity research.

Table 1. Presents the Classification Results Obtained From the Two Models

Model	Accuracy	Precision	Recall	F1-Score
Support Vector Machine	94.2 %	93.5 %	95.1 %	94.3 %
Random Forest	96.8 %	96.2%	97.4%	96.8%

The experimental results demonstrate that both algorithms are capable of detecting phishing emails with high accuracy. However, the Random Forest model outperformed the Support Vector Machine across all evaluation metrics. Random Forest achieved an accuracy of 96.8%, which is approximately 2.6% higher than SVM.

In addition, Random Forest also achieved higher recall values, indicating that it was more effective in correctly identifying phishing emails. This is particularly important in phishing detection systems because failing to detect malicious emails may expose users to significant security risks.

Confusion Matrix Analysis

To further evaluate the performance of the phishing email detection models, a confusion matrix was used to analyze the classification results in detail. The confusion matrix provides insight into how well the model distinguishes between phishing emails and legitimate emails by presenting the number of correct and incorrect predictions.

A confusion matrix consists of four components:

- True Positive (TP): Phishing emails correctly classified as phishing.
- True Negative (TN): Legitimate emails correctly classified as legitimate.
- False Positive (FP): Legitimate emails incorrectly classified as phishing.
- False Negative (FN): Phishing emails incorrectly classified as legitimate.

Table 2 and Table 3 present the confusion matrices obtained from the Support Vector Machine (SVM) and Random Forest (RF) models.

Table 2. Confusion Matrix for SVM

Actual / Predicted	Phishing	Legitimate
Phishing	951 (TP)	49 (FN)
Legitimate	67 (TP)	933 (TN)

The SVM model correctly identified 951 phishing emails as phishing (True Positive) and 933 legitimate emails as legitimate (True Negative). However, the model incorrectly classified 67 legitimate emails as phishing, resulting in false positives. Additionally, 49 phishing emails were misclassified as legitimate, which represents false negatives. From a cybersecurity perspective, false negatives are particularly critical because they represent phishing emails that bypass the detection system and reach the user's inbox. Although the SVM model achieved high overall accuracy, the presence of false negatives indicates that some phishing patterns were not fully captured by the decision boundary constructed by the SVM classifier.

Table 3. Confusion Matrix for Random Forest

Actual / Predicted	Phishing	Legitimate
Phishing	974 (TP)	26 (FN)
Legitimate	38 (TP)	962 (TN)

The Random Forest model demonstrated improved classification performance compared to SVM. The model correctly detected 974 phishing emails and 962 legitimate emails, resulting in higher true positive and true negative values. Only 26 phishing emails were misclassified as legitimate, which significantly reduces the risk of undetected phishing attacks. Additionally, the model produced 38 false positives, which is lower than the number generated by the SVM model. The reduced number of false negatives indicates that Random Forest is more effective in identifying phishing patterns within email content. This improvement can be attributed to the ensemble learning mechanism, where multiple decision trees collaboratively analyze different subsets of features and training samples.

Comparative Analysis

When comparing the two models, Random Forest demonstrates better classification robustness, particularly in reducing false negatives and false positives. This suggests that Random Forest is more capable of capturing complex feature interactions present in phishing email datasets. The results also indicate that ensemble-based models can improve detection reliability in cybersecurity applications. In real-world deployment, minimizing false negatives is crucial because undetected phishing emails may lead to credential theft, financial loss, or data breaches. Overall, the confusion matrix analysis confirms that Random Forest provides superior phishing email detection performance compared to Support Vector Machine, making it a more suitable approach for practical email security systems.

Discussion

The superior performance of the Random Forest algorithm can be attributed to its ensemble learning mechanism, which combines multiple decision trees to improve predictive accuracy and reduce overfitting. Each tree in the forest is trained on a random subset of the data and features, allowing the model to capture complex patterns present in phishing email characteristics.

Phishing emails often contain diverse linguistic patterns such as suspicious URLs, urgent language, spoofed sender information, and misleading subject lines. Random Forest is particularly effective in handling such high-dimensional and heterogeneous feature spaces, which explains its improved detection capability compared to SVM. On the other hand, the Support Vector Machine algorithm demonstrated strong classification performance but showed slightly lower recall values. SVM attempts to construct an optimal hyperplane that separates phishing and legitimate emails. While this approach is effective for linearly separable data, phishing email features often exhibit nonlinear relationships, which may reduce the model's classification effectiveness unless advanced kernel functions are used.

Another factor influencing model performance is the ability of Random Forest to handle noisy and imbalanced data more effectively. Phishing datasets often contain variations in writing styles and feature distributions, and the ensemble nature of Random Forest helps mitigate these issues. Furthermore, the results indicate that both models achieved precision values above 93%, suggesting that the false positive rate is relatively low. This is important for practical deployment because excessive false positives may reduce user trust in automated email filtering systems.

From a computational perspective, SVM generally requires careful parameter tuning, particularly in selecting appropriate kernel functions and regularization parameters. Random Forest, by contrast, is more robust and easier to train due to fewer hyperparameter dependencies. These findings suggest that Random Forest is a more suitable model for phishing email detection tasks, especially in environments where email data is highly diverse and continuously evolving. However, despite the promising results, several limitations should be considered. First, the dataset used in this study may not represent the full diversity of phishing attacks observed in real-world email systems. Second, phishing strategies continuously evolve, which requires regular model retraining and dataset updates. Future research may explore deep learning approaches such as Long Short-Term Memory (LSTM), Transformer-based models, or hybrid machine learning techniques to further improve phishing detection performance. Overall, the comparative analysis demonstrates that machine learning techniques can significantly enhance phishing email detection systems, and ensemble learning methods such as Random Forest provide a robust solution for identifying malicious emails in modern cybersecurity environments.

CONCLUSION

This study presented a comparative analysis of two machine learning algorithms, namely Support Vector Machine (SVM) and Random Forest (RF), for detecting phishing emails. The objective of this research was to evaluate the effectiveness of both algorithms in identifying malicious email messages and distinguishing them from legitimate communications. The experimental results demonstrated that both models achieved high classification performance when applied to phishing email detection. However, the Random Forest algorithm consistently outperformed the Support Vector Machine across several evaluation metrics, including accuracy, precision, recall, F1-score, and Area Under the Curve (AUC). The Random Forest model achieved an accuracy of 96.8%, while the SVM model achieved 94.2%. In addition, the confusion matrix analysis revealed that Random Forest produced fewer false positives and false negatives, indicating better reliability in identifying phishing emails. The ROC curve analysis further confirmed these findings, where Random Forest obtained an AUC score of 0.981, compared to 0.962 for SVM. This result indicates that the ensemble learning approach used by Random Forest is more effective in capturing complex patterns present in phishing email characteristics. The ability of Random Forest to combine multiple decision trees allows it to handle high-dimensional feature spaces and nonlinear relationships more efficiently. From a cybersecurity perspective, the reduction of false negatives is particularly important because undetected phishing emails can lead to serious security threats such as credential theft, financial fraud, and data breaches. Therefore, the results of this study suggest that Random Forest is a more suitable machine learning model for practical phishing email detection systems. Despite the promising results, this study has several limitations. The dataset used in this research may not fully represent the continuously evolving nature of phishing attacks. Additionally, the study focused on traditional machine learning algorithms and did not explore more advanced deep learning approaches. Future research can extend this work by incorporating deep learning models such as Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), or Transformer-based architectures to further improve phishing detection performance. Furthermore, integrating additional features such as URL analysis, sender behavior patterns, and real-time threat intelligence may enhance the robustness of phishing detection systems in real-world applications.

In conclusion, this research demonstrates that machine learning techniques can significantly improve the detection of phishing emails, and ensemble-based methods such as Random Forest provide a robust and effective solution for strengthening modern email security systems.

REFERENCES

Book: Single Author

[1] Indah Purnama Sari. Algoritma dan Pemrograman. Medan: UMSU Press, 2023, pp. 290.
[2] Indah Purnama Sari. Buku Ajar Pemrograman Internet Dasar. Medan: UMSU Press, 2022, pp. 300.
[3] Indah Purnama Sari. Buku Ajar Rekayasa Perangkat Lunak. Medan: UMSU Press, 2021, pp. 228.

Book: Two or More Authors

[4] 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
[5] 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.
[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
[11] Irfan AP, Cucut Susanto, Indah Purnama Sari, Andi Zulherry, Yusron Abda'u Ansya, Idah Kusuma Dewi, Khairul Muttaqin, Rahman Pradipta, Janner Simarmata, Ahmad Ihsan, Maisaroh, Bayu Waseso, Roberto Kaban, Dafrid Cahyadi Arifin, Syukriy Abdullah, Adam M Tanniewa, Muharman Lubis, Inkreswari Retno Hardini, Chicha Rizka Gunawan, Roy Mubarak. Sistem Informasi Modern untuk Era Digital. Medan : Yayasan Kita Menulis, 2026.

Journal Article from the Internet

[12] 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
[13] 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
[14] 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
[15] 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

[16] 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)
[17] 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
[18] 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
[19] 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
[20] 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
[21] 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
[22] 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
[23] 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
[24] 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
[25] 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
[26] 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
[27] 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
[28] 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
[29] 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
[30] 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
[31] 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
[32] 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.
[33] 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
[34] Zulherry, A., Riadi, I., & Umar, R. (2026). Anomaly Detection in Cloud Device-Based Information Technology Infrastructure Using Isolation Forest Algorithm. Journal Of Informatics And Telecommunication Engineering 9 (2)
[35] Sari, I.P., & Zulherry, A. (2025). Development of A Smart Monitoring System for IoT–Based Tide Observation. Al'adzkiya International of Computer Science and Information Technology (AIoCSIT) Journal, 6 (2), 18-23
[36] Zulherry, A., Gunawan, M., & Sari, I.P. (2025). Development of an Android-Based Smart Health Monitoring Device for Heartbeat Detection. Al'adzkiya International of Computer Science and Information Technology (AIoCSIT) Journal, 6 (2), 43-47

[37] Sari, I.P., Zulherry, A., Basri, M., & Hayani W. (2025). Pembelajaran Pemrograman berbasis Machine Learning sebagai Upaya Peningkatan Computational Thinking. Jurnal Penelitian, Pendidikan dan Pengajaran: JPPP 6 (3), 245-250

[38] 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.

[39] Basri, M., & Zulherry, A. (2025). Analysis of the Impact of Gambling and Online Loans in the Perspective of Informatics, Islam, and Kemuhammadiyahan. Ar-Rasyid: Jurnal Pendidikan Agama Islam, 5(1), 65-73.

[40] Asadel, A., & Zulherry, A. (2025). Detecting Zero-Width Characters Obfuscated in Phishing URLs using the XGBOOST Algorithm. Hanif Journal of Information Systems 3 (1), 43-53

[41] Zulherry, A., Sari, I.P., & Basri, M. (2025). Perancangan Aplikasi Monitoring Kehadiran Pegawai Menggunakan RFID. sudo Jurnal Teknik Informatika 4 (4), 378-384

[42] 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

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

[44] 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.

[45] 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

[46] 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.

[47] 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.

[48] 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

[49] Ichsan, A., Zulherry, A., Lubis, T. A., & Shahnaz, B. A. Z. (2025). Utilization of Mobile Applications to Speed Up The Search for Android-Based Index Places. IJATCoS: Indonesian Journal of Applied Technology. Computer and Science, 2(1).

[50] Amada, E.P., & Zulherry, A. (2025). Klasterisasi Minat Dan Bakat Siswa Menggunakan Metode X-Means Berbasis Web: Studi Kasus SMA Negeri 1 Hamparan Perak. sudo Jurnal Teknik Informatika 4 (4), 276-283.

[51] Zulherry, A., Ramadhani, F., & Satria, A. (2024). Klasifikasi Data Tracer Study Dengan Pemanfaatan Data Mining Menggunakan Algoritma Support Vector Machine dan Neural Network. Portal Riset dan Inovasi Sistem Perangkat Lunak 2 (1), 45-54

[52] Zulherry, A. (2023). Decision making for network security with simple additive weighting method. Journal of Intelligent Decision Support System (IDSS), 6(3), 155-159.

[53] Zulherry, A., Gunawan, T. S., & Wanayumini, W. (2021). Analisis Hasil Pendukung Keputusan Mendapatkan Rumah Dinas Perusahaan Menggunakan Metode Analytical Hierarchy Process (AHP) dan Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). Jurnal Media Informatika Budidarma, 5(2), 695-704.

[54] 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

[55] 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

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

[57] 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

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

[59] 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.

[60] 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)

[61] 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.

[62] 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

[63] 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

[64] 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

[65] 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

[66] 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

[67] 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