

b_URLs_phishing_detection_model_with_random_forest_algorithm.pdf

by Ahmad Faza

Submission date: 06-Oct-2026 01:57PM (UTC+0700)

Submission ID: 3052015546

File name: b_URLs_phishing_detection_model_with_random_forest_algorithm.pdf (312.21K)

Word count: 3398

Character count: 18483

Web URLs Phishing Detection Model with Random Forest Algorithm

1st Aulia Kharisma Putri
Information System
Universitas Multimedia Nusantara
Tangerang, Indonesia
aulia.putri@student.umn.ac.id

4th Santo Fernandi Wijaya
Information System
Universitas Multimedia Nusantara
Tangerang, Indonesia
santo.fernandi@umn.ac.id

6th Jansen Wiratama
Information System
Universitas Multimedia Nusantara
Tangerang, Indonesia
jansen.wiratama@umn.ac.id

14th Monika Evelin Johan
Information System
Universitas Multimedia Nusantara
Tangerang, Indonesia
monika.evelin@umn.ac.id

3rd Samuel Ady Sanjaya
Information System
Universitas Multimedia Nusantara
Tangerang, Indonesia
samuel.ady@umn.ac.id

6th Ahmad Faza
Information System
Universitas Multimedia Nusantara
Tangerang, Indonesia
ahmad.faza@umn.ac.id

Abstract— As internet users grow and technology evolves, so do the security risks, one example being phishing. Phishing is an attempt to obtain important information from someone, such as username, password, and other sensitive data, by providing a fake website that resembles the original. This research focuses on the problem of phishing website URLs that are increasing in number. Creating a model using an Algorithm that can detect phishing website URLs. The classification algorithm model that will be used in this research is Random Forest, which will be evaluated based on the confusion matrix value. The accuracy result is 99%. Second, the f1 score test result is 99.1%. The third result of recall testing is 99.3%. The last test result is a precision of 98.9%. With high accuracy, f1 score, recall, and precision values, the model created using the Random Forest algorithm can be applied well to applications in web URLs, phishing detecting, analyzing fake URL patterns, and identifying suspected links as fake web URLs.

Keywords— CRISP-DM, phishing detection, random forest, web URLs

I. INTRODUCTION

Information technology is developing rapidly along with the times, especially in the field of Technology and Information. Currently, it has experienced rapid progress and provides various benefits. This rapid improvement of technology makes almost all activities easier, especially using the internet, especially through website media [1]. On the other hand, this convenience can be an opportunity for security risks for those who are still not familiar with online transaction security procedures. The security risk can be exploited by internet criminals to obtain confidential information, such as personal data, e-mail passwords, and even financial information, such as credit card data and online banking accounts, without the internet user's knowledge [2].

As internet users grow and technology develops, risks to security are increasingly diverse. One example is the practice of phishing. Phishing itself is an attempt to obtain crucial information from an individual, such as usernames, passwords, and other sensitive data, by providing a fake website that resembles the original [3]. Phishing websites will be carefully designed by internet criminals to resemble genuine sites, including appearance, content, domain URLs, and other elements, with the aim of deceiving victims (internet users). The main goal is to make victims believe that they are accessing a legitimate website page.

The impact of phishing can include financial losses and data loss and cause significant losses to victims. Therefore, it is necessary to detect links or uniform resource locators (URLs). Phishing detection is a way to find out whether a website URL address is fake or real. There are several ways to assess how safe a URL is, such as using blacklists, allowlists, statistics, or machine learning technology. Among all, machine learning technology is better because it is more efficient and accurate. This technology uses a special algorithm model to understand dangerous URL patterns and is able to detect the type of URL, whether it is a phishing or a safe site, according to needs [4].

HTML content analysis allows monitoring suspicious changes in the structure of web pages, while domain authority evaluation provides an idea of the legitimacy of the information source. Additionally, observation of script code helps in detecting suspicious activity or manipulation attempts that could harm users. By combining these three aspects, data mining models help create a strong layer of protection to identify and prevent URL-based phishing attacks efficiently [5].

Some of the functions of data mining include analysis of associations between data, data classification, data clustering, prediction, and others. This research uses a data mining-based model that aims to increase accuracy in classifying phishing URLs in the context of web security. By utilizing a data mining model, this model is designed to accurately detect and classify URLs that have the characteristics of phishing attacks. Data classification is a process of finding a model or function that can explain and differentiate data classes and concepts [6]. In this research, the functionality applied is data classification and uses a Random Forest Algorithm.

II. METHODOLOGY

Based on the dataset obtained, the method for solving the problem was determined using the Random Forest (RF), Support Machine Vector (SVM), and K-Nearest Neighbor (KNN) algorithms to determine the most accurate results. The determination of this method or algorithm is based on relevant literature studies and adapted to the needs and characteristics of existing data [7], [8]. After determining the approach, the processing process begins with steps organized based on the CRISP-DM framework. CRISP-DM (Cross Industry Standard Process for Data Mining) is a framework that is widely used in various industries to carry out data science processes. This methodology details the stages and tasks

involved in a data science project and explains the relationship between each task [7].

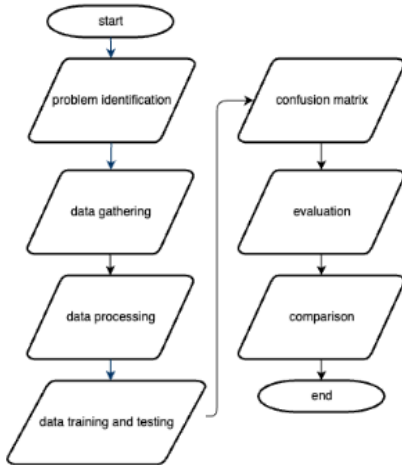


Fig 1. Research Flow

A. Business Understanding

The first stage of this research focuses on analyzing phishing cases in Indonesia and researching the corresponding data. Business Understanding in phishing web URL detection includes an in-depth understanding of web security threats, organizational needs regarding online security, and the impact of phishing attacks. The goal is to identify key aspects related to preventing, detecting, and protecting against phishing attacks on web URLs. This research aims to detect phishing website URLs using classification methods such as Random Forest, K-Nearest Neighbors, and Support Vector Machine.

Random Forest is a learning algorithm that is known for its ability to produce accurate predictions, especially in a framework with high-dimensional data [9]. The main advantage of Random Forest is its ability to overcome the overfitting problem that often occurs in complex models and its ability to perform well without requiring very specific parameter tuning [10]. This advantage makes Random Forest a popular choice in data analysis, especially in situations where high prediction accuracy is required without sacrificing model complexity or requiring complicated parameter tuning [11].

Support Vector Machine (SVM) is a machine learning algorithm used to create classification and regression models. The goal is to find the best-dividing line between two classes of data [12]. This algorithm has key parameters such as kernel parameters that allow data transformation into a higher dimensional space to increase class separation, as well as penalty parameters that regulate the trade-off between classification error on the training data and the separation margin between classes [13].

K-Nearest Neighbors is an algorithm in machine learning that classifies data based on similarity to previously existing training data [14]. This KNN method utilizes the principle that new data will be placed in the same class as the majority of classes of its nearest neighbors in feature space. This method is often used for solving classification problems, where data will be placed in categories based on the majority of categories from its nearest neighbors in feature space [15].

B. Data Understanding

The second stage of this research focuses on retrieving data from the Kaggle platform [24]. Kaggle is a data science platform that offers access to daily and weekly time series that include exogenous variables as well as business hierarchy information [16]. This process starts with automation on the platform to get listing links and carry out the scraping process. Kaggle was chosen as the data source because the datasets there have undergone curation and pre-processing, ensuring data quality and integrity.

TABLE I. SAMPLE OF PHISHING DATASET

Index	UsingIP	LongURL	ShortURL	Symbol@	Redirecting
0	1	1	1	1	1
1	1	1	0	1	1
2	1	0	1	1	1
3	1	0	-1	1	1
4	-1	0	-1	1	-1
5	1	0	-1	1	1
6	1	0	1	1	1
7	1	0	-1	1	1
8	1	1	-1	1	1
9	1	1	1	1	1
10	1	1	-1	1	1
11	-1	1	-1	1	-1
12	1	1	-1	1	1
13	1	1	-1	1	1
14	1	-1	-1	-1	1
15	1	-1	-1	1	1
16	1	-1	1	1	1

The result is a dataset with 34 attributes, including 32 attributes that can be used for analysis such as Class, Using IP, Long URL, Short URL, Symbol@, Redirecting, Prefix Suffix, Sub Domains, HTTPS, Domain Reg Len, Favicon, Non-Std Port, HTTPS Domain URL, Request URL, Anchor URL, Links In Script Tags, Server From Handler, Email Info, Abnormal URL, Website Forwarding, Status Bar Cust, Disable Right Click, Using Popup Window, Iframe Redirection, Age of Domain, DNS Recording, Website Traffic, Page Rank, Google Index, Links Pointing to Page, Stats Report. The total amount of data in this dataset reaches more than 11,000 data.

C. Data Preparation

In the third phase of the research, the focus was on data cleaning and processing. Here is a detailed breakdown of this phase: Missing Values Check: The dataset was examined for missing or incomplete values. This step is crucial to ensure the integrity and quality of the data. The missing values were checked using the ISNA () function in the dataset, and the sum of missing values for each attribute was calculated and displayed. The output shows no missing values for any attributes, as indicated by the zeroes next to each attribute name.

Data Description: The data was described to determine its types and attributes. This step involves identifying whether the data is numerical or categorical, which helps decide the appropriate processing techniques. The attributes in the dataset were listed, showing a total of 32 attributes. These attributes are shown in Fig 2:

```
#cek missing value

missing_values = data.isna().sum()
print(missing_values)

Index      0
UsingIP    0
LongURL    0
ShortURL   0
Symbol@    0
Redirecting// 0
PrefixSuffix- 0
SubDomains 0
HTTPS      0
DomainRegLen 0
Favicon    0
NonStdPort 0
HTTPSDomainURL 0
RequestURL 0
AnchorURL  0
LinksInScriptTags 0
ServerFormHandler 0
InfoEmail  0
AbnormalURL 0
WebsiteForwarding 0
```

Fig 2. Missing Value Checking

The dataset comprises 32 attributes and more than 11,000 data points. The data preparation process was divided into two main parts are shown in Fig 3:

```
# Data Split
X = data_satu.drop(["class"],axis = 1)
y = data_satu["class"]
```

Fig 3. Data Split

Data Cleansing involves removing any inconsistencies, errors, or irrelevant parts of the data. Given that no missing values were found, the data was already complete. Data Splitting involves dividing the dataset into subsets, typically for training and testing purposes in machine learning applications. Overall, the data preparation phase ensured that the dataset was ready for analysis and further processing, maintaining a high data quality and integrity standard.

D. Data Modelling

The fourth stage of this research includes modeling, which involves selecting the model and algorithm to be used, as well as implementing the algorithm. This research uses a modeling in the form of classification by displaying accuracy, precision, recall, and F1 score values. At this stage, the data model is adjusted as needed to achieve the desired results. In this research, three classification algorithms were chosen, namely Random Forest, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). The selection of the algorithm is based on evaluating the advantages, disadvantages, and information from previous research related to predicting phishing website URLs using data mining. The modeling stage was carried out using Google Colab tools. Google Collaboratory is a free cloud computing service provided by Google. The advantage of Google Colab lies in its ease of creating data visualizations because it does not require software installation on the user's computer [17].

E. Evaluation

The fifth stage of the research involved implementation using Python. Python, a programming language known for its ease of use, Python has gained immense popularity in both industrial and academic circles. The advantage of this

language lies in its ability to be interpreted and executed by computers easily and to be accessed for free [18]. Python has a number of advantages and features that differentiate it from other programming languages [19]. This research uses the metrics A (Accuracy), P (Precision), R (Recall), and F1 Score to assess the performance of the classification model that has been built. Confusion matrix is a useful evaluation technique in the classification and prediction process [20].

Accuracy (A) describes how precise the algorithm model is in making accurate predictions. This accuracy is the proportion of correct predictions (both positive and negative) compared to the overall data [21].

$$A = \frac{(TP + FN)}{(TP + FN + FP + FN)} \times 100\% \quad (1)$$

Precision (P) reflects the level of accuracy between the requested data and the model prediction results by comparing true positive predictions to the overall positive predicted results [22]. This precision aims to reduce the number of positive examples that are classified incorrectly.

$$P = \frac{TP}{FP + TP} \times 100\% \quad (2)$$

Recall (R), which is often referred to as sensitivity or true positive rate, evaluates the fraction of positive cases that are actually correctly identified by the model [21]. This recall becomes crucial when the cost of false negatives is high, as the focus is on reducing the number of missed positive examples.

$$R = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

F1 Score is a balanced measure that considers both precision and recall. This F1 Score is useful when dealing with imbalanced datasets, where one class may be more dominant than another [22]. The F1 score calculates the harmonic mean of precision and recall, providing a single value that balances between false positives (FP) and false negatives (FN).

$$F1 = 2 * \frac{(recall + precision)}{(recall + precision)} \quad (4)$$

```
features = ['AnchorURL', 'HTTPS', 'websiteTraffic', 'LinksInScriptTags', 'SubDomains']
target = 'class'

# Pisahkan data menjadi data training dan data testing
X_train, X_test, y_train, y_test = train_test_split(df_satu(features), df_satu(target), test_size=0.2, random_state=42)

# Data fitting untuk beberapa algoritma
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

Fig 4. Data Modelling Evaluation

Evaluation is the assessment stage of the model that has been created at the modeling stage. The main focus is to ensure that the results obtained are in line with previously established business objectives. By using these metrics, this research can measure the extent to which the model can provide accurate and relevant predictions in identifying phishing website URLs.

III. RESULTS AND DISCUSSION

After carrying out the stages in CRISP-DM, a confusion matrix value is produced to measure the evaluation performance of each machine learning model. The following are the resulting values:

```

SVM Metrics:
Accuracy: 0.9280868385345997
Precision: 0.9304
Recall: 0.9417004048582996
F1 Score: 0.9360160965794769

Random Forest Metrics:
Accuracy: 0.9371325192220714
Precision: 0.9328593996840442
Recall: 0.9562753036437247
F1 Score: 0.944222311075571

KNN Metrics:
Accuracy: 0.9271822704658526
Precision: 0.9182242990654206
Recall: 0.9546558704453442
F1 Score: 0.9360857483128225

```

Fig 5. Confusion Matrix Results

The evaluation results that have been presented illustrate the relative performance of the three models in classifying phishing URL detection. Analysis of these values provides a comprehensive view of the strengths and weaknesses of each model in the context of phishing URL detection.

TABLE II. RESULTS OF ALGORITHMS PERFORMANCE

Algorithms	Accuracy	Precision	Recall	F1 Score
SVM	0.928	0.934	0.941	0.936
RF	0.937	0.932	0.956	0.944
KNN	0.927	0.918	0.954	0.936

In the Random Forest model, there is an increase in performance with an accuracy of 0.937, indicating the correct level of predictions from this model. A precision of 0.933 indicates the model's level of accuracy in classifying positive data, while recall reaches 0.956 indicating the model's ability to find positive data instances. The F1 score of 0.944 shows a balance between precision and recall in the Random Forest model.

IV. CONCLUSION

In this research, we have investigated important aspects of detecting phishing websites, exploring various machine learning techniques and their effectiveness in dealing with the ever-growing cybersecurity threat of phishing attacks. The increasing level of fraudulent activity has become a major challenge for individuals and organizations around the world, driving the need to develop robust and efficient methods for identifying and preventing these fraudulent websites.

This analysis confirms that the superiority of Random Forest depends not only on the overall availability of attributes but also on its ability to adapt to the most relevant attributes. This result analysis indicates that Random Forest has a better capacity to understand patterns in data in a more adaptive way than SVM and KNN. In the context of developing a model for web URL phishing detection, the selection of appropriate attributes plays a very important role. Although all algorithms show good potential, Random Forest's ability to adapt to the most significant information underlines the importance of selecting relevant features in the

development of a reliable and effective phishing detection model. This model can be a basis for optimizing models that will be implemented in phishing detection practices in a wider environment.

ACKNOWLEDGMENT

We extend our heartfelt gratitude to Universitas Multimedia Nusantara for their invaluable support, which played a pivotal role in the successful completion of this research endeavor. Their substantial contribution was instrumental in achieving our objectives, and we are deeply grateful for their unwavering assistance.

REFERENCES

- [1] P. Subarkah and A. N. Ikhsan, "Identifikasi Website Phishing Menggunakan Algoritma Classification And Regression Trees (CART)," *J. Ilm. Inform.*, vol. 6, no. 2, pp. 127–136, 2021, doi: 10.35316/jimi.v6i2.1342.
- [2] N. B. Putri and A. W. Wijayanto, "Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing," *Komputika J. Sist. Komput.*, vol. 11, no. 1, pp. 59–66, 2022, doi: 10.34010/komputika.v11i1.4350.
- [3] A. N. I. Pungkas Subarkah, "Aplikasi Pendeteksi Website Phishing Menggunakan Machine Learning," *J. Ilm. Inform. Inform.*, 2020.
- [4] S. Saxena, A. Shrivastava, and V. Birchha, "A Proposal on Phishing URL Classification for Web Security," *Int. J. Comput. Appl.*, vol. 178, no. 39, pp. 47–49, 2019, doi: 10.5120/ijca2019919282.
- [5] F. Carroll, J. A. Adejobi, and R. Montasari, "How Good Are We at Detecting a Phishing Attack? Investigating the Evolving Phishing Attack Email and Why It Continues to Successfully Deceive Society," *SN Comput. Sci.*, vol. 3, no. 2, p. 170, 2022, doi: 10.1007/s42979-022-01069-1.
- [6] M. W. Shaukat, R. Amin, M. M. A. Muslam, A. H. Alshehri, and J. Xie, "A Hybrid Approach for Alluring Ads Phishing Attack Detection Using Machine Learning," *Sensors*, vol. 23, no. 19, pp. 1–27, 2023, doi: 10.3390/s23198070.
- [7] D. A. Kristiyanti and S. Hardani, "Sentiment Analysis of Public Acceptance of Covid-19 Vaccines Types in Indonesia using Naïve Bayes, Support Vector Machine, and Long Short-Term Memory (LSTM)," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 3 SE-Information Systems Engineering Articles, Jun. 2023, doi: 10.29207/resti.v7i3.4737.
- [8] M. H. Fadly and M. E. Johan, "Web-Based Heart Disease Prediction by Comparison and Implementation of SVM, AdaBoost, and Hybrid SVM-AdaBoost Algorithms," in *2023 7th International Conference on New Media Studies (CONMEDIA)*, 2023, pp. 257–262, doi: 10.1109/CONMEDIA60526.2023.10428512.
- [9] H. Wang and G. Wang, "Improving random forest algorithm by Lasso method," *J. Stat. Comput. Simul.*, vol. 91, no. 2, pp. 353–367, 2021, doi: 10.1080/00949655.2020.1814776.
- [10] W. Feng, C. Ma, G. Zhao, and R. Zhang, "FSRF: An Improved Random Forest for Classification," *Proc. 2020 IEEE Int. Conf. Adv. Electr. Eng. Comput. Appl. AEECA 2020*, pp. 173–178, 2020, doi: 10.1109/AEECA49918.2020.9213456.
- [11] C. Slimani, "RaFIO: A Random Forest I / O-Aware algorithm," *SIGAPP Symp. Appl. Comput.*, no. 1, pp. 521–528, 2021.
- [12] E. Ameisen and an O. M. C. Safari, *Building Machine Learning Powered Applications*. 2020.
- [13] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, no. xxxx, pp. 189–215, 2020, doi: 10.1016/j.neucom.2019.10.118.
- [14] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A brief review of nearest neighbor algorithm for learning and classification,"

- 2019 *Int. Conf. Intell. Comput. Control Syst. ICCS 2019*, no. Iciccs, pp. 1255–1260, 2019, doi: 10.1109/ICCS45141.2019.9065747.
- [15] S. Zhang, “Challenges in KNN Classification,” *IEEE Trans. Knowl. Data Eng.*, pp. 1–13, 2021, doi: 10.1109/TKDE.2021.3049250.
- [16] C. S. Bojer and J. P. Meldgaard, “Kaggle forecasting competitions: An overlooked learning opportunity,” *Int. J. Forecast.*, vol. 37, no. 2, pp. 587–603, 2021, doi: 10.1016/j.ijforecast.2020.07.007.
- [17] R. Gelar Guntara, “Visualisasi Data Laporan Penjualan Toko Online Melalui Pendekatan Data Science Menggunakan Google Colab,” *J. Ilm. Multidisiplin*, vol. 2, no. 6, pp. 2091–2100, 2023.
- [18] Muhammad Romzi and B. Kurniawan, “Pembelajaran Pemrograman Python Dengan Pendekatan Logika Algoritma,” *JTIM J. Tek. Inform. Mahakarya*, vol. 03, no. 2, pp. 37–44, 2020.
- [19] V. Thangarajah, “Python current trend applications-an overview,” Oct. 2019.
- [20] R. Ridho and H. Hendra, “Klasifikasi Diagnosis Penyakit Covid-19 Menggunakan Metode Decision Tree,” *JUST IT J. Sist. Informasi, Teknol. Inf. dan Komput.*, vol. 11, no. 3, pp. 69–75, 2022.
- [21] I. Sapuan, M. H. Fauzan, and C. Juliane, “Implementasi Data Mining untuk Klasterisasi dan Prediksi Kelompok Keluarga,” *JTERA (Jurnal Teknol. Rekayasa)*, vol. 7, no. 1, p. 149, 2022, doi: 10.31544/jtera.v7.i1.2022.149-156.
- [22] K. Omari, “Comparative Study of Machine Learning Algorithms for Phishing Website Detection,” *IJACS*, vol. 14, No 9, 2023, doi: 10.14569/IJACSA.2023.0140945.

ORIGINALITY REPORT

20%

SIMILARITY INDEX

14%

INTERNET SOURCES

15%

PUBLICATIONS

6%

STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universitas Multimedia Nusantara Student Paper	1%
2	sifisherlessciences.com Internet Source	1%
3	journal.binus.ac.id Internet Source	1%
4	Md. Abu Ashraf Siddiq, Mohammad Arifuzzaman, M. S. Islam. "Phishing Website Detection using Deep Learning", Proceedings of the 2nd International Conference on Computing Advancements, 2022 Publication	1%
5	Prismahardi Aji Riyantoko, Sugiarto, I Gede Susrama Mas Diyasa, Kraugusteeliana. "'F.Q.A.M" Feyn-QLattice Automation Modelling: Python Module of Machine Learning for Data Classification in Water Potability", 2021 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS, 2021 Publication	1%
6	Stevanus Pungky, Jansen Wiratama. "Integrating Machine Learning Based Sales Forecasting With Odoo Erp for Automated Inventory Management in a Retail Company", 2025 4th International Conference on Electronics Representation and Algorithm (ICERA), 2025 Publication	1%
7	www.ijraset.com Internet Source	1%
8	journal2.uad.ac.id Internet Source	1%
9	ai-terms-glossary.com Internet Source	1%
10	Umi Khultsum, Annida Purnamawati, Eka Rahmawati. "Comparative Study of Mobilenet and Densenet for Chest CT Scan-Based Lung Cancer Image Classification", 2025 International Conference on Information Technology Research and Innovation (ICITRI), 2025 Publication	1%

11	jurnal.unsil.ac.id Internet Source	1 %
12	Sibel Kapan, Efnan Sora Gunal. "Improved Phishing Attack Detection with Machine Learning: A Comprehensive Evaluation of Classifiers and Features", Applied Sciences, 2023 Publication	1 %
13	blog.coforge.com Internet Source	1 %
14	Dandi, Monika Evelin Johan. "Detecting Hate Speech on Memes using FixEfficientNet-L2", 2021 6th International Conference on New Media Studies (CONMEDIA), 2021 Publication	1 %
15	Kamal Omari, Ayoub Oukhatar. "Enhancing phishing website detection: a comparative study of SMOTETomek-XGB and SMOTEENN-XGB", IAES International Journal of Artificial Intelligence (IJ-AI), 2026 Publication	1 %
16	Alsharif, Bader. "Advances in Real-Time American Sign Language Recognition System Using Deep Learning Techniques for Enhanced Accessibility.", Florida Atlantic University Publication	<1 %
17	dergipark.org.tr Internet Source	<1 %
18	Rihab Fannouch, Saïd Tounsi. "AI-based analysis of key predictors of CO2 emissions across income groups", iScience, 2026 Publication	<1 %
19	Submitted to University of Sydney Student Paper	<1 %
20	Muhammad Hilman Fadly, Monika Evelin Johan. "Web-Based Heart Disease Prediction by Comparison and Implementation of SVM, AdaBoost, and Hybrid SVM-AdaBoost Algorithms", 2023 7th International Conference on New Media Studies (CONMEDIA), 2023 Publication	<1 %
21	jurnal.polibatam.ac.id Internet Source	<1 %
22	www.frontiersin.org Internet Source	<1 %
23	citedness.org Internet Source	<1 %

24	econpapers.repec.org Internet Source	<1 %
25	ramchandrapaudel.com Internet Source	<1 %
26	Bambang Susilo, Riri Fitri Sari. "Intrusion Detection in IoT Networks Using Deep Learning Algorithm", Information, 2020 Publication	<1 %
27	Rizki Rahman, Ferian Fauzi Abdulloh. "Performance of Various Naïve Bayes Using GridSearch Approach In Phishing Email Dataset", sinkron, 2023 Publication	<1 %
28	digital.library.unt.edu Internet Source	<1 %
29	ejournal.uniks.ac.id Internet Source	<1 %
30	ijream.org Internet Source	<1 %
31	www.dline.info Internet Source	<1 %
32	www.ijfmr.com Internet Source	<1 %
33	"Applications of Artificial Intelligence and Machine Learning", Springer Science and Business Media LLC, 2022 Publication	<1 %
34	dirros.openscience.si Internet Source	<1 %
35	ejournal.ikado.ac.id Internet Source	<1 %
36	ijcrt.org Internet Source	<1 %
37	journal.eng.unila.ac.id Internet Source	<1 %
38	www.mdpi.com Internet Source	<1 %

Exclude quotes

On

Exclude matches

< 7 words

Exclude bibliography

On