International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI:10.55041/ISJEM05975
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Machine Learning Based Malware Trend Analysis using Cyber Threat Intelligence Data

[1]M D Mahesh, [2]Sunnapu Munilakshmi

[1]Assistant Professor, [2]Postgraduate

[1]Department of MCA, Annamacharya Institute of Technology and Sciences, Tirupati, Andhra Pradesh, India.

[2]Department of MCA, Annamacharya Institute of Technology and Sciences, Tirupati, Andhra Pradesh, India.

Abstract

The number of cyberattacks and malware-related risks on a variety of online platforms has dramatically expanded due to the quick development of digital technology. Because cyber threats are always changing, traditional cybersecurity techniques frequently fail to identify new virus patterns. Threat detection systems can be improved by using Cyber Threat Intelligence (CTI), which offers useful information on malicious activity, attack behaviours, and threat actors. This study offers a machine learning-based method for utilising cyber threat intelligence data to analyse malware trends. The suggested approach makes use of a dataset that includes information about threats and uses data pretreatment methods to get the data ready for model training. To find harmful trends in the dataset, three machine learning algorithms—Random Forest, Support Vector Machine, and Decision Tree—are put into practice and assessed. Evaluation criteria like precision, recall, F1-score, and accuracy are used to gauge these models' performance. According to experimental results, the Random Forest model outperforms all other examined algorithms in terms of accuracy, indicating its efficacy in identifying dangers related to malware. Users can effectively carry out training and prediction tasks thanks to the system's deeper integration into a web-based interface. The suggested method helps security experts recognise new malware trends more successfully and advances cyber threat analysis.

Keywords

Cyber Threat Intelligence, Malware Detection, Machine Learning, Random Forest, Support Vector Machine, Decision Tree, Cybersecurity Analytics, Threat Pattern Analysis, Malware Trend Detection, Data-Driven Security Analysis.

I.Introduction

The threat landscape in contemporary computing systems has greatly increased due to the growing reliance on digital technology and internet-based services. Malware infections, ransomware assaults, phishing operations, and advanced persistent threats are examples of cyberattacks that have increased in frequency and sophistication. Individuals, organisations, and vital infrastructures are at grave risk from these threats. Conventional security systems mostly rely on signature-based detection techniques, which are frequently unsuccessful in spotting recently discovered or unidentified malware variants. Intelligent methods that can evaluate cyber threat data and identify harmful activity in a more flexible and effective way are therefore becoming more and more necessary.

By gathering and evaluating data about cyberattacks, threat actors, vulnerabilities, and hostile infrastructure, Cyber Threat Intelligence (CTI) plays a significant part in contemporary cybersecurity. Security analysts can better analyse attack trends and anticipate possible dangers by using the insightful information included in CTI data. However, manual analysis is challenging and time-consuming due to the vast amount and intricate structure of threat intelligence data. Therefore, in order to extract significant patterns from such data, automated methods are needed.

Because machine learning algorithms may learn from past data and find hidden patterns linked to cyberattacks, they have recently attracted a lot of attention in the field of cybersecurity. It is feasible to identify dangerous behaviours, categorise malware activity, and more thoroughly examine new threat patterns by using machine learning algorithms on cyber threat

intelligence data.

The main goal of this study is to assess malware trends by applying machine learning techniques to cyber threat intelligence data. To find the best model for malware detection, the suggested system applies and contrasts several classification techniques, such as Random Forest, Support Vector Machine, and Decision Tree.cThis study aims to enhance malware trend identification by utilising cyber threat intelligence data and machine learning methods. The outcomes of the experiment show how well the suggested method works to spot malicious tendencies and assist cybersecurity specialists in more effectively spotting possible threats.

II. Problem Statement

Cyberattacks and virus dangers have greatly expanded due to the quick growth of digital systems and internet-based services. Ransomware, trojans, and spyware are just a few of the malware threats that both individuals and organisations must constantly deal with. Conventional cybersecurity systems mostly use signature-based detection methods, which are only useful for known threats. Nevertheless, these techniques frequently fall short in identifying recently developed or altered malware strains.

Large amounts of information about cyberattacks, threat actors, vulnerabilities, and harmful activity are provided by Cyber Threat Intelligence (CTI). The breadth, complexity, and dynamic nature of this data make manual analysis challenging, even though it offers useful information for spotting virus trends. Automated solutions that can effectively process this data and identify dangerous trends are necessary for security analysts.

In order to identify threats and identify malware patterns, an intelligent system that can use machine learning techniques to analyse cyber threat intelligence data is required. By using machine learning techniques like Random Forest, Support Vector Machine, and Decision Tree to evaluate cyber threat data and more accurately categorise criminal activity, the suggested solution seeks to solve this issue.

III. Proposed System

The suggested approach offers a machine learning-based framework for utilising cyber threat intelligence data to analyse malware trends. The system's primary goal is to use sophisticated machine learning algorithms to automatically identify dangerous patterns and categorise possible malware activity. Cyber threat intelligence data is processed by the system and converted into an organised format that can be used for model analysis and training.

The dataset that contains information on cyber threats is first gathered and preprocessed in order to eliminate noise, deal with missing values, and get the data ready for additional analysis. Following preprocessing, the textual and structured threat information is transformed into numerical representations that machine learning models can exploit by applying feature extraction techniques.

The system analyses the dataset and finds patterns connected to malware using a variety of classification algorithms, such as Random Forest, Support Vector Machine, and Decision Tree. The processed dataset is used to train these models, and performance metrics including accuracy, precision, recall, and F1-score are used to assess them. A web-based interface is created utilising HTML-based frontend templates and a backend framework to make the system interactive and user-friendly. This interface allows users to make predictions using cyber threat intelligence data and train machine learning models. The suggested system facilitates proactive cybersecurity analysis and makes it easier for security professionals to identify malware trends.

IV. Dataset Description

This study uses a Cyber Threat Intelligence (CTI) dataset, which includes data on malware activity, attack patterns, and cyber threats. The dataset, which includes structured records detailing various cybersecurity incidents and threat indicators, is kept in a CSV file called "Cyber-Threat-Intelligence-Custom-Data_new_processed.csv." A number of attributes, including id, text, relations, diagnosis, and solutions, are present in every record in the collection and offer thorough

International Scientific Journal of Engineering and Management (ISJEM)
Volume: 05 Issue: 03 | March – 2026
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
DOI:10.55041/ISJEM05975

explanations of cyber threat incidents. The primary threat intelligence description is found in the text column; related threat entities and security details are shown in additional columns. Attack patterns, malware, software, URLs, infrastructure, identities, tools, campaigns, and vulnerabilities are among the labelled entity data included in the collection.

The dataset also includes other entity-related columns, including id_1, label_1, start_offset, end_offset, id_2, label_2, start_offset_2, end_offset_2, id_3, label_3, start_offset_3, and end_offset_3. The location and kind of security entities contained in the threat description text are indicated by these columns. Structured threat analysis is made possible by the offset values, which show where particular items are located in the text. The dataset goes through a preprocessing step where irrelevant values are eliminated and textual data is transformed into numerical representations before the machine learning models are trained. The threat intelligence data is converted into a format that machine learning algorithms can use by using feature extraction techniques. To assess the effectiveness of the system's classification models, the processed dataset is subsequently split into training and testing subsets.

This dataset enables the proposed system to learn patterns related to malware behavior and cyber threat indicators, which helps in identifying emerging malware trends and improving cybersecurity analysis.

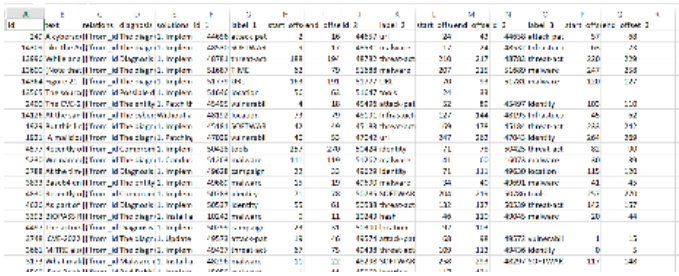

Figure 1: Dataset

V. Model Performance Evaluation

The training outcomes of the machine learning models utilised in the suggested cyber threat intelligence system are shown in the above image. Performance criteria like precision, recall, F1-score, and accuracy were used to train and assess three classification algorithms: Random Forest, Support Vector Machine, and Decision Tree. The Random Forest model outperformed the other models, as seen by its maximum accuracy of 95.52%. The Decision Tree model attained an accuracy of 92.10%, but the Support Vector Machine model earned an accuracy of roughly 90.57%. These findings show that ensemble-based techniques, such as Random Forest, are superior for spotting malware trends and analysing cyber threat intelligence data.

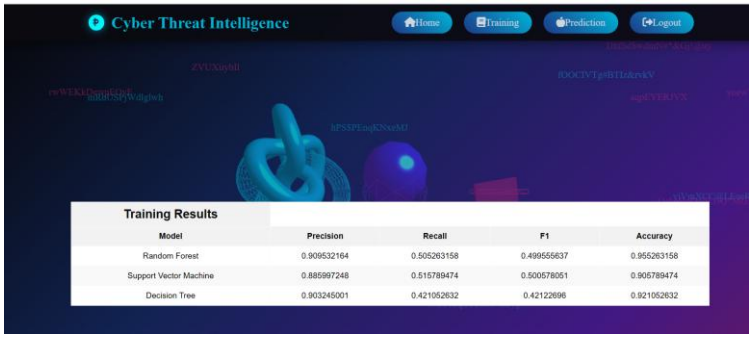

Figure 2: Machine Learning Model Training Results

VI. Conclusion

This study offers a machine learning-based method for utilising Cyber Threat Intelligence data to analyse malware trends. The suggested system analyses cyber threat data and finds dangerous patterns using classification techniques including Random Forest, Support Vector Machine, and Decision Tree. In order to prepare the dataset for model training and assessment, it underwent preprocessing.
According to the trial results, the Random Forest model outperformed the other algorithms in identifying malware-related dangers from cyber intelligence data. Additionally, the system has a web-based interface that makes it simple for users to train models and provide threat predictions.

All things considered, the suggested method enhances cyber threat analysis and aids cybersecurity experts in more effectively spotting new malware patterns. A viable way to improve contemporary cybersecurity systems is to combine machine learning with cyber threat intelligence data.

References

[1] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[2] S. Dua and X. Du, Data Mining and Machine Learning in Cybersecurity. Boca Raton, FL, USA: CRC Press, 2016.

[3] N. Sommer and V. Paxson, “Outside the closed world: On using machine learning for network intrusion detection,” in Proceedings of the IEEE Symposium on Security and Privacy, Oakland, CA, USA, 2010, pp. 305–316.

[4] M. Conti, A. Dehghantanha, K. Franke, and S. Watson, “Internet of Things security and forensics: Challenges and opportunities,” Future Generation Computer Systems, vol. 78, pp. 544–546, 2018.

[5] J. Zhang and M. Zulkernine, “Anomaly based network intrusion detection with unsupervised outlier detection,” in Proceedings of the IEEE International Conference on Communications, 2006, pp. 2388–2393.