DETEKSI SERANGAN DISTRIBUTED DENIAL OF SERVICE (DDoS) PADA APACHE SPARK MENGGUNAKAN METODE K-MEANS CLUSTERING

TUGAS AKHIR

OLEH :

MOCHAMMAD RAFII NANDA WICAKSANA
09011281823053

JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2023

LEMBAR PENGESAHAN

DETEKSI SERANGAN DISTRIBUTED DENIAL OF SERVICE (DDOS) PADA APACHE SPARK MENGGUNAKAN METODE K-MEANS CLUSTERING

TUGAS AKHIR

Program Studi Sistem Komputer
Jenjang S1

Oleh :

MOCHAMMAD RAFII NANDA WICAKSANA
09011281823053

Indralaya, 8 Mei 2023

Mengetahui

Ketua Jurusan Sistem Komputer	Pembimbing Tugas Akhir,
Dr. Ir. H. Sukemi, M.T.	**Ahmad Heryanto, S.Kom., M.T.**
NIP. 196612032006041001	**NIP. 198701222015041002**

HALAMAN PERSETUJUAN

Telah diuji dan lulus pada :

Hari : Jum'at

Tanggal : 14 April 2023

Tim Penguji :

1. **Ketua** : **Sarmayanta Sembiring, M.T.**
2. **Sekretaris** : **Nurul Afifah, M.Kom.**
3. **Penguji** : **Ahmad Faii Oklilas, M.T.**
4. **Pembimbing I** : **Ahmad Heryanto, S.Kom., M.T.**

Mengetahui, 9/5/23
Ketua Jurusan Sistem Komputer

Dr.Ir.H. Sukemi, M.T.
NIP. 196612032006041001

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Mochammad Rafii Nanda Wicaksana

NIM : 09011281823053

Judul : Deteksi Serangan Distributed Denial Of Service (DDoS) Pada Apache Spark Menggunakan Metode K-Means Clustering

Hasil Pengecekan Software *iThenticate/Turnitin* : 18%

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan atau plagiat. Apabila ditemukan unsru penjiplakan atau plagiat dalam laporan tugas akhir ini, maka saya bersedia menerima sanksi akademik dari universitas sriwijaya

Demikian, pernyatan ini saya buat dalam keadaan sadar dan tanpa paksaan dari siapapun.

Indralaya, Mei 2023

Mochammad Rafii Nanda W.

NIM.09011281823053

KATA PENGANTAR

Assalamualaikum Warahmatullahi Wabarakatuh.

Puji dan syukur atas kehadiran Allah Subhanahu Wa ta'ala yang telah memberikan rahmat dan hidayah-Nya lah sehingga penulis dapat menyelesaikan penyusunan Tugas Akhir ini yang berjudul "**Deteksi Serangan Distributed Denial of Service (DDoS) Pada Apache Spark Menggunakan Metode K-Means Clustering**".

Pada kesempatan ini penulis mengucapkan terima kasih kepada pihak yang telah memberikan bantuan, dorongan, motivasi, semangat dan bimbingan dalam menyelesaikan penyusunan Tugas Akhir ini. Penulis mengucapkan terima kasih kepada :

1. Allah Subhanahu Wa ta'ala yang memberikan rahmat dan hidayah-Nya serta nikmat yang tak terhitung.
2. Kedua orangtua saya dan saudara yang telah membantu.
3. Bapak Dr. Jaidan Jauhari, S.Pd., M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya
4. Bapak Dr. Ir. H. Sukemi, M.T. selaku Ketua Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Bapak Ahmad Heryanto, S.Kom., M.T. selaku Pembimbing Tugas Akhir yang telah berkenan meluangkan waktunya untuk membimbing, memberikan saran dan motivasi serta bimbingan terbaik untuk penulisan dalam menyelesaikan Tugas Akhir ini.
6. Bapak Ahmad Fali Oklilas, M.T. selaku Dosen Pembimbing Akademik saya.
7. Dwi Lingga Hanayuda, Jepi Sujana, Bima Gusti Syauqi, Daffa Bima Perdana dan Agung Al hafizin selaku rekan yang membantu Menyusun dan menyelesaikan penulisan Tugas Akhir ini.
8. Prazna Paramitha Avi, Rani Octaviani dan Muhammad Imam Rafi yang telah membantu dalam penulisan Tugas Akhir ini.

9. Mbak Renny selaku admin Jurusan Sistem Komputer yang telah membantu mengurus seluruh berkas
10. Teman-teman Sistem Komputer Angkatan 2018 Indralaya.

Dalam penyusunan Tugas Akhir ini penulis menyadari sepenuhnya masih jauh dari kata sempurna, oleh karena itu penulis mengharapkan saran dan kritik dari semua pihak yang berkenan agar menjadi bahan evaluasi yang lebih baik lagi.

Akhir kata saya harap semoga Laporan Tugas Akhir ini dapat bermanfaat serta dapat menambah pengetahuan dan wawasan bagi yang membutuhkannya.

Wassalamualakum Warahmatullahi Wabarakatuh

Indralaya, Mei 2023

Penulis

Mochammad Rafii Nanda W.

09011281823053

DETEKSI SERANGAN DISTRIBUTED DENIAL OF SERVICE (DDoS) PADA APACHE SPARK MENGGUNAKAN METODE K-MEANS CLUSTERING

Mochammad Rafii Nanda Wicaksana (09011281823053)
Department of Computer System, Faculty of Computer Science, University of Sriwijaya
Palembang, Indonesia
Email : mochammadrafii82@gmail.com

ABSTRAK

Distributed Denial-of-Service (DDoS) merupakan sekumpulan dari serangan denial-of-service yang dilakukan dengan menjalankan perintah dari komputer master ke sejumlah botnet yang mana merupakan host yang telah terinfeksi untuk menyerang target tertentu. Dikarenakan hal tersebut, deteksi serangan DDoS menjadi yang pertama dan yang paling penting untuk melawan serangan DDoS. Dasar untuk melakukan pendekatan deteksi yaitu menggunakan machine learning. K-means clustering adalah algoritma analisis clustering yang paling sederhana dan paling terkenal dalam memecahkan masalah clustering. Algoritma ini dikenal efisien untuk dataset yang besar. Pada jurnal ini mengusulkan deteksi serangan DDoS dengan menggunakan salah satu metode unsupervised learning yaitu K-means clustering pada apache spark. Penelitian ini menggunakan dataset CIC-DDoS2019 dari University of New Brunswick (UNB) untuk melatih dan melakukan percobaan pada sistem deteksi yang digunakan. Model ini menghasilkan hasil evaluasi terbaik dengan nilai recall, presisi, spesifisitas, akurasi dan F1 sebesar 99.99%, 99.99%, 88.24%, 99.98%, dan 99.99%.

Kata Kunci : DDoS, K-means, Apache Spark, Keamanan Jaringan, Machine Learning.

DISTRIBUTED DENIAL OF SERVICE (DDoS) ATTACK DETECTION ON APACHE SPARK USING K-MEANS CLUSTERING METHOD

Mochammad Rafii Nanda Wicaksana (09011281823053)
Department of Computer System, Faculty of Computer Science, University of Sriwijaya
Palembang, Indonesia
Email : mochammadrafii82@gmail.com

ABSTRACT

Distributed Denial-of-Service (DDoS) is a collection of denial-of-service attacks that are carried out by executing commands from the master computer to a number of botnets which are infected hosts to attack certain targets. Because of this, DDoS attack detection is the first and most important way to counter DDoS attacks. The basis for carrying out the detection approach is using machine learning. K-means clustering is the simplest and most well-known clustering analysis algorithm in solving clustering problems. This algorithm is known to be efficient for large datasets. This paper proposes detecting DDoS attacks using an unsupervised learning method, namely K-means clustering on Apache Spark. This study used the CIC-DDoS2019 dataset from the University of New Brunswick (UNB) to train and perform experiments on the detection system used. This model produces the best evaluation results with the recall of 99,99%, precision of 99,99%, specificity of 88.24%, the accuracy of 99.98%, and F1 score of 99.99%.

Keyword : DDoS, K-means, Apache Spark, Network Security, Machine Learning.

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Distributed Denial-of-Service (DDoS) merupakan sekumpulan dari serangan denial-of-service yang dilakukan dengan menjalankan perintah dari komputer master ke sejumlah botnet yang mana merupakan host yang telah terinfeksi untuk menyerang target tertentu[1]. Serangan DDoS ini pada umumnya akan mengirimkan paket dalam jumlah yang besar yang mana hal ini bertujuan membuat target menjadi kelebihan beban sehingga tidak dapat diakses. Tujuan utama dari serangan DDoS yaitu membuat layanan host korban menjadi mati[2] dengan memenuhi lalu lintas jaringan dengan traffic yang tinggi sehingga akan menolak akses ke layanan untuk pengguna normal lainnya[3].

Terdapat banyak metode dan alat untuk melakukan serangan DDoS. Secara tradisional, serangan DDoS seperti ICMP flooding, SYN flooding dan UDP flooding dilakukan pada *layer network* dan *layer transport*[4]. Akibat yang ditimbulkan dari serangan DDoS menyebabkan kerugian yang besar sehingga menjadi suatu masalah serius dan dikatakan sebagai pelanggaran terhadap kebijakan penggunaan internet. Serangan ini biasanya akan menyerang aktifitas yang terjadi pada internet seperti pada website, bisnis, dan kompetisi game. Pada tahun 2015, serangan DDoS menyerang *Github*, yang merupakan penyedia layanan berbasis cloud yang berdampak pada layanan selama seminggu dan pada tahun 2016 terjadi hal yang sama pada perusahaan *Dyn*, yang menyediakan layanan server DNS sehingga hal ini berdampak pada *Netflix* dan *Spotify*[5].

Dikarenakan hal tersebut, deteksi serangan DDoS menjadi yang pertama dan yang paling penting untuk melawan serangan DDoS. Terdapat dua teknik untuk mendeteksi DDoS yaitu dengan deteksi penyalahgunaan dan deteksi anomali[6]. Teknik deteksi penyalahgunaan yaitu mendeteksi serangan dengan membandingkan aktivitas jaringan tujuan saat ini dengan karakteristik serangan yang telah diketahui. Tetapi cara ini sulit untuk mendeteksi serangan baru. Oleh karena itu, teknik deteksi anomali digunakan untuk mendeteksi serangan yang

belum diketahui dengan membandingkan aktivitas jaringan tujuan saat ini dengan aktivitas normal yang telah ditetapkan. Dasar untuk melakukan pendekatan deteksi yaitu menggunakan *machine learning* untuk membuat model aktivitas normal dan membandingkannya dengan aktivitas baru.

K-means clustering adalah algoritma analisis clustering yang mengelompokkan objek berdasarkan nilai fiturnya ke dalam K cluster yang terpisah-pisah. Objek yang diklasifikasikan ke dalam cluster yang sama memiliki nilai fitur yang serupa. K adalah nilai positif yang menentukan jumlah cluster dan harus diberikan terlebih dahulu[7]. Proses k-means clustering menghasilkan cluster *centroids* untuk lalu lintas normal dan anomali yang dapat digunakan untuk mendeteksi anomali. Menurut [8] K-means merupakan algoritma yang paling sederhana dan paling terkenal dalam memecahkan masalah clustering. Algoritma ini dikenal efisien untuk dataset yang besar.

Pada [1], [2] dan [4] telah melakukan penelitian menggunakan metode K-means yang mana pada penelitian ini menunjukkan hasil yang baik dan berhasil untuk mendeteksi anomali dan DDoS. Pada [7] mendeteksi traffic anomali menggunakan k-means. Pada penelitian tersebut membagi menjadi 2 cluster yaitu normal dan anomali yang mana anomali terdeteki delapan kali lebih tinggi dari normal. Dengan menambah algoritma clustering akan meningkatkan kualitas deteksi.

Pada [9] melakukan penelitian anlisa kinerja Apache Spark menggunakan k-means. Hasil yang didapat dalam penelitian tersebut meunjukkan bahwa Spark jauh lebih cepat di mana setiap ukuran dari dataset menghasilkan penurunan waktu tiga kali lipat dibandingkan Map Reduce yang berarti bahwa Spark dapat digunakan dalam pemrosesan big data.

Dari uraian tersebut dapat menjadi landasan bagi penulis untuk meningkatkan performa model yang telah dibuat dan mengusulkan deteksi serangan DDoS dengan menggunakan salah satu metode *unsupervised learning* yaitu K-means clustering pada apache spark.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang dijelaskan, maka perumusan masalah yang akan dibahas adalah sebagai berikut :

1. Bagaimana serangan dari DDoS dapat dikenali dan dideteksi?
2. Bagaimana model K-means dapat mendeteksi dan menghitung akurasi deteksi serangan DDoS?
3. Apa Pengaruh dari Spark dalam melakukan pemrosesan data?

1.3 Tujuan

Adapun tujuan dari penulisan Tugas Akhir ini antara lain :

1) Membangun simulasi program untuk mengenali dan mendeteksi pola serangan dari DDoS pada apache spark.
2) Mengklasifikasi serangan DDoS menggunakan metode K-means.
3) Menghitung akurasi, presisi, *recall*, dan F1 Score dari deteksi serangan DDoS.
4) Mengetahui pengaruh spark dalam melakukan pemrosesan data.

1.4 Manfaat

Manfaat dari penulisan Tugas Akhir ini, antara lain :

1. Dapat membantu dalam mendeteksi serangan DDoS pada lalu lintas jaringan.
2. Dapat menerangkan proses terjadinya penyerangan yang dilakukan oleh pelaku pada sistem korban.
3. Dapat membantu mengurangi waktu dan biaya yang diperlukan dalam mendeteksi serangan DDoS dengan menggunakan Apache Spark.

1.5 Batasan Masalah

Batasan Masalah pada Tugas Akhir ini, antara lain :

1. Penelitian ini menggunakan dataset CIC-DDoS2019 dari *University of New Brunswick* (UNB).

2. Metode yang digunakan untuk klasifikasi serangan DDoS menggunakan K-means Clustering.
3. Indikator peneliti yaitu mengukur nilai akurasi, presisi, *recall*, dan F1 Score sebagai hasil dari penelitian.

1.6 Sistematika Penulisan

Sistematika yang akan digunakan dalam penulisan tugas akhir adalah sebagai berikut :

BAB I PENDAHULUAN

Bab pertama akan memaparkan sistematis mengenai latar belakang, tujuan penelitian, rumusan masalah, serta bentuk sistematika penelitian.

BAB II TINJAUAN PUSTAKA

Bab kedua akan menjelaskan teori-teori dasar yang akan menjadi landasan dari penelitian ini.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan proses dan rangkaian kegiatan dalam penelitian.

BAB IV HASIL DAN ANALISIS

Bab ini akan memaparkan hasil pengujian yang diperoleh dan menjelaskan analisa terhadap hasil penelitian sementara yang telah dilakukan.

BAB V KESIMPULAN

Bab ini akan memaparkan kesimpulan sementara dari hasil yang telah didapat dari penelitian.

Daftar Pustaka

[1] M. I. W. Pramana, Y. Purwanto, and F. Y. Suratman, "DDoS detection using modified K-means clustering with chain initialization over landmark window," *ICCEREC 2015 - Int. Conf. Control. Electron. Renew. Energy Commun.*, pp. 7–11, 2015, doi: 10.1109/ICCEREC.2015.7337056.

[2] T.-Y. Hsieh, Chang-Jung. Chan, "Detection DDoS Attacks Based on Neural-Network Using Apache Spark," pp. 1–4, 2016.

[3] A. Alsirhani, S. Sampalli, and P. Bodorik, "DDoS Detection System: Using a Set of Classification Algorithms Controlled by Fuzzy Logic System in Apache Spark," *IEEE Trans. Netw. Serv. Manag.*, vol. PP, no. c, p. 1, 2019, doi: 10.1109/TNSM.2019.2929425.

[4] C. She, W. Wen, K. Zheng, and Y. Lyu, "Application-Layer DDoS Detection by K-means Algorithm," vol. 50, no. Iceeecs, pp. 75–78, 2016, doi: 10.2991/iceeecs-16.2016.16.

[5] A. E. Cil, K. Yildiz, and A. Buldu, "Detection of DDoS attacks with feed forward based deep neural network model," *Expert Syst. Appl.*, vol. 169, no. December 2020, p. 114520, 2021, doi: 10.1016/j.eswa.2020.114520.

[6] Y. Gu, K. Li, Z. Guo, and Y. Wang, "Semi-supervised k-means ddos detection method using hybrid feature selection algorithm," *IEEE Access*, vol. 7, pp. 64351–64365, 2019, doi: 10.1109/ACCESS.2019.2917532.

[7] G. Münz, S. Li, and G. Carle, "Traffic Anomaly Detection Using K-Means Clustering," *GI/ITG Work. MMBnet*, no. January, pp. 13--14, 2007.

[8] A. Kumar, Y. S. Ingle, P. Abhijit, and P. Dhule, "Canopy Clustering : A Review on Pre-Clustering Approach to K-Means Clustering," *Int. J. Innov. Adv. Comput. Sci.*, vol. 3, no. 5, pp. 22–29, 2014.

[9] S. Gopalani and R. Arora, "Comparing Apache Spark and Map Reduce with Performance Analysis using K-Means," *Int. J. Comput. Appl.*, vol.

113, no. 1, pp. 8–11, 2015, doi: 10.5120/19788-0531.

[10] M. Haggag, M. M. Tantawy, and M. M. S. El-Soudani, “Implementing a deep learning model for intrusion detection on apache spark platform,” *IEEE Access*, vol. 8, no. 1, pp. 163660–163672, 2020, doi: 10.1109/ACCESS.2020.3019931.

[11] S. V. Sivareddy and S. Saravanan, “Performance evaluation of classification algorithms in the Design of apache spark based intrusion detection system,” *Proc. 5th Int. Conf. Commun. Electron. Syst. ICCES 2020*, no. Icces 2020, pp. 443–447, 2020, doi: 10.1109/ICCES48766.2020.09138066.

[12] N. V. Patil, C. Rama Krishna, and K. Kumar, “S-DDoS: Apache spark based real-time DDoS detection system,” *J. Intell. Fuzzy Syst.*, vol. 38, no. 5, pp. 6527–6535, 2020, doi: 10.3233/JIFS-179733.

[13] S. Gumaste, D. G. Narayan, S. Shinde, and K. Amit, “Detection of DDoS Attacks in OpenStack-based Private Cloud Using Apache Spark,” 2020.

[14] S. Pande, A. Khamparia, D. Gupta, and D. N. H. Thanh, *DDOS Detection Using Machine*, no. October. Springer Singapore, 2020.

[15] C. P. Chang, W. C. Hsu, and I. E. Liao, “Anomaly detection for industrial control systems using k-means and convolutional autoencoder,” *2019 27th Int. Conf. Software, Telecommun. Comput. Networks, SoftCOM 2019*, pp. 1–6, 2019, doi: 10.23919/SOFTCOM.2019.8903886.

[16] P. H. Pwint and T. Shwe, “Network Traffic Anomaly Detection based on Apache Spark,” *2019 Int. Conf. Adv. Inf. Technol. ICAIT 2019*, pp. 222–226, 2019, doi: 10.1109/AITC.2019.8920897.

[17] L. Chen, Y. Zhang, Q. Zhao, G. Geng, and Z. Yan, “Detection of DNS DDoS Attacks with Random Forest Algorithm on Spark,” *Procedia Comput. Sci.*, vol. 134, pp. 310–315, 2018, doi: 10.1016/j.procs.2018.07.177.

[18] A. Alsirhani, S. Sampalli, and P. Bodorik, “DDoS Detection System: Utilizing Gradient Boosting Algorithm and Apache Spark,” *Can. Conf. Electr. Comput. Eng.*, vol. 2018-May, pp. 1–6, 2018, doi: 10.1109/CCECE.2018.8447671.

[19] T. Chang and C. Hsieh, “Detection and Analysis of Distributed Denial-of-service in Internet of Things — Employing Artificial Neural Network and Apache Spark Platform,” vol. 30, no. 4, pp. 857–867, 2018.

[20] M. Belouch, S. El Hadaj, and M. Idlianmiad, “Performance evaluation of intrusion detection based on machine learning using apache spark,” *Procedia Comput. Sci.*, vol. 127, pp. 1–6, 2018, doi: 10.1016/j.procs.2018.01.091.

[21] H. Zhang, S. Dai, Y. Li, and W. Zhang, “Real-time Distributed-Random-Forest-Based Network Intrusion Detection System Using Apache Spark,” *2018 IEEE 37th Int. Perform. Comput. Commun. Conf. IPCCC 2018*, pp. 1–7, 2018, doi: 10.1109/PCCC.2018.8711068.

[22] M. Assefi, E. Behravesh, G. Liu, and A. P. Tafti, “Big data machine learning using apache spark MLlib,” *Proc. - 2017 IEEE Int. Conf. Big Data, Big Data 2017*, vol. 2018-Janua, pp. 3492–3498, 2017, doi: 10.1109/BigData.2017.8258338.

[23] R. Yusof, N. I. Udzir, and A. Selamat, “An Evaluation on KNN-SVM Algorithm for Detection and Prediction of DDoS Attack An Evaluation on KNN-SVM Algorithm for Detection and Prediction of DDoS Attack,” no. August, 2016, doi: 10.1007/978-3-319-42007-3.

[24] X. Qin, T. Xu, and C. Wang, “DDoS attack detection using flow entropy and clustering technique,” *Proc. - 2015 11th Int. Conf. Comput. Intell. Secur. CIS 2015*, pp. 412–415, 2016, doi: 10.1109/CIS.2015.105.

[25] A. Rai and R. K. Challa, “Survey on recent DDoS mitigation techniques and comparative analysis,” *Proc. - 2016 2nd Int. Conf. Comput. Intell.*

Commun. Technol. CICT 2016, pp. 96–101, 2016, doi: 10.1109/CICT.2016.27.

[26] R. V. Deshmukh and K. K. Devadkar, "Understanding DDoS attack & its effect in cloud environment," *Procedia Comput. Sci.*, vol. 49, no. 1, pp. 202–210, 2015, doi: 10.1016/j.procs.2015.04.245.

[27] A. S. and R. M., "A Review of Hadoop Ecosystem for BigData," *Int. J. Comput. Appl.*, vol. 180, no. 14, pp. 35–40, 2018, doi: 10.5120/ijca2018916273.

[28] A. Raj and R. D'Souza, "A Review on Hadoop Eco System for Big Data," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, no. June, pp. 343–348, 2019, doi: 10.32628/cseit195172.

[29] S. Salloum, R. Dautov, X. Chen, P. X. Peng, and J. Z. Huang, "Big data analytics on Apache Spark," *Int. J. Data Sci. Anal.*, vol. 1, no. 3–4, pp. 145–164, 2016, doi: 10.1007/s41060-016-0027-9.

[30] G. A. Shoro and T. R. Soomro, "Big Data Analysis: Apache Spark Perspective," *Glob. J. Comput. Sci. Technol.*, vol. 15, no. 1, pp. 7–14, 2015, [Online]. Available: http://www.computerresearch.org/index.php/computer/article/viewFile/1137/1124.

[31] A. Bhattacharya and S. Bhatnagar, "Big Data and Apache Spark: A Review," *Int. J. Eng. Res. Sci. ISSN*, vol. 2, no. 5, pp. 206–210, 2016.

[32] S. Jonnalagadda, P. Srikanth, K. Thumati,] Sri, H. Nallamala, and A. Professors, "A Review Study of Apache Spark in Big Data Processing," *Int. J. Comput. Sci. Trends Technol.*, vol. 4, 2013.

[33] H. Luu, *Beginning Apache Spark 2*. 2018.

[34] E. E. Drakonaki and G. M. Allen, "Spark: Cluster Computing with Working Sets," *Skeletal Radiol.*, vol. 39, no. 4, pp. 391–396, 2010, doi:

10.1007/s00256-009-0861-0.

[35] H. Luu, *Beginning Apache Spark 2*. 2018.

[36] Ayon Dey, “Machine Learning Algorithms: A Review,” *Int. J. Comput. Sci. Inf. Technol.*, vol. 7, no. 3, pp. 1174–1179, 2016, doi: 10.21275/ART20203995.

[37] Q. Liu and Y. Wu, “Encyclopedia of the Sciences of Learning,” *Encycl. Sci. Learn.*, no. April, 2012, doi: 10.1007/978-1-4419-1428-6.

[38] E. S. Han and A. goleman, daniel; boyatzis, Richard; Mckee, “Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification,” *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2019.

[39] S. Dwivedi and L. K. P.Bhaiya, “A Systematic Review on K-Means Clustering Techniques,” *Int. J. Sci. Res. Eng. Trends*, vol. 5, no. 3, pp. 750–752, 2019.

[40] M. E. Celebi, H. A. Kingravi, and P. A. Vela, “A comparative study of efficient initialization methods for the k-means clustering algorithm,” *Expert Syst. Appl.*, vol. 40, no. 1, pp. 200–210, 2013, doi: 10.1016/j.eswa.2012.07.021.

[41] S. Shukla, “A Review ON K-means DATA Clustering APPROACH,” *Int. J. Inf. Comput. Technol.*, vol. 4, no. 17, pp. 1847–1860, 2014, [Online]. Available: http://www.irphouse.com.

[42] A. Singh, A. Yadav, and A. Rana, “K-means with Three different Distance Metrics,” *Int. J. Comput. Appl.*, vol. 67, no. 10, pp. 13–17, 2013, doi: 10.5120/11430-6785.

[43] T. M. Kodinariya and P. R. Makwana, “Review on determining of cluster in K-means,” *Int. J. Adv. Res. Comput. Sci. Manag. Stud.*, vol. 1, no. 6, pp. 90–95, 2013, [Online]. Available: https://www.researchgate.net/publication/313554124.

[44] G. K. Armah, G. Luo, and K. Qin, “A Deep Analysis of the Precision Formula for Imbalanced Class Distribution,” *Int. J. Mach. Learn. Comput.*, vol. 4, no. 5, pp. 417–422, 2014, doi: 10.7763/ijmlc.2014.v4.447.

[45] Maria Navin JR and Pankaja R, “Performance Analysis of Text Classification Algorithm using Confusion Matrix,” *Int. J. Eng. Tech. Res.* , vol. 6, no. 4, pp. 75–78, 2016.

[46] M. Hossin and M. . Sulaiman, “a Review on Evaluation Metrics for Data,” *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, no. 2, pp. 1–11, 2015.

[47] I. Sharafaldin, A. H. Lashkari, S. Hakak, and A. A. Ghorbani, “Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy,” *Proc. - Int. Carnahan Conf. Secur. Technol.*, vol. 2019-Octob, no. December, 2019, doi: 10.1109/CCST.2019.8888419.

[48] A. Naveen and T. Velmurugan, “Identification of calcification in MRI brain images by k-means algorithm,” *Indian J. Sci. Technol.*, vol. 8, no. 29, 2015, doi: 10.17485/ijst/2015/v8i29/83379.

[49] C. Zhu, C. Uwa, and W. Feng, “Informatics in Medicine Unlocked Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques,” *Informatics Med. Unlocked*, no. March, p. 100179, 2019, doi: 10.1016/j.imu.2019.100179.