iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.59078

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

Application of Robust Software Modelling Toolfor Web Attacks Detection

A. Mounika Rajeswari[1], Kalluri Rishita[2], Lanka Shriya[3], Balla Ganesh[4]
[1]*Assistance Professor,* [2, 3, 4]*UG Student, Department of Computer Science & Engineering, CMR College of Engineering & Technology, Hyderabad, India*

Abstract: ***The current inquiry as an examination of web attacks where proliferation of web-based applications has brought about a concurrent rise in cyber threats, particularly the form of web attacks targeting vulnerable systems. Approaches to web attack detection often rely on rule-based or signature-based methods, which struggle to change with the increasing landscape of attacks. In response, this study proposes an innovative approach leveraging DL techniques for web attacks. By harnessing the capability of DL, especially CNN and recurrent neural networks (RNNs), our proposed system learns directly from raw web traffic data, eliminating the need for manual feature engineering. This end-to-end approach not only streamlines the detection process but also enhances the system's ability to generalize across different types of attacks and adapt to new threats. To evaluate the effectiveness of our approach, we conducted extensive experiments on diverse datasets containing both benign and malicious web traffic. Our results demonstrate the superiority of end-to-end deep learning over traditional methods, achieving higher detection accuracy and robustness against adversarial attacks. In conclusion, our study highlights the promise of end-to-end deep learning as a viable approach related to web attacks, offering enhanced detection capabilities in the phase of evolving cyber threats.***

Keywords: Deep Learning, Web Attack Detection

I. INTRODUCTION

A web assault alludes to any malevolent movement or activity carried out with the deliberate of compromising the security, astuteness, or availabilities of websites, web applications, web servers, or there clients. Theses assaults misuse vulnerabilities or shortcomings in web advances, conventions, or arrangements to attain different evil goals. Web assaults can show in various shapes and can target distinctive layers of the internet stack, counting the application layer, organize layer, and server framework. A few common sorts of web assaults Cross-site Scripting (XSS), Cross-Site Ask Forgerty (CSRF), Denny of Benefit (DoS) Man-in-the-middle (MITM), Phishing attacks, Probe, R2L, U2R.

Halfond, W. G., Viegas, J., & Orso, A. [1], Web applications are medium to cyber-attacks, including common ones like SQL injection Wassermann, G., & Su, Z. [2], and inaccessible code execution. In spite of the advancement of countermeasures like firewalls and interruption location frameworks Raponi, S., Caprolu, M., & Di Pietro, R. [3], web assaults remain a critical danger. Investigation appears that over half of web applications amid a 2015-2016 filter contained noteworthy security vulnerabilities. Wrong positive confinements Pietraszek, T. [4] require manual choice of attack-specific highlights and tall untrue positive rates, making it basic to diminish these frameworks. An foundation that requires less mastery and labeled preparing information is required to address these challenges. Fu, X., Lu, X., Peltsverger, B., Chen, S., Qian, K., & Tao, L. [5] battle due to workforce impediments, classification impediments, and wrong positive restrictions. Workforce impediments include in-depth space information of web security, whereas classification restrictions include huge sums of labeled preparing information and the trouble of getting it for subjective custom applications. Su, T., Sun, H., Zhu, J., Wang, S., & Li, Y. [6] Consideration instrument is utilized to screen the organize stream vector composed of parcel vectors produced by the BLSTM demonstrate, which can get the key highlights for arrange activity classification. Numerous convolutional layers are utilized to handle information tests strategies.

II. LITERATURE SURVEY

Zargar, S. [7] In this paper they have discusses about the ceaseless flourishing of the financial advertise, MasterCard volume has until the end of time been impacting these a long time. The blackmail organizations are moreover rising rapidly. Beneath this circumstance, blackmail disclosure has turned into an progressively more critical issue. Be that as it may, the degree o f the distortion is completely much lower than the virtuoso trade, so the unevenness dataset makes this issue altogether more testing. In this paper we mainly promptho w to adjust to the Visa distortion distinguishing proof issue by utilizing supporting procedures and moreover gave a commitment ofthe brief examination between these making a difference methods.

Liu, J., Kantarci, B., & Adams, C. [8] This paper digs into the security vulnerabilities experienced by a wide cluster of Web of Things (IoT) gadgets and applications. The differing nature of IoT systems postures challenges for utilizing common benchmarks just like the NSL-KDD dataset to assess distinctive Arrange Interruption Discovery Frameworks (NIDS). To address this crevice, the paper analyzes particular assaults inside the NSL-KDD dataset that seem affect sensor hubs and systems in IoT situations. Moreover, it assesses eleven machine learning calculations to distinguish these assaults, displaying the comes about of their examination. The consider uncovers that tree-based strategies and gathering strategies perform way better than other machine learning approaches. Eminently, XG Boost rises as the top-performing administered calculation with 97curacy, a Matthews relationship coefficient (MCC) of90.5%, and an Zone Beneath the Bend (AUC) of 99.6%. Moreover, a critical finding is that the Expectation-Maximization (EM) calculation, an unsupervised strategy, too illustrates solid execution in recognizing assaults inside the NSL-KDD dataset, outperforming the exactness of the Naïve Bayes classifier by 22.0%.

Bisong, E. [9] This paper deals about the Education organized to create the information of machine learning, profound learning, information science, and cloud computing effortlessly open Prepares you with aptitudes to construct and send large - scale learning models on Google Cloud Stage Covers the programming abilities fundamental for machine learning and profound learning modeling utilizing the Python stack Incorporates bundles such as Numpy , Tensorflow, Matplotlib, Keras, Pandas and Scikit-learn.

III. STRUCTURE OF LSTM

The LSTM (Long Short-Term Memory) calculation offers unmistakable focuses of intrigued for recognizing web attacks in an end-to- end way compared to other calculations such as CNN, RNN, and customary machine learning calculations. LSTM basically bargains with Long-Term, Conditions, Continuous Modeling, Memory Cells, End-to-End Learning, Capturing Worldly Conditions, Learning Relevant Data.

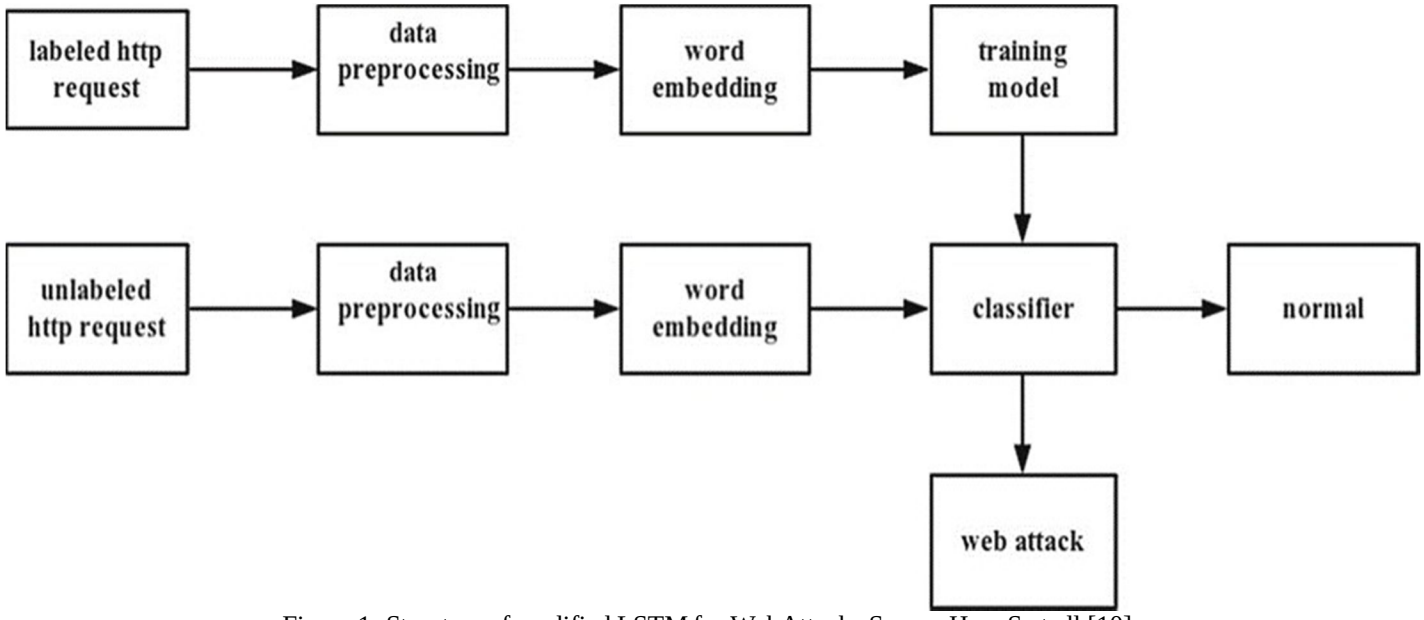


Figure 1**:** Structure of modified LSTM for WebAttacks Source Hao, S.et all [10]

IV. METHODOLOGY

A. Classification of Attacks

We address a assortment of assaults in our organize security endeavors. These incorporate Dissent of Benefit (DoS) assaults like apache2, back, arrive, neptune, mailbomb, unit, processtable, smurf, tear, udpstorm, and worm. Moreover, we center on Test assaults such as ipsweep, mscan, nmap, portsweep, holy person, and satan. Moreover, we consider Unauthorized Get to to Nearby Superuser (U2R) assaults like buffer_overflow, loadmodule, perl, ps, rootkit, sqlattack, and xterm. Finally, we address Unauthorized Get to from a Farther Machine (R2L) assaults like ftp_write, http_tunnel, imap, named, phf, sendmail, snmpgetattack, snmpguess, spy,warezmaster, xsnoop. These categories offer assistance us classify and get it the nature of arrange interruptions, permitting us to create compelling defense instruments.

B. Dataset

The NSL-KDD dataset serves as an upgraded form of the KDD'99 dataset, advertising a profitable asset for analysts within the field of interruption location frameworks (IDS) and organize security. Its primary reason is to supply an compelling benchmark for comparing different interruption discovery strategies. Categorized into four primary classes—Denial of Benefit (DoS), Test, Unauthorized Get to from a Farther Machine (R2L), and Unauthorized Get to to Nearby Superuser Benefits (U2R)—the dataset includes a assorted extendof cyber assaults experienced in arrange situations. Analysts utilize the NSL-KDD dataset to create and assess interruption discovery methods, pointing to upgrade the discovery and moderation of security dangers inside computer systems.

C. Data Analysis

Exploratory Data Examination (EDA) is an principal step in understanding and analyzing a dataset comprehensively. It incorporates a couple of key assignments pointed at picking up bits of information into the data's structure and characteristics. At to begin with, labeling the column names is essential, since it distributes critical identifiers to each incorporate, empowering less requesting explanation and examination. Taking after this, checking for invalid values ensures that there are no misplaced areas inside the dataset, which might something else skew examination comes approximately or obstruct appear execution. Along these lines, data visualization techniques such as making plots and charts are utilized to apparently talk to the transport of data and explore associations between differing highlights. These visualizations offer assistance in recognizing plans, designs, and peculiarities inside the dataset, in this way enlightening following steps inside the data examination get ready.

D. Feature Selection

Deciding the foremost relevant highlights may be a essential angle of information investigation, supporting in recognizing those traits that contribute most to anticipating the target variable or course name. In this setting, the recorded highlights are positioned based on their relationship with the target course. Highlights like 'dst_host_srv_count', 'logged_in', and 'dst_host_diff_srv_rate' display generally solid relationships with the target lesson, showing their potential centrality in recognizing between distinctive classes of organize activity. On the other hand, highlights such as 'num_shells' and 'urgent' appear weaker relationships with the target course, recommending they may have less prescient control or significance in this setting. Understanding the quality of these relationships guides the selection of highlights for building prescient models, guaranteeing that as it were the foremost enlightening traits are utilized, subsequently improving demonstrate exactness and productivity. Also, the nonattendance of relationship for 'num_outbound_cmds' with the target course highlights its negligible impact in separating between diverse classes of organize activity.

E. Algorithms

In our extent, we utilize a differing extend of calculations to address different angles of our issue. These calculations incorporate Choice Relapse, Bolster Vector Machines (SVM),Calculated Trees, Gaussian Naïve Bayes, as well as profound learning models suchas LSTM ,GRU, CNN, and RNN. Each calculation offers interesting qualities and capabilities suited to diverse sorts of information and assignments inside our venture. By leveraging this combination of conventional machine learning and profound learning strategies, we point to viably address the complexities and challenges show in our issue space, eventually moving forward the precision and strength of our arrangements.

F. Implementation Block Diagram

The flowchart starts with the introductory setup, where an application is opened and essential bundles are imported to encourage the advancement handle. Taking after this, the dataset is investigated and experiences information preprocessing, which regularly includes errands like cleaning the information, taking care of lost values, and changing the information into a appropriate organize for investigation. Another, the flowchart delineates a few key steps within the include designing handle, counting include era, include choice, and name encoding. These steps offer assistance in planning the information for encourage investigation and modeling. The flowchart at that point moves to the preparing and testing stages of the venture. Within the preparing stage, different machine learning calculations are connected, counting KFold cross-validation, calculated relapse, back vector machines (SVM), credulous Bayes, irregular timberlands, stacking classifiers, and voting classifiers. Also, profound learning procedures such as CNN, LSTM, gated repetitive units (GRU), and repetitive neural systems (RNN) are utilized amid the testing stage to assess the execution of the models.

At long last, the flowchart concludes with the execution of user registration and login functionalities, allowing clients to supply input and get the ultimate yield or result from the online location. This organized approach laid out within the flowchart guarantees a precise and organized advancement handle, driving to the creation of a useful and user-friendly online stage.

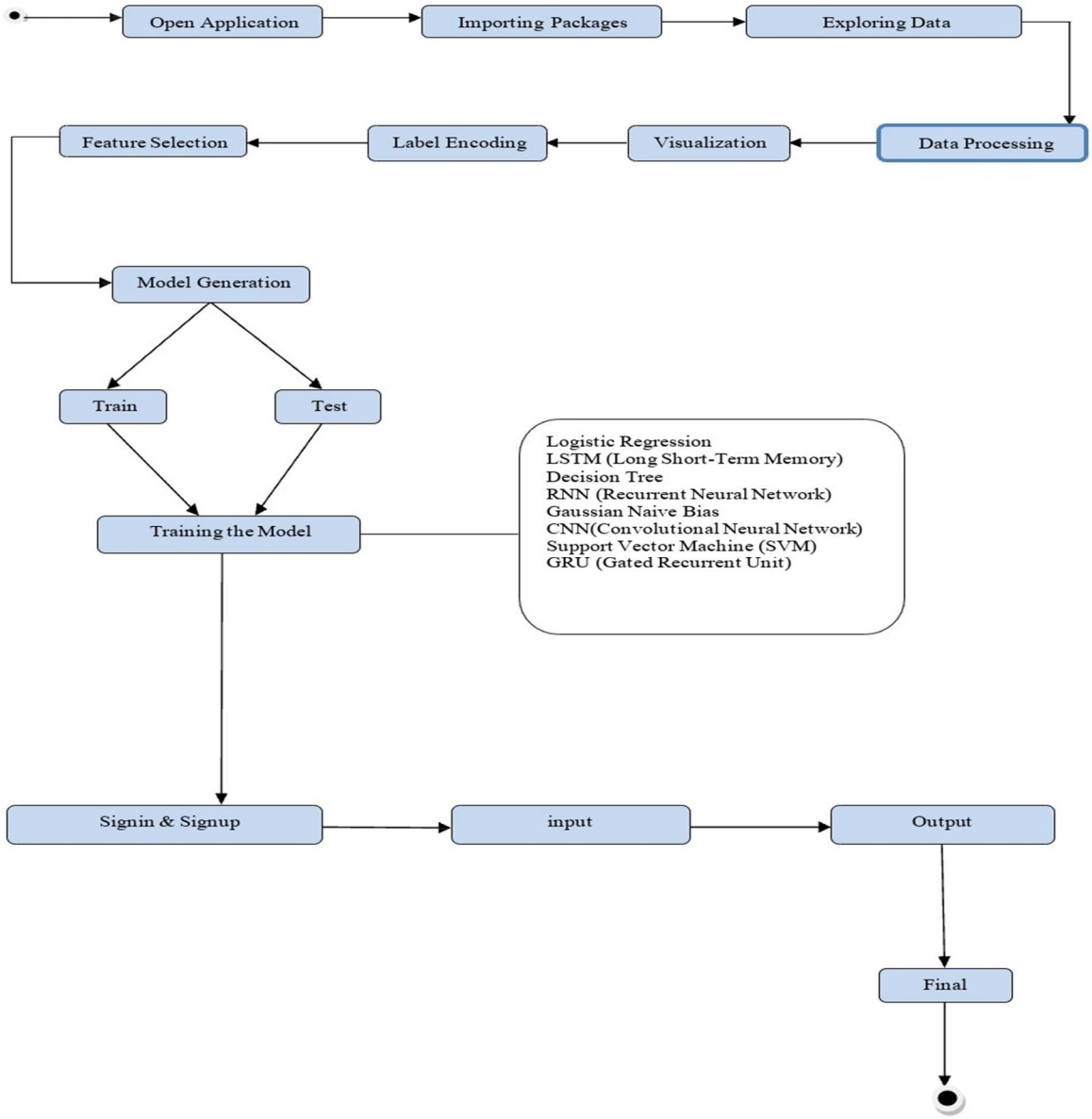


Figure 2**:** Block Diagram of Proposed System's Implementation

V. PERFORMANCE VALIDATION

A. *Performance Validation Of Machine Learning Algorithms*

Table 1: Precision

Algorithm	Normal-Attack	DoS-Attack	R2L-Attack	Probe-Attack	U2R-Attack
Decision Tree	0.967720	0.972861	0.744722	0.826865	0.0
Logistic Regression	0.888980	0.944034	0.0	0.857955	0.0
SVM	0.169973	0.716787	0.057082	0.169648	0.0
GNB	0.946300	0.974133	0.043536	0.341962	0.002609

Table 2: Recall

Algorithm	Normal-Attack	DoS-Attack	R2L-Attack	Probe-Attack	U2R-Attack
Decision Tree	0.954000	0.968885	0.487437	0.936335	0.0
Logistic Regression	0.980076	0.939841	0.0	0.439190	0.0
SVM	0.087330	0.097022	0.440955	0.611871	0.0
GNB	0.678349	0.746969	0.204774	0.576322	0.974359

Table 3: F1-Score

Algorithm	Normal-Attack	DoS-Attack	R2L-Attack	Probe-Attack	U2R-Attack
Decision Tree	0.960811	0.970869	0.589218	0.878202	0.0
Logistic Regression	0.932308	0.941933	0.0	0.580976	0.0
SVM	0.115380	0.170910	0.101080	0.265644	0.0
GNB	0.790228	0.845560	0.071806	0.429236	0.005204

Table 4: Support

Algorithm	Normal-Attack	DoS-Attack	R2L-Attack	Probe-Attack	U2R-Attack
Decision Tree	53956.000000	36703.000000	796.000000	9283.000000	39.0
Logistic Regression	53956.000000	36703.000000	796.0	9283.000000	39.0
SVM	53956.000000	36703.000000	796.000000	9283.000000	39.0
GNB	53956.000000	36703.000000	796.000000	9283.000000	39.000000

B. Performance Validation Of Deep Learning Algorithms

Table 5: Average Accuracy across k-Folds (k=10)

Algorithm	Average Accuracy across k-folds (K=10)
GRU	0.9850327432155609
LSTM	0.9848382592201232
RNN	0.9822266280651093
CNN	0.9752728700637817

VI. PERFORMANCE VALIDATION

A. Comparison Of Algorithms

Table 6: Comparison of Algorithms

Algorithm	Accuracy	Precision	F1-Score	Recall	Sensitivity	Specificity
Decision Tree	95.336376	70.186724	67.715002	66.592677	97.558317	98.265495
Logistic Regression	90.208375	53.396384	48.559363	46.693533	93.774488	99.010873
SVM	14.054376	21.989918	12.874339	24.240124	15.304756	75.351641
Gaussian Naïve Bayes	68.779520	46.208900	42.779500	62.191868	94.777563	98.068995
GRU	98.654495	72.935717	74.018918	75.251725	99.337605	99.660557
LSTM	98.765628	74.384986	74.783166	75.194744	99.207469	99.856809
RNN	98.543362	76.799996	72.906212	70.491976	99.197049	99.706723
CNN	97.983727	75.900331	71.622521	68.963413	99.043270	99.586435

Table 6 interprets the data of the comparison of the accuracy, precision, f1-score, recall, sensitivity and specificity of Decision Tree, Logistic Regression, SVM, Gaussian Naïve Bayes, GRU, LSTM, RNN and CNN algorithms. We can have a clear view that in terms of accuracy, f1-score and specificity LSTM has the high performance and in precision RNN, in recall and sensitivity GRU.

B. Figures

Figure 3 shows the graphical representation of the accuracy, precision, f1-score, recall, sensitivity and specificity of the algorithms.

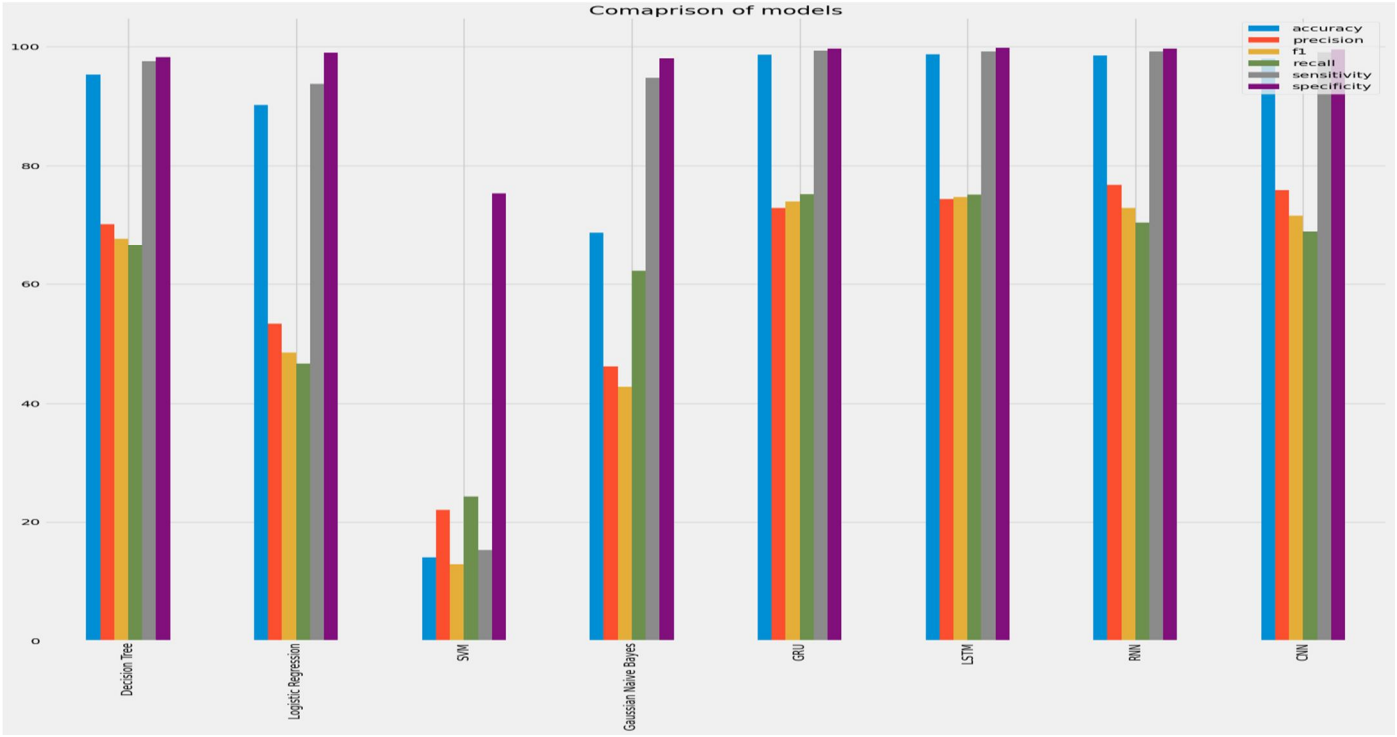


Figure 3: Graphical representation of Comparison of models

Figure 4 shows the graphical representation of the comparison of cross validation accuracy of algorithms.

Cross Validation Accuracy: It is a robust technique used as a performance metric to compare the efficiency of different models.

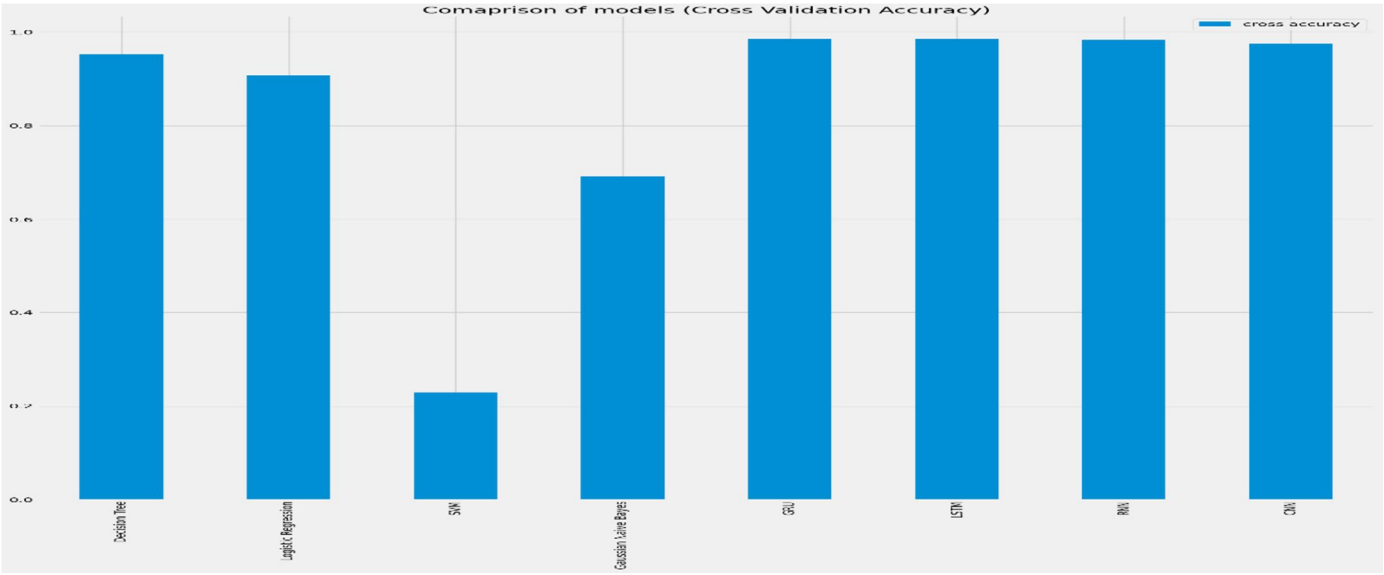


Figure 4: Graphical representation of Cross Validation Accuracy

Figure 5 shows the execution time of different models to train and test the dataset. We can generalize that the Deep Learning models takes longer time to train than the ML algorithms. LSTM and GAN take the maximum time to train the model.

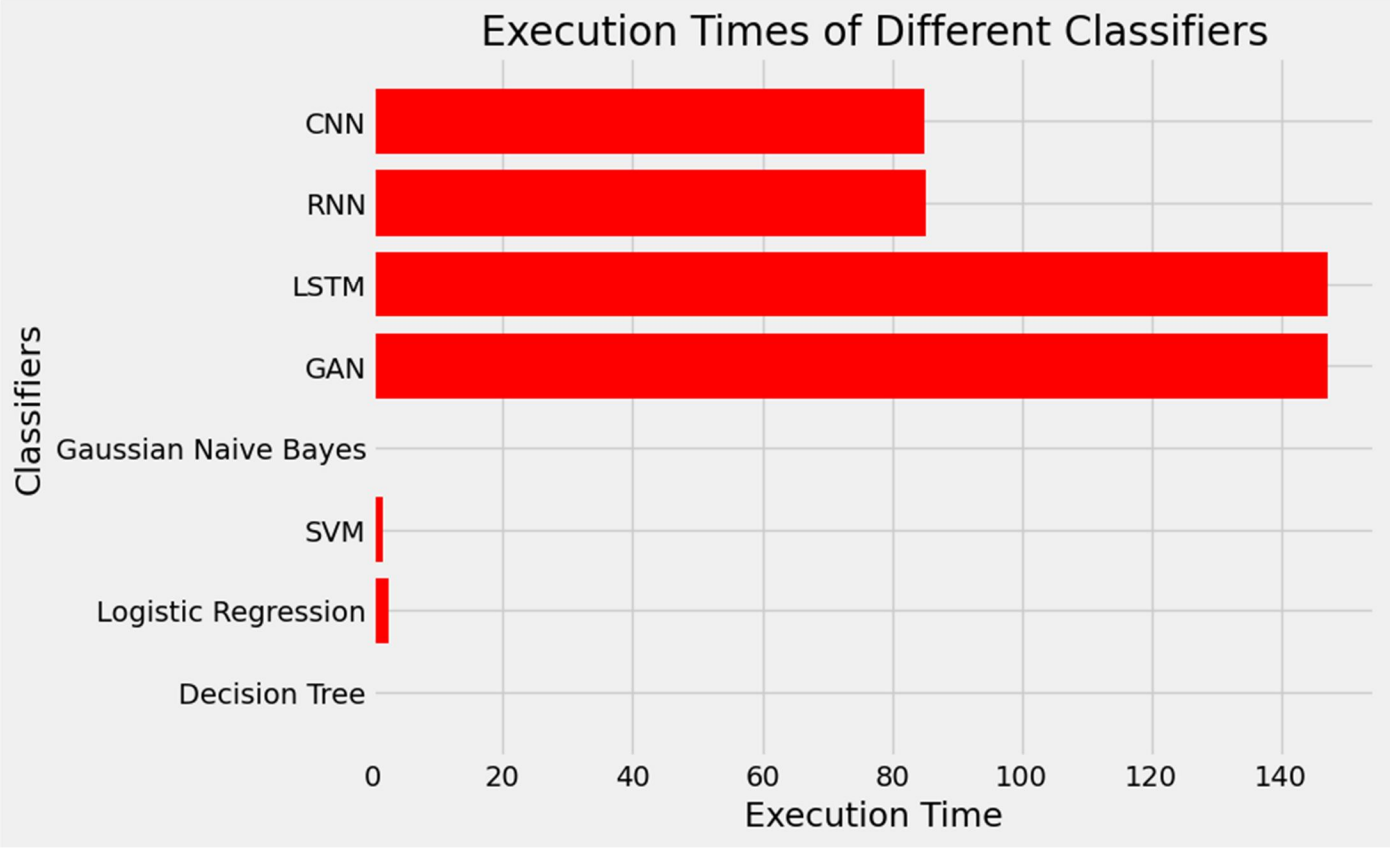


Figure 5: Graphical Representation of Execution Times of algorithms

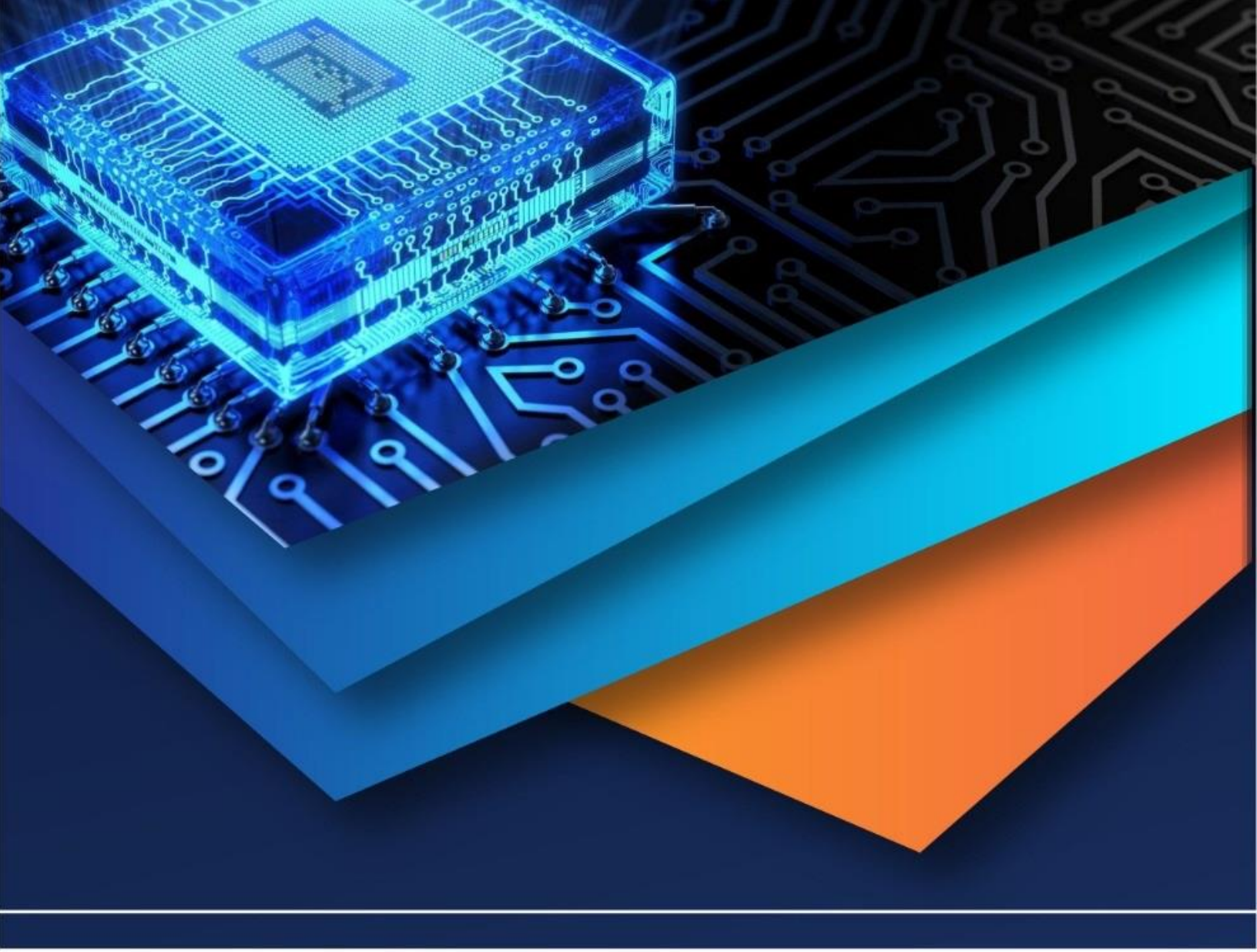
VII. CONCLUSION

The ponder presents a novel approach for arrange assault discovery, leveraging profound methods like LSTM nearby conventional machine learning models. Through tests conducted on NSL-KDD dataset, both LSTM and different machine learning calculations wereutilized for execution assessment. The discoveries of this investigation not as it were illustrating the adequacy of the LSTM demonstrate but too highlight its predominance over existing state-of-the-art approaches, counting machine learning calculations. This approval through execution comparison underscores the potential of LSTM as a strong arrangement for organize assault discovery in real-world scenarios. In future endeavors, the center will be on optimizing the computational productivity of the LSTM demonstrate. This involves refining its design to diminish computational costs without compromising discovery precision. Furthermore, the proposed demonstrate will experience encourage preparing on assorted sorts of assaults to guarantee its adequacy in tending to modern and advancing dangers. In outline, the consider presents a promising progression in arrange assault location, advertising a profound learning approach with LSTM that outperforms conventional machine learning strategies in terms of execution. The commitment to future enhancements, counting computational optimization and improved versatility to rising dangers, implies a proactive position in progressing cybersecurity measures for arrange defense.

REFERENCES

[1] Halfond, W. G., Viegas, J., & Orso, A. (2006, March). A classification of SQL-injection attacks and countermeasures. In Proceedings of the IEEE international symposium on secure software engineering (Vol. 1, pp. 13-15). IEEE.
[2] Wassermann, G., & Su, Z. (2008, May). Static detection of cross-site scripting vulnerabilities. In Proceedings of the 30th international conference on Software engineering (pp. 171-180).
[3] Raponi, S., Caprolu, M., & Di Pietro, R. (2019). Intrusion detection at the network edge: Solutions, limitations, and future directions. In Edge Computing–EDGE 2019: Third International Conference, Held as Part of the Services Conference Federation, SCF 2019, San Diego, CA, USA, June 25–30, 2019, Proceedings 3 (pp. 59- 75). Springer International Publishing.

[4] Pietraszek, T. (2004). Using adaptive alert classification to reduce false positives in intrusion detection. In Recent Advances in Intrusion Detection: 7th International Symposium, RAID 2004, Sophia Antipolis, France, September 15-17, 2004. Proceedings 7 (pp. 102-124). Springer Berlin Heidelberg.
[5] Fu, X., Lu, X., Peltsverger, B., Chen, S., Qian, K., & Tao, L. (2007, July). A static analysis framework for detecting SQL injection vulnerabilities. In 31st annual international computer software and applications conference (COMPSAC 2007) (Vol. 1, pp. 87-96). IEEE.
[6] Su, T., Sun, H., Zhu, J., Wang, S., & Li, Y. (2020). BAT: Deep learning methods on network intrusion detection using NSL-KDD dataset. IEEE Access, 8, 29575- 29585.
[7] Zargar, S. (2021). Introduction to sequence learning models: RNN, LSTM, GRU. Department of Mechanical and Aerospace Engineering, North Carolina State University, Raleigh, North Carolina, 27606.
[8] Liu, J., Kantarci, B., & Adams, C. (2020, July). Machine learning-driven intrusion detection for Contiki-NG-based IoT networks exposed to NSL-KDD dataset. In Proceedings of the 2nd ACM workshop on wireless security and machine learning (pp. 25-30).
[9] Bisong, E. (2019). Building machine learning and deep learning models on Google cloud platform (pp. 59-64). Berkeley, CA: Apress.
[10] Hao, S., Long, J., & Yang, Y. (2019, April). Bl-ids: Detecting web attacks using bi-lstm model based on deep learning. In International conference on security and privacy in new computing environments (pp. 551-563). Cham: Springer International Publishing.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY