IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Survey Paper on Cyber Attack Detection Model (CADM) Using Machine Learning

Network Attack Detection

[1]Ganesh Mankar, [2]Rakesh Oza, [3]Saurabh Sapkal, [4]Ayush Tilekar, [5]Prof. Rakhi Punwatkar
[1]Student, [2]Student, [3]Student, [4]Student, [5]Professor
Computer Department
Zeal College of Engineering and Research, Pune, Maharashtra, India

Abstract: In this survey study, the Cyber Attack Detection Model (CADM) is specifically highlighted as it examines the field of machine learning applications in Cyber Attack detection. In a time where technology is king, strong defenses are required due to the growing threat of Cyber Attacks. The study examines the literature on Cyber Attack detection, highlighting the development of machine learning approaches and the function of intrusion detection systems (IDS). As an extensive case study, the CADM project demonstrates the complexities of its design, feature selection, preprocessing, data gathering, and ensemble integration techniques. Metrics for performance assessment, including as recall, accuracy, precision, and AUROC, are examined along with comparisons to alternative systems. The report sheds light on the difficulties in detecting Cyber Attacks and suggests future paths for the field's advancement. In the end, the CADM project proves to be a noteworthy addition to the improvement of Cybersecurity by means of inventive machine learning techniques.

Keywords **–** Cyber Attack, LASSO, Random Forest, Gradient Boosting

INTRODUCTION

Cybersecurity stands as a paramount concern in today's technologically driven landscape, with the proliferation of networked devices intensifying the need for robust defense mechanisms. Intrusion Detection Systems (IDS) play a pivotal role in safeguarding against Cyber threats, and the integration of machine learning has emerged as a promising avenue for bolstering their capabilities. This survey paper delves into the realm of machine learning applications in Cyber Attack detection, focusing on the Cyber Attack Detection Model (CADM) as an exemplar. As Cyber adversaries become increasingly sophisticated, a comprehensive understanding of the methodologies employed in modern detection systems becomes imperative. The survey commences with an exploration of the broader context of Cybersecurity, highlighting the challenges posed by Cyber Attacks. It then transitions to an overview of existing research in the field, emphasizing the evolution of machine learning techniques in Intrusion Detection Systems. The CADM project, a noteworthy contribution to this domain, is introduced as a case study, offering insights into its architecture, data collection strategies, preprocessing methodologies, feature selection techniques, and ensemble integration methods. The paper aims to provide a holistic understanding of the intricacies involved in utilizing machine learning for Cyber Attack detection, with CADM serving as an illustrative example of innovative advancements in this critical domain.

The rising frequency and sophistication of Cyber Attacks have fueled the development of sophisticated detection systems in the everchanging field of Cybersecurity. The Cyber Attack Detection Model (CADM) is the focal point of this survey paper's investigation into the field of machine learning applications for Cyber Attack detection. Cybersecurity is a critical component of contemporary technology that is essential for protecting networks from harmful activity. In this environment, intrusion detection systems (IDS) have become essential tools, and machine learning approaches are becoming essential to maximize their effectiveness.

The CADM project offers a comprehensive method for detecting Cyber Attacks, making it an insightful case study. This survey walks you through the architecture of the project, explaining its nuances in terms of feature selection, data collection, preprocessing, and ensemble integration techniques. In training and testing the CADM model, the importance of datasets like NSL-KDD, KDD Cup 99, UNSW-NB15, URL 2016, and CICIDS 2017 is examined. The importance of pertinent features in building an effective machine learning model is highlighted through the discussion of feature selection approaches, such as LASSO regularization. Metrics for performance evaluation, including recall, accuracy, precision, and AUROC, are carefully studied to give a thorough picture of CADM's capabilities. Analyses conducted in comparison with current systems highlight the model's superior accuracy and precision. In-depth discussion of the difficulties in detecting Cyber Attacks is provided by the survey, which also opens the door to possible future research directions.

LITERATURE REVIEW

The study paper suggests a machine learning-based Cyber Attack Detection Model (CADM). The CADM uses an ensemble classification method to analyze patterns in network data and categorize cyberattacks. The model works with large datasets, applies visualization features, and extracts features using LASSO. For classification, an ensemble approach consisting of the Random Forest and Gradient Boosting algorithms is employed. Using a variety of datasets, such as NSL-KDD, KDD Cup 99, UNSW-NB15, URL 2016, and CICIDS 2017, the study assesses the model.

The article highlights the difficulties in developing CADM and stresses the significance of picking the right preprocessing mechanism, relevant feature extraction, and classification algorithm. The research has produced an automated CADM using DBSCAN and LASSO, identified intricate attack patterns, suggested a machine learning-based system for unknown threats, and developed a model that can be applied to several datasets.

Data collection from five benchmark datasets, dataset merging, outlier removal preprocessing, LASSO feature selection, and ensemble classification using Random Forest and Gradient Boosting are the steps of the suggested methodology. True Positive Rate, False Positive Rate, Average Accuracy, Precision, Recall, and AUROC are all assessed as part of the performance analysis.

The results show that the classifiers using Random Forest and Gradient Boosting perform well and achieve high classification accuracy. The suggested model performs better in terms of accuracy, precision, recall, false positive rate, and F1-score when compared to current systems, as the paper shows. In a number of datasets, such as NSL-KDD, KDD Cup 99, UNSW-NB15, URL 2016, and CICIDS 2017, the CADM model exhibits better performance.

The paper concludes with a thorough methodology for utilizing machine learning techniques to construct a Cyber Attack Detection Model. The authors propose that future work could focus on improving accuracy for particular attacks, exploring more attacks in publicly available datasets, and developing a prevention system based on the CADM. The proposed model exhibits promising results across multiple datasets.

RESEARCH METHODOLOGY

Literature Review: Carry out a thorough analysis of the body of knowledge regarding machine learning applications in Cyber Attack detection. Find important research papers, articles, and conference proceedings on subjects like machine learning algorithms for threat detection, Cybersecurity, and intrusion detection systems.

Selection of Relevant Works: Focus only on literature that addresses machine learning in the context of Cyber Attack detection. Give top priority to research on feature selection, ensemble methods, and integrating machine learning into practical applications.

Case Study on CADM: Thorough examination of the CADM project with a focus on its methods, architecture, and constituent parts. Examine the goals of the project, the datasets (NSL-KDD, KDD Cup 99, UNSW-NB15, URL 2016, CICIDS 2017) that were used, and the reasoning behind the selection of each dataset.

Data Collection and Preprocessing: Examine the significance of preprocessing and gathering datasets for CADM. Examine the ways in which CADM handles categorical features, outliers, and missing values during the data preprocessing stage.

Feature Selection and Extraction: Examine CADM's feature selection procedure, paying particular attention to LASSO regularization and other methods. Recognize how certain features contribute to increasing the accuracy and effectiveness of the model.

Ensemble Integration Methods: Examine how ensemble approaches, in particular the Random Forest and Gradient Boosting algorithms, can be used in CADM. Examine the ways in which these techniques enhance the Cyber Attack detection model's overall classification accuracy and robustness.

Performance Evaluation Metrics: Examine the AUROC, accuracy, precision, recall, and other performance evaluation metrics used in CADM. Examine these metrics in comparison to industry norms and talk about their importance.

Comparative Analysis: Compare and contrast CADM with the current Cyber Attack detection systems. Describe the advantages, disadvantages, and improvements of CADM over other models in terms of precision, accuracy, and adaptability.

Challenges and Future Directions: Examine the difficulties CADM has detecting Cyber Attacks. Give possible directions for further study and development, taking into account the dynamic nature of Cyber threats.

Conclusions and Synthesis: Combine the results of the comparative analysis, case study on CADM, and literature review. Using CADM as an example, draw general conclusions regarding the efficacy of machine learning techniques in CyberAttack detection.

Presentation of Results: Present the survey's results in an understandable and well-organized way, making appropriate use of visuals like graphs and charts. Make sure there is a roadmap in the methodology section.

RESULTS

Comprehensive Analysis: A thorough analysis of the body of research on machine learning applications in CyberAttack detection, offering a comprehensive picture of the field's current status.

CADM Case Study Analysis: Comprehensive explanations of the architecture, data gathering, preprocessing, feature selection, and ensemble integration techniques of the CADM project.

Dataset Significance: The importance of the datasets used in CADM (NSL-KDD, KDD Cup 99, UNSW-NB15, URL 2016, CICIDS 2017) for testing and training, as well as their effect on the model's performance, are clearly explained.

Data Preprocessing Techniques: Gain an understanding of how CADM addresses issues with data preprocessing, including missing values, outliers, and categorical features, all of which strengthen the model's resilience.

Feature Selection Insights: Insights into the CADM feature selection procedure, with a focus on the use of LASSO regularization and how it enhances model accuracy.

Ensemble Integration Effectiveness: An assessment of the ensemble techniques' (Random Forest and Gradient Boosting) performance in CADM, emphasizing their contributions to precise Cyber Attack classification.

Performance Evaluation Metrics: The model's ability to produce dependable results is demonstrated by the clear presentation and analysis of the performance evaluation metrics (accuracy, precision, recall, and AUROC) used in CADM.

Comparative Analysis: A comparison of current Cyber Attack detection systems with CADM, highlighting the advantages and improvements of CADM with regard to precision, accuracy, and adaptability.
Identification of Challenges: This section identifies and discusses the difficulties CADM has encountered when detecting Cyber Attacks, offering suggestions for future research directions.
Future Research Directions: Building on the knowledge gained from CADM, possible future research avenues and advancements in the field of machine learning for Cyber Attack detection are suggested.
Synthesized Conclusions: Using CADM as a prominent case study, a synthesis of the data results in general conclusions regarding the efficacy of machine learning techniques in Cyber Attack detection.
Contribution to Knowledge: Significant knowledge and insight addition to the body of literature, with the survey report acting as a thorough guide for academics, industry professionals, and Cybersecurity enthusiasts.

CONCLUSION

In conclusion, this project has shown that machine learning can be used to detect network and CyberAttacks. The suggested model demonstrated a low false positive rate of less than 1% while achieving high accuracy in identifying known as well as unknown Attacks. In order to safeguard networks from harm, the model could also identify Attacks in real time. The primary benefit of utilizing machine learning for network Attack detection is its adaptability to new and developing Attack types. The model can be re-trained on fresh data to pick up on new Attack patterns when the threat landscape shifts. Because of this, machine learning is an effective tool for defending networks against intrusions.
However, there are certain difficulties in applying machine learning to the detection of network Attacks. The requirement to train the model on a sizable dataset of tagged Attack and non-Attack traffic presents one difficulty. The collection of this data can be costly and time-consuming. The requirement for the model to detect Attacks in real time, which can be computationally demanding, presents another difficulty.
Machine learning is a promising method for detecting CyberAttacks in spite of these difficulties. Even in real time, it is possible to detect network Attacks with high accuracy, as demonstrated by the suggested model. It is expected that machine learning will become even more important in defending networks against intrusions as the field of machine learning technology advances.

References

[1] H. Alqahtani, I. Sarker, A. Kalim, S. Minhaz Hossain, S. Ikhlaq and S. Hossain, "Cyber Intrusion Detection Using Machine Learning Classification Techniques," In Proc. International Conference on Communications in Computer and Information Science, pp. 121-131, 2020. doi: 10.1007/978- 981-15-6648-6 10 [Accessed 27 October 2020].
[2] P. Negandhi, Y. Trivedi and R. Mangrulkar, "Intrusion Detection System Using Random Forest on the NSL-KDD Dataset," Emerging Research in Computing, Information, Communication and Applications, pp. 519-531, 2019. doi: 10.1007/978-981-13-6001-5_43 [Accessed 12 July 2020].
[3] B. Ahmad, W. Jian and Z. Anwar Ali, "Role of Machine Learning and Data Mining in Internet Security: Standing State with Future Directions," Journal of Computer Networks and Communications, vol. 2018, pp. 1-10, 2018. doi: 10.1155/2018/6383145 [Accessed 2 October 2020].
[4] A. Gupta, G. Prasad and S. Nayak, "A New and Secure Intrusion Detecting System for Detection of Anomalies Within the Big Data," Studies in Big Data, pp. 177-190, 2018. doi: 10.1007/978-3-030-03359-0_8 [Accessed 30 August 2020].
[5] T. Tang, D. McLernon, L. Mhamdi, S. Zaidi and M. Ghogho, "Intrusion Detection in SDN-Based Networks: Deep Recurrent Neural Network Approach," Deep Learning Applications fo r Cyber Security, pp. 175-195, 2019. doi: 10.1007/978-3-030-13057-2_8' [Accessed 30 August 2020].
[6] C. Gayathri Harshitha, M. Kameswara Rao and P. Neelesh Kumar, "A Novel Mechanism for Host-Based Intrusion Detection System," In Proc. First International Conference on Sustainable Technologies for Computational Intelligence, pp. 527-536, 2019. doi: 10.1007/978-981-15- 0029-9 42 [Accessed 21 June 2020].
[7] A. Ahmim, M. Ferrag, L. Maglaras, M. Derdour and H. Janicke, "A Detailed Analysis of Using Supervised Machine Learning for Intrusion Detection," Strategic Innovative M arketing and Tourism, pp. 629-639, 2020. doi: 10.1007/978-3-030-36126-6 70 [Accessed 7 August 2020].
[8] R. Jaiswal and S. Lokhande, "Analysis of Early Traffic Processing and Comparison of Machine Learning Algorithms for Real Time Internet Traffic Identification Using Statistical Approach," Advanced Computing, Networking and Informatics, vol. 2, Sm art Innovation, Systems and Technologies, vol 28, pp. 577-587, 2014. doi: 10.1007/978-3-319-07350- 7 64 [Accessed 24 September 2020].
[9] W. Zong, Y. Chow and W. Susilo, "Interactive three-dimensional visualization of network intrusion detection data for machine learning," Future Generation Computer Systems, vol. 102, pp. 292-306, 2020. doi: 10.1016/j.future.2019.07.045 [Accessed 24 September 2020].
[10] H. Liu and A. Gegov, "Collaborative Decision Making by Ensemble Rule Based Classification Systems," Studies in B ig Data, pp. 245-264, 2015. doi: 10.1007/978-3-319-16829-6_10 [Accessed 20 September 2020].