

# Deep learning approach to DDoS attack with imbalanced data at the application layer

Rahmad Gunawan<sup>1</sup>, Hadhrami Ab Ghani<sup>2</sup>, Nurulaqilla Khamis<sup>3</sup>, Januar Al Amien<sup>1</sup>, Edi Ismanto<sup>4</sup>

<sup>1</sup>Departement of Informatics Engineering, Faculty of Computer Sciences, Universitas Muhammadiyah Riau, Pekanbaru, Indonesia

<sup>2</sup>Department of Data Science, Faculty of Data Science and Computing, Universiti Malaysia Kelantan, Kota Bharu, Malaysia

<sup>3</sup>Department of Control and Mechatronic, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia

<sup>4</sup>Department of Informatics Education, Faculty of Teacher Training and Education, Universitas Muhammadiyah Riau, Pekanbaru, Indonesia

## Article Info

### Article history:

Received Nov 30, 2022

Revised Mar 08, 2023

Accepted Mar 25, 2023

### Keywords:

ADASYN  
Application layer  
DDoS  
Deep learning  
LDAP  
SMOTE

## ABSTRACT

A distributed denial of service (DDoS) attack is where one or more computers attack or target a server computer, by flooding internet traffic to the server. As a result, the server cannot be accessed by legitimate users. A result of this attack causes enormous losses for a company because it can reduce the level of user trust, and reduce the company's reputation to lose customers due to downtime. One of the services at the application layer that can be accessed by users is a web-based lightweight directory access protocol (LDAP) service that can provide safe and easy services to access directory applications. We used a deep learning approach to detect DDoS attacks on the CICDDoS 2019 dataset on a complex computer network at the application layer to get fast and accurate results for dealing with unbalanced data. Based on the results obtained, it is observed that DDoS attack detection using a deep learning approach on imbalanced data performs better when implemented using synthetic minority oversampling technique (SMOTE) method for binary classes. On the other hand, the proposed deep learning approach performs better for detecting DDoS attacks in multiclass when implemented using the adaptive synthetic (ADASYN) method.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



## Corresponding Author:

Rahmad Gunawan

Departement of Informatics Engineering, Faculty of Computer Sciences

Universitas Muhammadiyah Riau, Jalan Tuanku Tambusai, Kota Pekanbaru, Provinsi Riau, Indonesia

Email: goengoen78@umri.ac.id

## 1. INTRODUCTION

More than 20% of denial-of-service attacks involve a form that utilizes a large number of devices to overload and disrupt a targeted system or network occurs in enterprises worldwide, according to a report published by Kaspersky in 2020 [1], the number of such attacks tripled in the second quarter of 2020 compared to the same quarter in 2019. A distributed denial of service (DDoS) attack is an attack that involves multiple computers that attack one main computer by flooding the internet network traffic so that legitimate users cannot access the main computer because the computer crashes. DDoS attacks are divided into two types of categories: attacks that occur at the network/transport layer by opening half connections on transmission control protocol (TCP), user datagram protocol (UDP), internet control message protocol (ICMP), and domain name system (DNS) and sending large packets or flooding the internet network traffic, and attacks that occur at the application layer only doing very little bandwidth requests, and tends to have a hidden nature. Recently, a new class of DDoS attacks that are referred to as application layer attacks has begun to increase in popularity. This attack exploits vulnerabilities and vulnerabilities in protocols operating at the application layer [2]. These

attacks will consume resources by overloading application servers, their purpose of the attack is more specific such as attacks on lightweight directory access protocol (LDAP), hypertext transfer protocol (HTTP), and DNS applications, by interfering with legitimate user services [3]. As a result, traditional defense systems are unable to cope with DDoS attacks at the application layer that use asymmetric computing between clients and servers, due to requests from protocols and computer network traffic. Attacks that flood network traffic fall into two categories, reflection/amplification-based attacks using DNS queries with fake source IPs by triggering heavy internet traffic resulting in server system crashes. Serengan which is HTTP-based or application-based can be divided into four types; request flooding attacks refer to a cyber attack technique in which a large number of requests are sent to a server or application, consuming its resources and causing it to malfunction or crash on network traffic [3].

In this study, researchers used one type of DDoS attack, namely LDAP. LDAP attack This refers to a DDoS attack related to the exploitation of the LDAP protocol. The attackers flood susceptible LDAP servers with a massive volume of LDAP requests pretending to be real LDAP clients using fake internet protocol (IP) addresses. The LDAP server becomes too busy to make a response for the attacker and becomes unable to respond to the actual LDAP client.

In various problem domains such as genetic engineering [4], [5], text mining [5], [6], picture recognition [7], financial fraud [8], web mining to text categorization [9], and imbalanced data classification has been advocated by researchers [10]. Today, the performance of machine learning (ML), particularly deep learning (DL), can evaluate large amounts of data [11], [12] to differentiate benign from malicious DDoS/DoS assaults rapidly, precisely, and reliably. DL, which is comprised of multiple DNN designs such as recurrent neural network (RNN), convolutional neural network (CNN), and long short-term memory (LSTM) network, offers numerous benefits for classification and prediction issues over standard ML models [8]. RNN employing LSTM units partially resolves the missing gradient problem [13] since LSTM units permit gradients to flow unaffected. However, LSTM networks may still encounter the issue of gradients exploding. This LSTM was created and utilized by several researchers.

In this research, a deep learning model is suggested to resolve data imbalances for identifying and forecasting DDoS/DoS assaults. Using synthetic minority oversampling technique (SMOTE) and adaptive synthetic (ADASYN) approaches, our primary contribution to this research is a novel method for identifying DDoS assaults using imbalanced datasets. In addition, this article provides the most recent assessment of multiclass and binary classes in the DL model for detecting DDoS assaults at the application layer of the new framework utilizing public datasets.

## 2. LITERATURE REVIEW

### 2.1. DDoS

Based on reflection distributed a type of cyber attack that disrupts normal network traffic a technique that allows attacks to be carried out while concealing the IP address of the computer that is attacking by making use of another machine that is legal. This allows the IP address of the computer that is attacking to remain confidential. Because of this, the IP address of the machine that is responsible for carrying out the assault might stay hidden. The data packet's transfer to the reflector server will make use of the source IP address that the attacker has previously provided for use in that transmission. When they initiate their assaults, attackers aim for application layer protocols like TCP and UDP, or a mix of the two at the very least, if not both. Some examples of attacks that fall within the TCP category [14] are those that make use of Microsoft SQL (MSSQL) hand files and simple service discovery protocol (SSDP) hand files.

Attacks such as CharGen, network time protocol (NTP), and trivial file transfer protocol (TFTP) are all examples of those that utilize UDP. It is possible to carry out some attacks, such as those utilizing DNS, LDAP, network basic input output system (NetBIOS), and simple network management protocol (SNMP), using either TCP or UDP to connect with the target to carry out the attack. These assaults are not dependent on or coordinated with one another in any way. The second variant of DDoS attack is known as an exploit-based DDoS attack, and it is an attack that still manages to hide the IP address of the attacker by making use of a real computer that belongs to a third party. This attack is commonly referred to as a DDoS attack that is based on exploiting a vulnerability in a system.

### 2.2. Classification

Binary classification is labeling the output into two groups on a dataset. In the case of our dataset with unbalanced data, our binary classification must have the ability to determine whether the labeling is in the form of an attack or not. Therefore, we group the labeling into two categories: normal and attack [15]. Multiclass classification results in three or more classes in a dataset. Multiclass is caused by data imbalance problems and has several classes.

### 2.3. ADASYN

The over-sampling approach known as sampling ADASYN [16] is a method for achieving a balanced distribution of classes by the random replication of examples in minority classes [17]. Because it replicates the original occurrence in its entirety, over-sampling raises the risk of overfitting. In essence, can adaptively create minority data samples by the data distribution. When there is there has been a rise in the production of synthetic data that pertains to the underrepresented group, it becomes more difficult to research. Although this strategy is unable to eliminate the learning bias caused by an unequal distribution of data, it can adaptively change the limits of judgment in place more emphasis on samples that are more challenging to investigate.

- Step 1: determine the number, and  $G$  of the samples for synthesis as [17].

$$G = (n_b - n_s)\beta, \quad (1)$$

Which is obtained from the difference between the majority sample,  $n_b$  and the minority sample,  $n_s$  where  $\beta \in (0, 1)$ .

- Step 2: for the current sample,  $i$ , the proportion,  $r_i$ , of the current majority class sample,  $k_i$ , from the current (number of) neighbors,  $K_i$ , is written as.

$$r_i = k_i/K_i \quad (2)$$

- Step 3: hence, the number of specimens,  $g$  to be synthesized is computed as:

$$g = Gr_i \quad (3)$$

And the synthesized new sample is expressed as:

$$Z_i = X_i + X_{Z_i} - Xi\lambda \quad (4)$$

Where  $X_i$  and  $X_{Z_i}$  represent the instantaneous minority sample and the random sample respectively with  $X_i, \lambda \in (0, 1)$  [13].

### 2.4. SMOTE

An approach known as the SMOTE involves the insertion of a minority class to generate an additional minority sample for achieving class balance [18] by undersampling the majority class. This method creates a new minority sample to balance the dataset. To do this, it forms a new instance of the minority class by combining neighboring instances. This ensures that the dataset is balanced without being overfit. The SMOTE procedure is detailed in Algorithm 1, which may be found below. The number of minority class samples ( $T$ ), the oversampling rate ( $N\%$ ), and the number of nearest neighbours are the three factors that determine how many synthetic samples ( $S$ ) will be generated for the minority classes ( $k$ ).

When  $N$  is less than 100%, randomization of the samples from the minority classes occurs. Only for the minority classes do we compute the  $k$ -nearest neighbor distances. This is a function of  $N$ , the current sample size from the minority class  $l$  the integral multiples of 100 that are present in  $N(j)$ , and an array of random integers ( $nn$  array).  $Z$  is an array of the original samples that came from the minority class,  $r$  is the number of synthetic samples that were created, and  $V$  is an array of the synthetic samples [18]. Synthetic minority over-sampling approach is the most effective oversampling technique due to its wide handling and practice, including applications in gender analysis, bioengineering [19], medical examination [20], and fraud identification. And we believe that the use of The investigation of distributed SMOTE is a current and lively research field [21].

## 3. METHOD

### 3.1. Dataset

The Canadian Institute for Cybersecurity is the origin of the dataset that will be utilized in the CICDDoS 2019 competition [22]. The dataset includes both fake and real-time DDoS assaults; the attacks are modeled after genuine data from the real world planar capacitor (PCAP). Also included are the findings of a network traffic analysis performed using CICFlowMeter-V3, complete with labeled streams that are organized according to timestamps, source and destination IP addresses, source and destination ports, protocols, and assaults (CSV file).

One variety of DDoS assaults was utilized in this investigation by the researchers. This attack was carried out at the application layer using LDAP, which is dependent on the TCP protocol. This massive document contains 80 columns and 2113234 rows. The dataset includes legitimate and the most recent

examples of frequently distributed denial of service attacks, which are analogous to real-world data PCAP. The raw data, which includes the network traffic PCAP and event logs (both Windows event logs and Ubuntu event logs) for each system, each of which has been recorded as a CSV file. The dataset has been arranged in a day-by-day format. Every day, we captured the raw data, which included the machine's network traffic PCAP and event logs (both Windows and Ubuntu event Logs).

### 3.2. Data preparation

At this stage hardware and software are needed to analyze the dataset. Meanwhile, we analyzed the dataset using virtual machine hardware with specifications Intel® Xeon® Gold 6134 CPU 3.20 Ghz, virtual processor 22, speed 3.19 Ghz, 32 Gb RAM with @Jupyter 6.4.12 software. Figure 1 shows the data architecture process flow, with an explanation of the stages as:

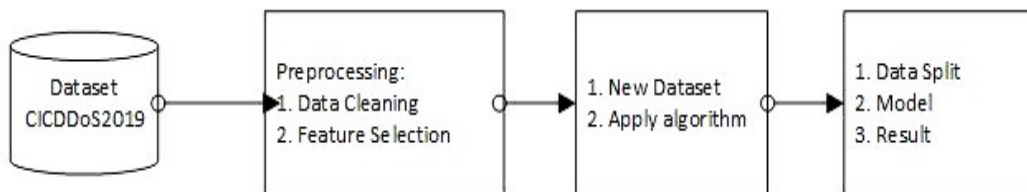


Figure 1. Process data architecture

- a) Input dataset CICDDoS2019: the first step is to obtain the required dataset on which the proposed model will be trained, tested and validated.
- b) Data cleaning: data cleansing is considered to be a crucial aspect step of data pre-processing. Data cleaning is the process of cleaning data by removing features that have identical values and also removing noise from irrelevant data [23], [24]. This data-cleaning process greatly affects performance because the data to be handled will reduce noise and complexity.
- c) Feature selection: the process of selecting features is among the most important phases contributing to deep learning model results and performance. Selecting and reducing features from the data set can increase training and testing speed, classifier accuracy, and computational modeling costs [25]. The goal is to reduce the dimensionality of the data and improve the quality of predictive models.
- d) New dataset: after preprocessing, a new dataset is obtained from the results of data cleaning and feature selection which will be classified at a later stage. A new dataset is a data set that can be new data that has never been used before, a subset of the original dataset, or a dataset that has undergone transformation or preprocessing. The next step is to choose the right model or algorithm for the data analysis you want to do.
- e) Apply the algorithm: we analyzed the dataset and found that the data was imbalanced, there are 3 labels in the "LDAP.csv" file, namely Benign and LDAP, and NetBIOS. We label [0,1,1] for multiclass and [0,1] for binary. We categorize binaries into two types: benign and LDAP. The author conducted testing of binary classification and multiclass classification. In this study, the author will compare the two tests, namely multiclass classification, and binary classification. Submission of ADASYN algorithm and SMOTE algorithm to handle imbalanced data.
- f) Data split: after getting the new dataset, we split the dataset for testing, namely training data and test data, with a distribution of 80% for training data and 20% for test data. The purpose of data split is to divide the dataset into different subsets for use in different stages of the data analysis process. Data splits are generally performed before training a machine learning model or evaluating model performance.
- g) Deep neural network: it is the framework that should be used to get knowledge from the dataset that we entered. In most cases, it is composed of the following three primary layers: the data that are received from the dataset are the responsibility of the input layer. In most cases, one node corresponds to a single dimension or the total number of characteristics that are included in the dataset. The data from the input layer is sent on to the hidden layer, which is the layer that is concealed. The information received from the previous layer is transmitted to the output layer, which is the last layer in the chain.
- h) Evaluation: using the following indications taken from the standard matrix, we can evaluate how effective the deep learning model will be in detecting DDoS attacks: 1) correctness refers to the overall accuracy of the model; 2) the term "precision" refers to the likelihood that the model would correctly identify the assault; 3) the chance that the model can identify attacks out of the total number of assaults is referred to as the recall; and 4) F-measure, also known as F1-Score, is a harmonic mean that combines recall and accuracy. The bottom equation contains the formulae that are used to calculate the values [26].

$$Accuracy = (TP + TN)/(TP + TN + FP + FN)$$

$$Precision = TP/(TP + FP)$$

$$Recall = TP/(TP + FN)$$

$$F1 - Score = 2((Precision \times Recall)/(Precision + Recall))$$

Where  $TP$ ,  $TN$ ,  $FP$ , and  $FN$  represent true positives, true negatives, false positives, and false negatives, respectively.

- $TP$  = true positive
- $TN$  = true negative
- $FP$  = false positive
- $FN$  = false negative

#### 4. RESULTS AND DISCUSSION

During the course of this experiment, the performance of the suggested network was assessed by the utilization of normal, LDAP, and NetBIOS probes, in that specific sequence, for the binary and multiclassification trials, respectively. The findings of the experiment revealed that the majority of the samples could be appropriately categorized in their respective categories. Because of this, the samples were able to be placed on the diagonal, which demonstrates that the classification performance was enhanced. The effectiveness of the suggested model, on the other hand, was significantly diminished when it was subjected to multiclassification trials as opposed to binary classification experiments, as can be seen by comparing the two figures. This was the case when the model was tested with both types of classifications. The fact that the suggested model showed poor performance in the studies that incorporated multiclassification is evidence that this is the case.

By comparing the ADASYN algorithm and the SMOTE algorithm. Figure 2 and Figure 3 show the results of the binary class comparison experiment between ADASYN and SMOTE in the confusion matrix binary model. Experiments show how to correctly classify numbers diagonally with good performance for SMOTE. Table 1 describes the results of the comparison between ADASYN and SMOTE, where the F1 value for benign is 0.9983 and for attack is 0.9983, while SMOTE shows F1 benign 0.9997 and attack is 0.9997. This shows that the SMOTE technique has better performance than ADASYN for the binary class.

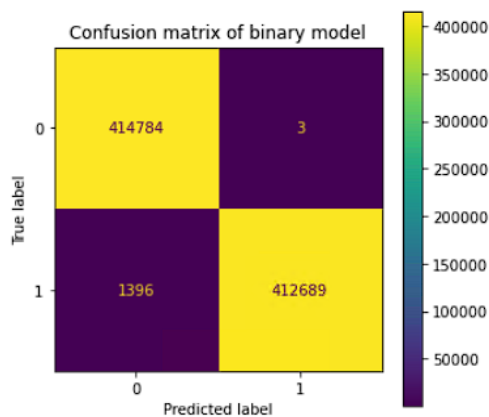


Figure 2. ADASYN binary

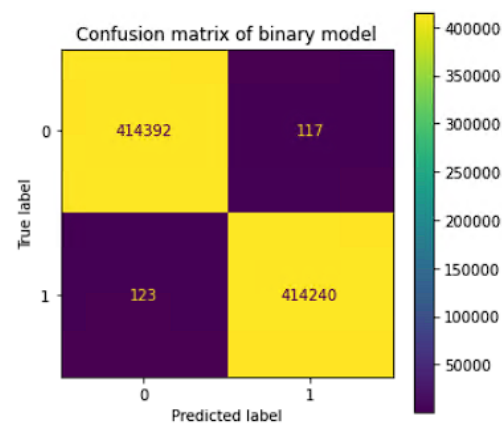


Figure 3. SMOTE binary

Table 1. Binary data comparison

| Type/c | ADASYN     |            | SMOTE      |            |
|--------|------------|------------|------------|------------|
|        | Benign (0) | Attack (1) | Benign (0) | Attack (1) |
| Acc    | 0.9983     | 0.9983     | 0.9997     | 0.9997     |
| Pre    | 0.9966     | 1.0        | 0.9997     | 0.9997     |
| Rec    | 1.0        | 0.9966     | 0.9997     | 0.9997     |
| F1     | 0.9983     | 0.9983     | 0.9997     | 0.9997     |

Figure 4 and Figure 5 show the results of a multiclass comparison experiment between ADASYN and SMOTE in the model multiclassification confusion matrix. The experiment shows correctly multiclassification diagonally with good performance for ADASYN. Table 2 describes the results of the comparison of ADASYN and SMOTE, where the F1 value for benign is 1.0, LDAP is 0.9999 and NetBIOS is 0.9999, on SMOTE it shows F1 benign 0.9999, LDAP is 0.9999 and netbois is 0.9998, this shows that the ADASYN technique has better performance than SMOTE for multiclassification.

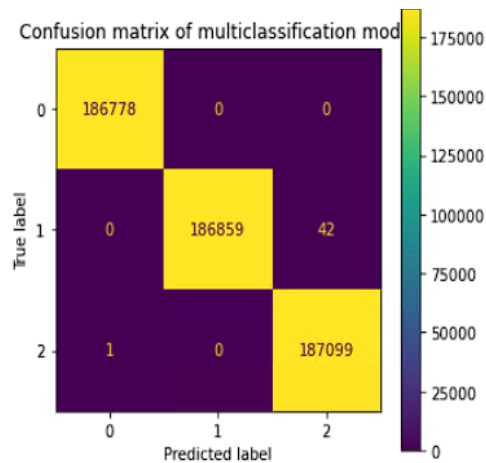


Figure 4. ADASYN

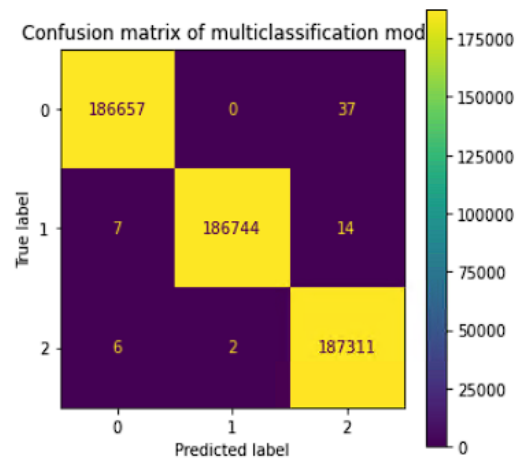


Figure 5. SMOTE

Table 2. Multiclass data comparison

| Type/classes | ADASYN     |          |             | SMOTE      |          |             |
|--------------|------------|----------|-------------|------------|----------|-------------|
|              | Benign (0) | LDAP (1) | NetBIOS (2) | Benign (0) | LDAP (1) | NetBIOS (2) |
| Acc          | 1.0        | 0.9999   | 0.9999      | 0.9999     | 1.0      | 0.9999      |
| Pre          | 1.0        | 1.0      | 0.9998      | 0.9999     | 1.0      | 0.9997      |
| Rec          | 1.0        | 0.9998   | 1.0         | 0.9998     | 0.9999   | 1.0         |
| F1           | 1.0        | 0.9999   | 0.9999      | 0.9999     | 0.9999   | 0.9998      |

## 5. CONCLUSION

From the stages of work that have been carried out in the research above, the results obtained can be concluded. The results of the performance comparison on the two algorithms, namely the ADASYN and SMOTE algorithms, show high accuracy performance in overcoming the problem of data imbalance, both in the binary category, namely benign and abnormal, and for multiclass into three classes, namely benign, LDAP and Netbois. The experimental results that we conducted show that SMOTE is better than ADASYN for binaries and ADASYN is better than SMOTE for multiclass in overcoming the problem of unbalanced data.




## REFERENCES

- [1] www.kaspersky.com, "The Kaspersky Q2 2020 DDoS attacks report," 2020. [Online]. Available: [https://www.kaspersky.com/about/press-releases/2020\\_no-summer-vacation-ddos-attacks-tripled-year-on-year-in-q2-2020](https://www.kaspersky.com/about/press-releases/2020_no-summer-vacation-ddos-attacks-tripled-year-on-year-in-q2-2020)
- [2] N. Tripathi and N. Hubballi, "Application layer denial-of-service attacks and defense mechanisms: A survey," *ACM Computing Surveys*, vol. 54, no. 4, pp. 1-33, 2021, doi: 10.1145/3448291.
- [3] I. Sreeram and V. P. K. Vuppala, "HTTP flood attack detection in application layer using machine learning metrics and bio inspired bat algorithm," *Applied Computing and Informatics*, vol. 15, no. 1, pp. 59–66, 2019, doi: 10.1016/j.aci.2017.10.003.
- [4] Y. Liu, Z. Yu, C. Chen, Y. Han, and B. Yu, "Prediction of protein crotonylation sites through LightGBM classifier based on SMOTE and elastic net," *Analytical Biochemistry*, vol. 609, 2020, doi: 10.1016/j.ab.2020.113903.
- [5] Y. Li, H. Guo, Q. Zhang, M. Gu, and J. Yang, "Imbalanced text sentiment classification using universal and domain-specific knowledge," *Knowledge-Based Systems*, vol. 160, pp. 1–15, 2018, doi: 10.1016/j.knosys.2018.06.019.
- [6] R. Panigrahi and S. Borah, "Dual-stage intrusion detection for class imbalance scenarios," *Computer Fraud & Security*, vol. 2019, no. 12, pp. 12–19, 2021, doi: 10.1016/S1361-3723(19)30128-9.
- [7] L. Wang and C. Wu, "Dynamic imbalanced business credit evaluation based on Learn++ with sliding time window and weight sampling and FCM with multiple kernels," *Information Sciences*, vol. 520, pp. 305–323, 2020, doi: 10.1016/j.ins.2020.02.011.
- [8] M. E. El-Telbany, "Prediction of the Electrical Load for Egyptian Energy Management Systems: Deep Learning Approach," *The International Conference on Artificial Intelligence and Computer Vision (AICV2020)*, 2020, vol. 1153, doi: 10.1007/978-3-030-44289-7\_23.
- [9] G. Wang, J. Chen, and L. T. Yang, "Security, Privacy, and Anonymity in Computation, Communication, and Storage," *11th International Conference and Satellite Workshops, SpaCCS 2018*, 2018, doi: 10.1007/978-3-319-72395-2.




- [10] Y. Sun, A. K. C. Wong, and M. S. Kamel, "Classification of imbalanced data: A review," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 23, no. 4, pp. 687–719, 2009, doi: 10.1142/S0218001409007326.
- [11] K. Kambatla, G. Kollias, V. Kumar, and A. Grama, "Trends in big data analytics," *Journal of Parallel and Distributed Computing*, vol. 74, no. 7, pp. 2561–2573, 2014, doi: 10.1016/j.jpdc.2014.01.003.
- [12] Sowmya R. and Suneetha K. R., "Data Mining with Big Data," *2017 11th International Conference on Intelligent Systems and Control (ISCO)*, 2017, pp. 246-250, doi: 10.1109/ISCO.2017.7855990.
- [13] T. Khempetch and P. Wuttidittachotti, "Ddos attack detection using deep learning," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 2, pp. 382–388, 2021, doi: 10.11591/ijai.v10.i2.pp382-388.
- [14] I. Sharafaldin, A. H. Lashkari, S. Hakak, and A. A. Ghorbani, "Developing Realistic Distributed Denial of Service (DDoS) Attack Dataset and Taxonomy," *2019 International Carnahan Conference on Security Technology (ICCST)*, 2019, pp. 1-8, doi: 10.1109/CCST.2019.8888419.
- [15] F. E. Laghrissi, S. Douzi, K. Douzi, and B. Hssina, "Intrusion detection systems using long short-term memory (LSTM)," *Journal of Big Data*, vol. 8, no. 65, 2021, doi: 10.1186/s40537-021-00448-4.
- [16] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008, pp. 1322-1328, doi: 10.1109/IJCNN.2008.4633969.
- [17] Y. Fu, Y. Du, Z. Cao, Q. Li, and W. Xiang, "A Deep Learning Model for Network Intrusion Detection with Imbalanced Data," *Electronics*, vol. 11, no. 6, 2022, doi: 10.3390/electronics11060898.
- [18] S. I. Popoola, B. Adebisi, R. Ande, M. Hammoudeh, K. Anoh, and A. A. Atayero, "SMOTE-drrn: A deep learning algorithm for botnet detection in the internet-of-things networks," *Sensors*, vol. 21, no. 9, 2021, doi: 10.3390/s21092985.
- [19] C. Liu, J. Wu, L. Mirador, Y. Song, and W. Hou, "Classifying DNA Methylation Imbalance Data in Cancer Risk Prediction Using SMOTE and Tomek Link Methods," *International Conference of Pioneering Computer Scientists, Engineers and Educators*, 2018, vol. 902, pp. 1-9, doi: 10.1007/978-981-13-2206-8\_1.
- [20] M. Nakamura, Y. Kajiwara, A. Otsuka, and H. Kimura, "LVQ-SMOTE - Learning Vector Quantization based Synthetic Minority Over-sampling Technique for biomedical data," *BioData Mining*, vol. 6, no. 16, 2013, doi: 10.1186/1756-0381-6-16.
- [21] S. Hooda and S. Mann, "Distributed synthetic minority oversampling technique," *International Journal of Computational Intelligence Systems*, vol. 12, no. 2, pp. 929–936, 2019, doi: 10.2991/ijcis.d.190719.001.
- [22] University of New Brunswick (UNB), *DDoS Evaluation Dataset (CIC-DDoS2019)*, Canadian Institute for Cybersecurity, 2019. [Online]. Available: <https://www.unb.ca/cic/datasets/ddos-2019.html>
- [23] H. Xiong, G. Pandey, M. Steinbach, and V. Kumar, "Enhancing data analysis with noise removal," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 3, pp. 304-319, 2006, doi: 10.1109/TKDE.2006.46.
- [24] M. Jupri and R. Sarno, "Data mining, fuzzy AHP and TOPSIS for optimizing taxpayer supervision," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 18, no. 1, pp. 75–87, 2020, doi: 10.11591/ijeecs.v18.i1.pp75-87.
- [25] O. Thorat, N. Parekh, and R. Mangrulkar, "TaxoDaCML: Taxonomy based Divide and Conquer using machine learning approach for DDoS attack classification," *International Journal of Information Management Data Insights*, vol. 1, no. 2, 2021, doi: 10.1016/j.jjime.2021.100048.
- [26] M. S. Rana, C. Gudla, and A. H. Sung, "Evaluating machine learning models for android malware detection: A comparison study," *ICNCC '18: Proceedings of the 2018 VII International Conference on Network, Communication and Computing*, 2018, pp. 17–21, doi: 10.1145/3301326.3301390.

## BIOGRAPHIES OF AUTHORS






**Rahmad Gunawan**    graduated with a bachelor's degree at Gunadarma University with a major in Information Management, a master's degree with Gunadarma University majoring in Electrical Telecommunication. And now works as a lecturer at the Faculty of Computer Science, University of Muhammadiyah Riau. With research interests in the field of Machine learning algorithms and AI. He can be contacted at email: goengoen78@umri.ac.id.






**Hadhrami Ab Ghani**    received his bachelor degree in electronics engineering from Multimedia University Malaysia (MMU) in 2002. In 2004, he completed his masters degree in Telecommunication Engineering at The University of Melbourne. He then pursued his Ph.D. at Imperial College London in intelligent network systems and completed his Ph.D. in 2011. He can be contacted at email: hadhrami.ag@umk.edu.my.






**Nurulaqilla Khamis**    received her bachelor degree in electrical and electronics engineering from Universiti Tenaga Nasional in 2012. In 2015, she completed her masters degree in Artificial Intelligence at Universiti Teknologi Malaysia. She then pursued her Ph.D at Universiti Teknologi Malaysia in Artificial Intelligence and completed her Ph.D in 2020. Her current research work focuses on Machine Learning, Deep Learning and Swarm Intelligence Optimization. She can be contacted at email: nurulaqilla@utm.my.



**Januar Al Amien**    completed education bachelor's degree in the Informatics Engineering Department, STMIK-AMIK Riau. And master's degree in Master of Information Technology at Putra Indonesia University Padang. Now working as a lecturer in the Department of Computer Science, University Muhammadiyah of Riau. With research interests in the field of Machine learning algorithms and AI. He can be contacted at email: januaralamien@umri.ac.id.



**Edi Ismanto**    completed education bachelor's degree in the Informatics Engineering Department, State Islamic University of Sultan Syarif Kasim Riau. And master's degree in Master of Computer Science at Putra Indonesia University Padang. Now working as a lecturer in the Department of Informatics, University Muhammadiyah of Riau. With research interests in the field of Machine learning algorithms and AI. He can be contacted at email: edi.ismanto@umri.ac.id.