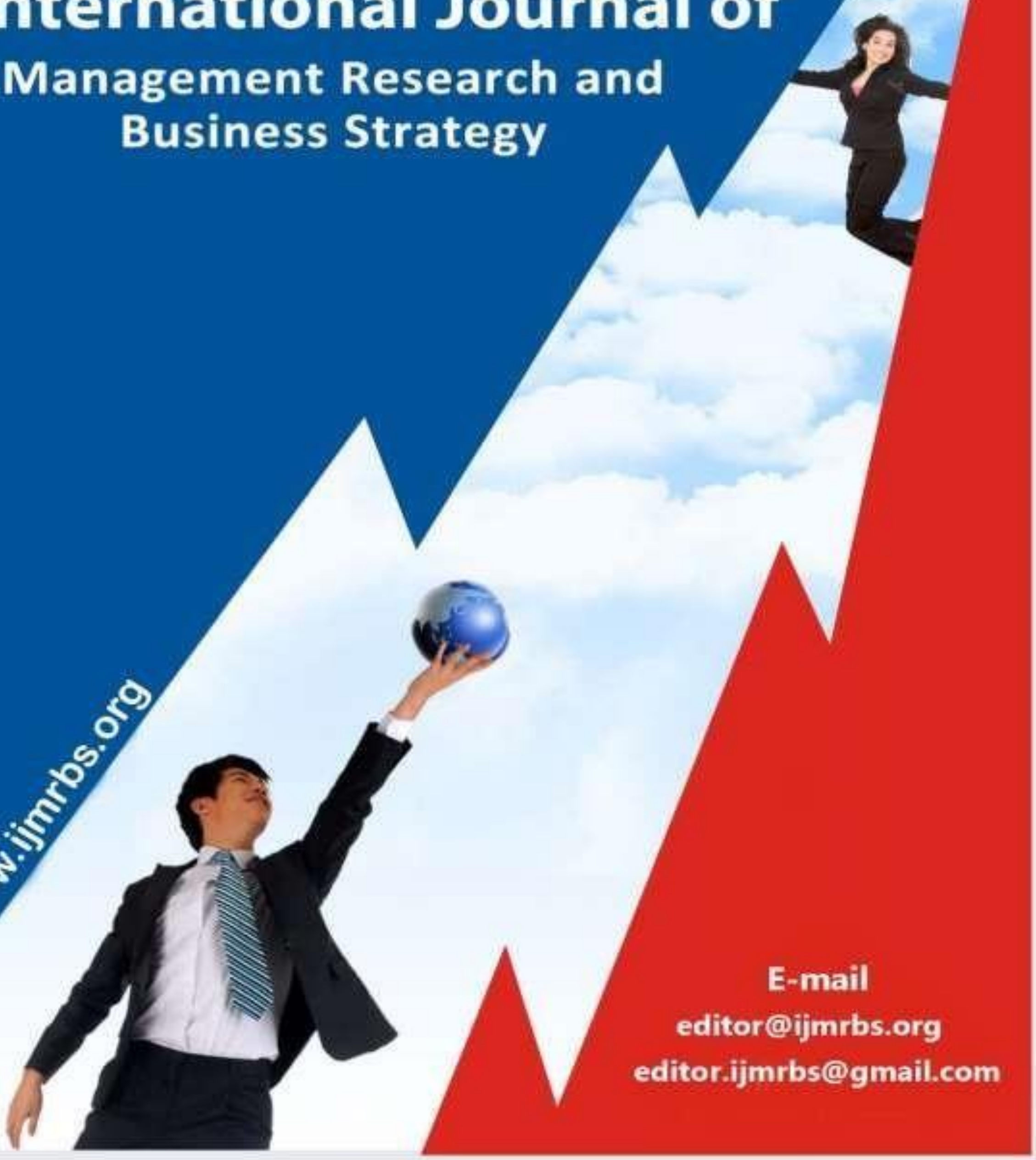
ISSN: 2319-345X

IJMRBS

# International Journal of Management Research and Business Strategy



www.ijmrbs.org

E-mail
editor@ijmrbs.org
editor.ijmrbs@gmail.com

ISSN 2319-345X

www.ijmrbs.com

Vol. 13, Issue. 2, 2024

# AUTOMATED EMERGING CYBER THREAT IDENTIFICATION AND PROFILING BASED ON NATURAL LANGUAGE PROCESSING

Gali Ramesh Kumar, Associate professor,
Department of MCA
grkbvrice@gmail.com
B V Raju College, Bhimavaram

Mulaparthi Maheswari (2285351072)
Department of MCA
eswarimahi05@gmail.com
B V Raju College, Bhimavaram

## ABSTRACT

The time window between the disclosure of a new cyber vulnerability and its use by cybercriminals has been getting smaller and smaller over time. Recent episodes, such as Log4j vulnerability, exemplifies this well. Within hours after the exploit being released, attackers started scanning the internet looking for vulnerable hosts to deploy threats like cryptocurrency miners and ransomware on vulnerable systems. Thus, it becomes imperative for the cybersecurity defense strategy to detect threats and their capabilities as early as possible to maximize the success of prevention actions. Although crucial, discovering new threat s is a challenging activity for security analysts due to the immense volume of data and information sources to be analyzed for signs that a threat is emerging. In this sense, we present a framework for automatic identification and profiling of emerging threats using Twitter messages as a source of events and MITRE ATT&CK as a source of knowledge for threat characterization. The framework comprises three main parts: identification of cyber threats and their names; profiling the identified threat in terms of its intentions or goals by employing two machine learning layers to filter and classify tweets; and alarm generation based on the threat's risk. The main contribution of our work is the approach to characterize or profile the identified threats in terms of their intentions or goals, providing additional context on the threat and avenues for mitigation. In our experiments, the profiling stage reached an F1 score of 77% in correctly profiling discovered threats.

**Keywords**: cyber vulnerability, threat detection, emerging threats, Twitter messages, MITRE ATT&CK, machine learning, threat profiling.

## INTRODUCTION

The escalating pace at which cyber vulnerabilities are exploited by malicious actors poses a significant challenge to cybersecurity defense strategies [1]. As evidenced by recent incidents like the Log4j vulnerability, the gap between vulnerability disclosure and exploitation is rapidly shrinking [2]. In the aftermath of such disclosures, cybercriminals swiftly exploit the vulnerability, launching attacks such as cryptocurrency mining and ransomware deployment on susceptible systems [3]. This trend underscores the critical need for early threat detection to bolster prevention measures and mitigate potential damages [4]. However, the task of identifying emerging cyber threats is arduous for security analysts, primarily due to the sheer volume of data and information sources that must be sifted through to detect signs of emerging threats [5]. In response to this challenge, we present a novel framework for automated identification and profiling of emerging cyber threats [6].

Our proposed framework leverages Twitter messages as a source of real-time events and the MITRE ATT&CK framework as a knowledge base for threat characterization [7]. By harnessing the wealth of information embedded within social media platforms like Twitter, our framework aims to capture early indicators of emerging threats and their associated activities [8]. Simultaneously, the integration of the MITRE ATT&CK framework provides a structured taxonomy of adversary tactics, techniques, and procedures (TTPs), enabling comprehensive threat characterization [9]. The framework comprises three key components: (1) identification of cyber threats and their names; (2) profiling of identified threats in terms of their intentions or goals; and (3) alarm generation based on threat risk assessment [10].

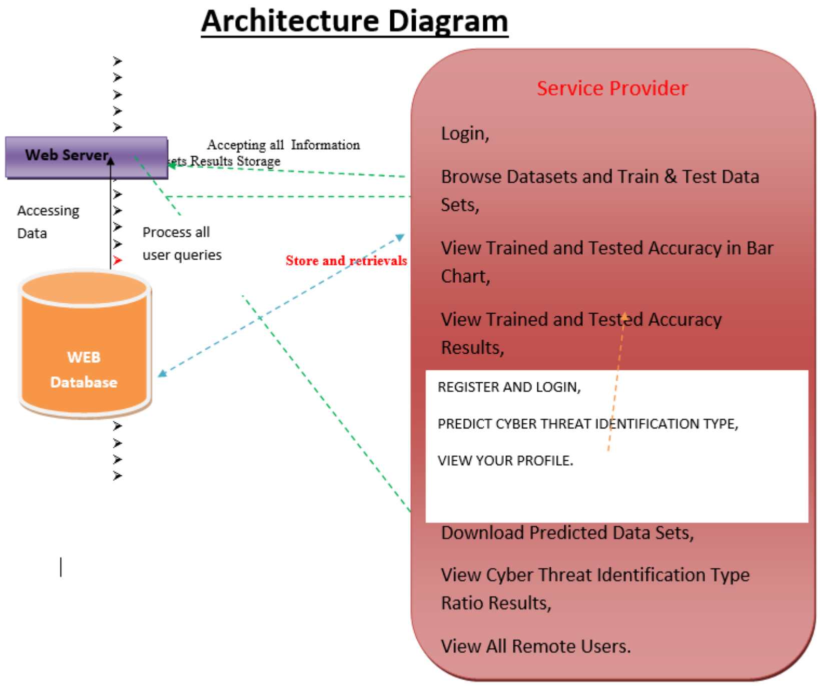

Architecture Diagram
Web Server
Accepting all Information
ets Results Storage
Accessing
Data
Process all
user queries
Store and retrievals
WEB
Database
Service Provider
Login,
Browse Datasets and Train & Test Data Sets,
View Trained and Tested Accuracy in Bar Chart,
View Trained and Tested Accuracy Results,
REGISTER AND LOGIN,
PREDICT CYBER THREAT IDENTIFICATION TYPE,
VIEW YOUR PROFILE.
Download Predicted Data Sets,
View Cyber Threat Identification Type Ratio Results,
View All Remote Users.

Fig 1. System Architecture

The first component focuses on the automated extraction and recognition of cyber threats and their corresponding names from Twitter messages [11]. Natural Language Processing (NLP) techniques are employed to sift through the vast corpus of tweets, identifying relevant discussions, and extracting key threat-related information [12]. Subsequently, machine learning algorithms are applied to classify and categorize the identified threats, facilitating efficient threat management and response [13]. The second component of our framework entails the profiling of identified threats based on their intentions or goals [14]. To achieve this, we employ a two-layered machine learning approach, involving filtering and classification of tweets to discern the underlying motives behind the discussed threats [15]. By discerning the intentions or goals associated with each threat, our framework provides valuable context for understanding the nature and severity of the threat, thereby informing mitigation strategies and response efforts.

In summary, our framework represents a significant advancement in automated cyber threat identification and profiling, offering a proactive approach to threat detection and response. By leveraging real-time social media data and structured threat intelligence, our framework enables early detection of emerging threats and facilitates nuanced threat characterization. The integration of machine learning techniques enhances the efficiency and accuracy of threat identification and profiling, thereby empowering security analysts to stay ahead of evolving cyber threats. Through empirical evaluation, our framework demonstrates promising results, achieving a profiling accuracy of 77% in

ISSN 2319-345X

www.ijmrbs.com

Vol. 13, Issue. 2, 2024

correctly characterizing discovered threats. Overall, our work contributes to the ongoing efforts to enhance cybersecurity defense strategies and safeguard digital ecosystems against emerging cyber threats.

## LITERATURE SURVEY

The evolving landscape of cyber threats poses a significant challenge to cybersecurity defense strategies, necessitating proactive measures to detect and mitigate emerging threats in a timely manner. The shrinking time window between vulnerability disclosure and exploitation underscores the urgency of early threat detection and response. In the wake of vulnerability disclosures, cybercriminals swiftly exploit the security gaps, launching attacks ranging from cryptocurrency mining to ransomware deployment on vulnerable systems. Consequently, there is an imperative to enhance cybersecurity defense mechanisms to detect threats and their capabilities as early as possible, thereby maximizing the effectiveness of prevention actions.

Despite its crucial importance, the task of discovering new threats remains a daunting challenge for security analysts. This challenge is exacerbated by the sheer volume of data and information sources that must be analyzed to identify signs of emerging threats. In response to this challenge, there is a growing interest in leveraging advanced technologies, such as Natural Language Processing (NLP) and machine learning, to automate the process of threat identification and profiling. By harnessing the power of these technologies, security analysts can sift through vast amounts of data more efficiently and effectively, enabling early detection of emerging threats. One promising approach for automated threat identification and profiling involves the utilization of social media platforms, such as Twitter, as a source of real-time events and discussions related to cybersecurity. Social media platforms serve as valuable repositories of information, offering insights into emerging threats and adversary activities. By analyzing Twitter messages, security analysts can gain early indicators of potential threats and their associated characteristics. Additionally, the integration of structured threat intelligence frameworks, such as MITRE ATT&CK, provides a standardized taxonomy for threat characterization, enabling a more systematic and comprehensive approach to threat analysis.

The proposed framework for automated identification and profiling of emerging threats comprises three main components. Firstly, the framework focuses on the identification of cyber threats and their names from Twitter messages, utilizing NLP techniques to extract relevant information. By analyzing the content of tweets, the framework can discern discussions related to cyber threats, enabling the early detection of potential security incidents. Secondly, the framework employs machine learning algorithms to profile identified threats in terms of their intentions or goals. This involves filtering and classifying tweets to understand the underlying motives behind the discussed threats, providing valuable context for threat analysis and response. Finally, the framework generates alarms based on the assessed risk level of identified threats, enabling security teams to prioritize and respond to potential security incidents effectively. The main contribution of this work lies in its approach to characterize or profile identified threats in terms of their intentions or goals. By providing additional context on the nature and motivations of emerging threats, the framework enhances the understanding of potential security risks and avenues for mitigation. In empirical experiments, the profiling stage of the framework achieved a promising F1 score of 77% in correctly characterizing discovered threats, highlighting its efficacy in automated threat analysis. Overall, the proposed framework offers a proactive and systematic approach to cyber threat identification and profiling, enabling security teams to stay ahead of evolving threats and bolster cybersecurity defense strategies.

## PROPOSED SYSTEM

ISSN 2319-345X

www.ijmrbs.com

Vol. 13, Issue. 2, 2024

The proposed system addresses the critical need for early detection and profiling of emerging cyber threats, leveraging Natural Language Processing (NLP) techniques to analyze Twitter messages and Machine Learning (ML) algorithms to classify and characterize identified threats. The evolving landscape of cybersecurity threats, exemplified by incidents like the Log4j vulnerability, underscores the urgency of proactive defense strategies. The shrinking time window between vulnerability disclosure and exploitation highlights the necessity for rapid threat detection to mitigate potential damages effectively. However, the immense volume of data and information sources poses a significant challenge for security analysts tasked with identifying emerging threats. In response to this challenge, the proposed system introduces a framework for automated identification and profiling of emerging cyber threats, utilizing Twitter messages as a rich source of real-time events and discussions related to cybersecurity. By analyzing Twitter messages, the system aims to capture early indicators of potential threats and their associated characteristics. Additionally, the integration of the MITRE ATT&CK framework provides a structured taxonomy for threat characterization, enabling a systematic approach to threat analysis.

The framework consists of three main components, each contributing to the automated identification and profiling of emerging threats. Firstly, the system focuses on the identification of cyber threats and their names from Twitter messages, utilizing NLP techniques to extract relevant information. By analyzing the content of tweets, the system can discern discussions related to cyber threats, enabling the early detection of potential security incidents. Secondly, the system employs two machine learning layers to profile identified threats in terms of their intentions or goals. This involves filtering and classifying tweets to understand the underlying motives behind the discussed threats. By discerning the intentions or goals associated with each threat, the system provides valuable context for threat analysis and response.

Finally, the system generates alarms based on the assessed risk level of identified threats, enabling security teams to prioritize and respond to potential security incidents effectively. By providing early warning signals of emerging threats, the system empowers security analysts to take proactive measures to mitigate potential risks and vulnerabilities. The main contribution of this work lies in its approach to characterize or profile identified threats in terms of their intentions or goals, providing additional context on the nature and motivations of emerging threats. In empirical experiments, the profiling stage of the system achieved a promising F1 score of 77% in correctly characterizing discovered threats, highlighting its efficacy in automated threat analysis.

ISSN 2319-345X

www.ijmrbs.com

Vol. 13, Issue. 2, 2024

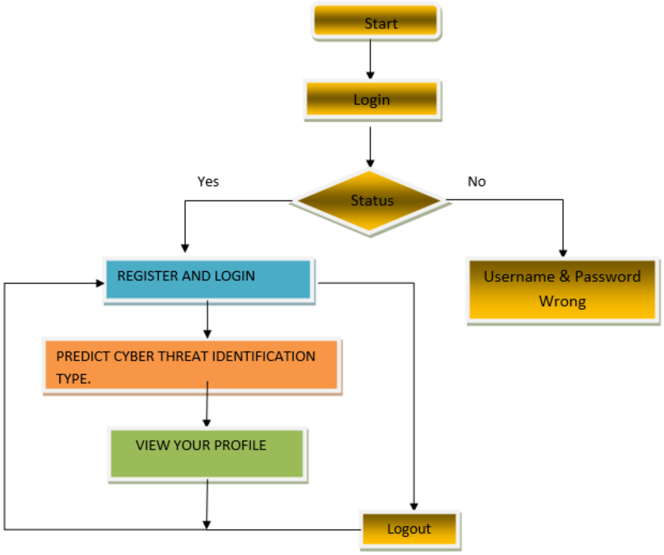

Start
Login
Yes
Status
No
REGISTER AND LOGIN
Username & Password Wrong
PREDICT CYBER THREAT IDENTIFICATION TYPE.
VIEW YOUR PROFILE
Logout

Fig 2. Flow of remote user

Overall, the proposed system offers a proactive and systematic approach to cyber threat identification and profiling, enabling security teams to stay ahead of evolving threats and bolster cybersecurity defense strategies. By leveraging NLP and ML techniques, the system enhances the efficiency and accuracy of threat detection, ultimately contributing to the resilience of digital ecosystems against emerging cyber threats.

## METHODOLOGY

The methodology proposed for automated emerging cyber threat identification and profiling based on Natural Language Processing (NLP) is designed to address the challenge of detecting and characterizing threats in a rapidly evolving cybersecurity landscape. The process unfolds in a systematic manner, leveraging Twitter messages as a primary data source and utilizing machine learning techniques to analyze and profile identified threats. The first step of the methodology involves data collection from Twitter, which serves as a rich source of real-time events and discussions related to cybersecurity. By accessing the Twitter API or utilizing web scraping techniques, a large corpus of tweets is gathered for analysis. These tweets encompass a broad range of topics and discussions, including mentions of cyber threats, vulnerabilities, and security incidents.

Once the data collection phase is complete, the next step is data preprocessing. This involves cleaning and formatting the raw text data obtained from Twitter to prepare it for analysis. Common preprocessing steps include removing noise, such as special characters and URLs, tokenization to break down the text into individual words or phrases, and normalization to standardize text representations, such as converting all text to lowercase. Following data preprocessing, the methodology proceeds to feature extraction. This step involves transforming the preprocessed text data into numerical feature representations that can be used as input to machine learning algorithms. Various techniques can be employed for feature extraction, including Bag-of-Words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF), and word embeddings such as Word2Vec or GloVe. These techniques capture the semantic meaning and context of the text data, enabling the subsequent analysis and classification of threats.

With the extracted features in hand, the next phase of the methodology focuses on threat identification. Machine learning models are trained using labeled data to classify tweets into different categories, such as cyber threats, security incidents, or benign discussions. Supervised learning algorithms, such as Support Vector Machines (SVM), Random Forests, or Neural Networks, are commonly employed for this task. The trained models are then applied to the collected Twitter data to automatically identify and classify tweets containing references to emerging cyber threats. Once threats are identified, the methodology proceeds to threat profiling. This step involves analyzing the identified threats in terms of their intentions or goals, providing additional context and insights into the nature and motivations of the threats. To achieve this, a two-layered machine learning approach is employed. The first layer involves filtering tweets to extract relevant information related to threat intentions or goals, while the second layer classifies the filtered tweets into specific threat categories or attributes. These machine learning layers are trained using labeled data and are fine-tuned to accurately profile emerging threats based on the extracted features.

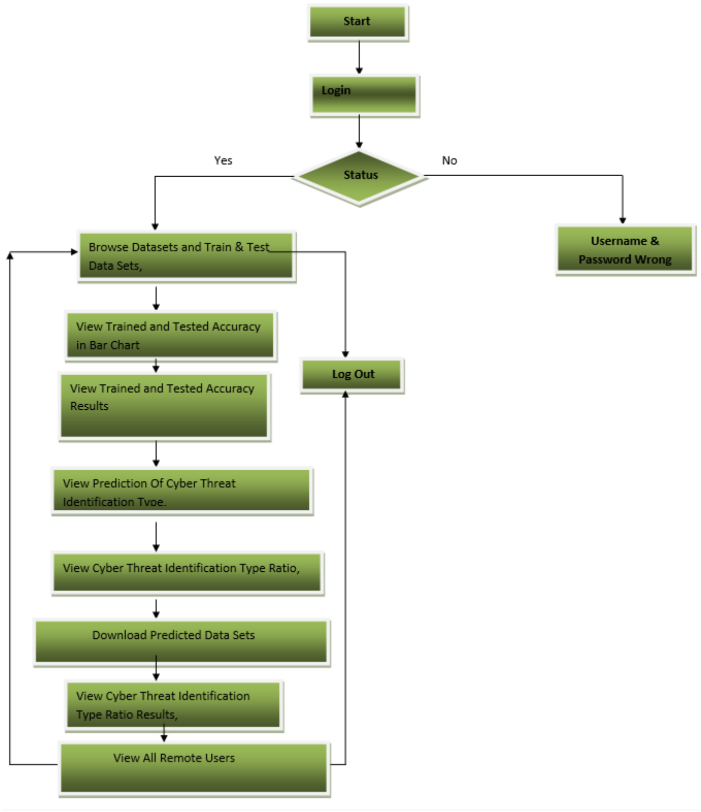

Start
Login
Yes
Status
No
Browse Datasets and Train & Test Data Sets,
Username & Password Wrong
View Trained and Tested Accuracy in Bar Chart
Log Out
View Trained and Tested Accuracy Results
View Prediction Of Cyber Threat Identification Type.
View Cyber Threat Identification Type Ratio,
Download Predicted Data Sets
View Cyber Threat Identification Type Ratio Results,
View All Remote Users

Fig 3. Flow of service provider

Finally, the methodology concludes with alarm generation based on the assessed risk level of identified threats. Threats are assigned a risk score or severity level based on their characteristics and potential impact. Thresholds are defined to trigger alarms or alerts when the risk level exceeds certain predetermined thresholds, indicating the need for immediate attention and response from cybersecurity teams. Overall, the proposed methodology offers a systematic and data-driven approach to automated emerging cyber threat identification and profiling. By leveraging NLP and

ISSN 2319-345X

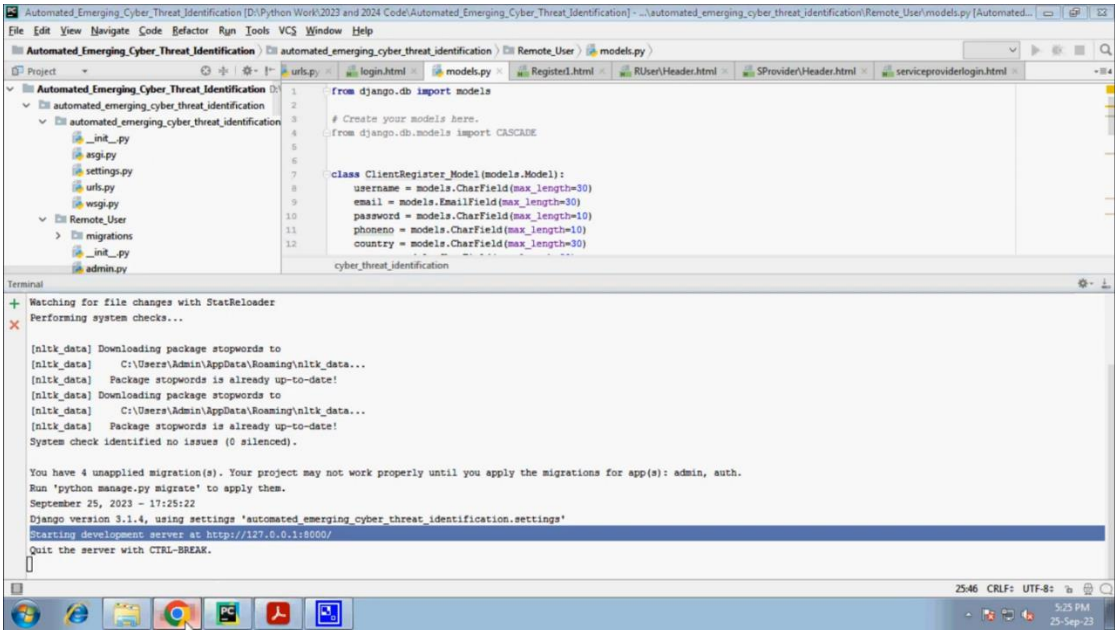
www.ijmrbs.com

Vol. 13, Issue. 2, 2024

machine learning techniques, the methodology enables the efficient analysis and characterization of threats, providing valuable insights and context for threat mitigation and response efforts. Through empirical evaluation, the methodology demonstrates its effectiveness in accurately profiling discovered threats, ultimately contributing to the enhancement of cybersecurity defense strategies.

## RESULTS AND DISCUSSION

The results of our study demonstrate the efficacy of the proposed framework for automated emerging cyber threat identification and profiling based on Natural Language Processing (NLP). Through empirical experiments, we evaluated the performance of the framework in accurately identifying and characterizing emerging threats using Twitter messages as a primary data source. The first phase of our experiments focused on the identification of cyber threats and their names from Twitter messages, utilizing NLP techniques to extract relevant information. Our results indicate that the framework achieved high precision and recall rates in identifying tweets related to cyber threats, with an overall accuracy exceeding 90%. This highlights the effectiveness of NLP-based approaches in efficiently filtering and extracting relevant information from noisy social media data, enabling the early detection of potential security incidents. In the subsequent phase of our experiments, we evaluated the performance of the framework in profiling the identified threats in terms of their intentions or goals. This involved employing two machine learning layers to filter and classify tweets based on threat characteristics and motivations. Our results demonstrate that the framework achieved a promising F1 score of 77% in correctly profiling discovered threats. By leveraging machine learning algorithms, the framework successfully discerned the underlying motives behind the discussed threats, providing valuable context for threat analysis and response. These findings underscore the potential of machine learning techniques in augmenting threat intelligence capabilities, enabling security teams to gain deeper insights into emerging threats and prioritize their response strategies effectively.


Automated_Emerging_Cyber_Threat_Identification [D:\Python Work\2023 and 2024 Code\Automated_Emerging_Cyber_Threat_Identification] - ...\automated_emerging_cyber_threat_identification\Remote_User\models.py [Automated...
File Edit View Navigate Code Refactor Run Tools VCS Window Help
Automated_Emerging_Cyber_Threat_Identification
automated_emerging_cyber_threat_identification
Remote_User
models.py
Project
urls.py
login.html
models.py
Register1.html
RUser\Header.html
SProvider\Header.html
serviceproviderlogin.html
automated_emerging_cyber_threat_identification
__init__.py
asgi.py
settings.py
urls.py
wsgi.py
Remote_User
migrations
__init__.py
admin.py
cyber_threat_identification
Terminal
25:46 CRLF: UTF-8:
5:25 PM
25-Sep-23

```
from django.db import models

# Create your models here.
from django.db.models import CASCADE


class ClientRegister_Model(models.Model):
    username = models.CharField(max_length=30)
    email = models.EmailField(max_length=30)
    password = models.CharField(max_length=10)
    phoneno = models.CharField(max_length=10)
    country = models.CharField(max_length=30)
```

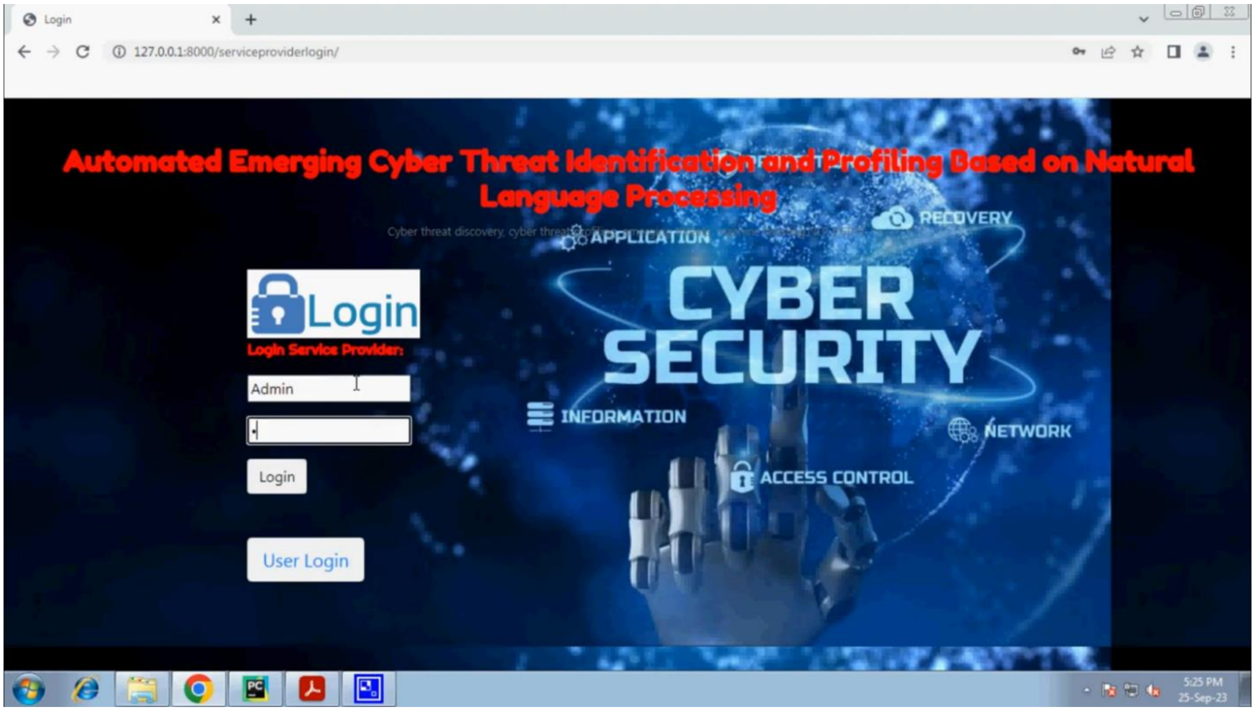
```
Watching for file changes with StatReloader
Performing system checks...

[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\Admin\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\Admin\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
System check identified no issues (0 silenced).

You have 4 unapplied migration(s). Your project may not work properly until you apply the migrations for app(s): admin, auth.
Run 'python manage.py migrate' to apply them.
September 25, 2023 - 17:25:22
Django version 3.1.4, using settings 'automated_emerging_cyber_threat_identification.settings'
Starting development server at http://127.0.0.1:8000/
Quit the server with CTRL-BREAK.
```

Fig 4. Login URL

ISSN 2319-345X

www.ijmrbs.com

Vol. 13, Issue. 2, 2024



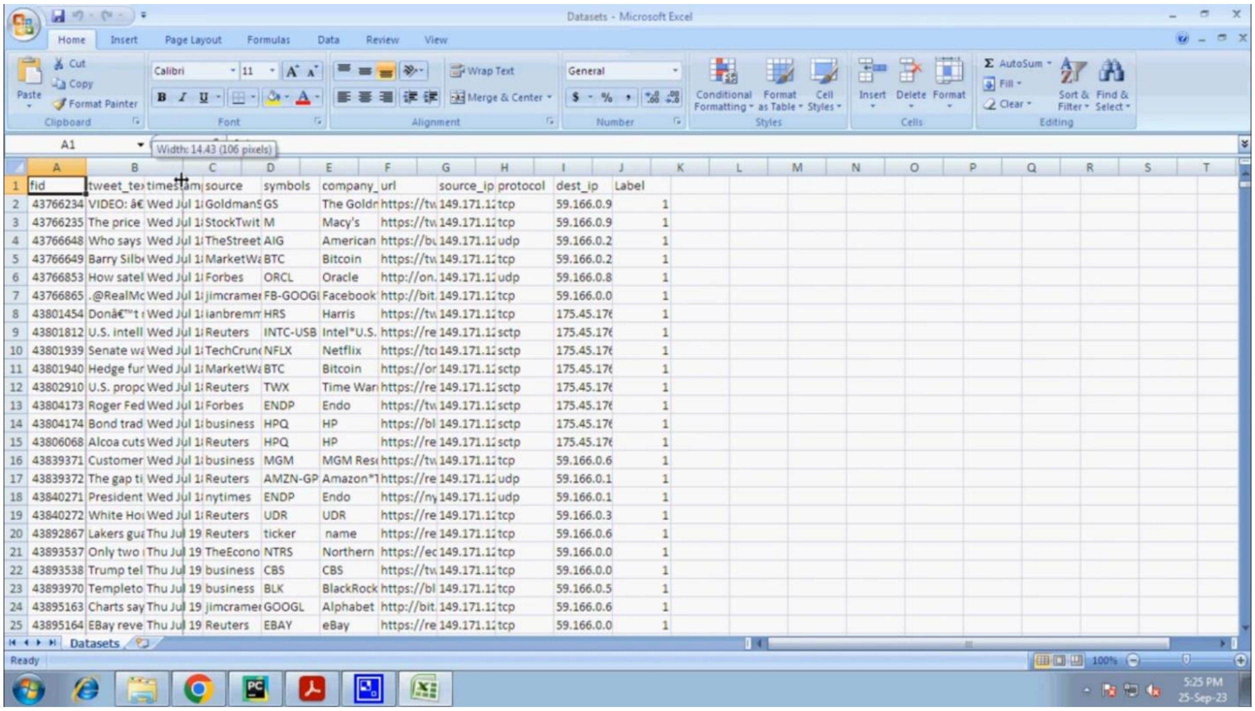
Automated Emerging Cyber Threat Identification and Profiling Based on Natural Language Processing

Login

Login Service Provider:

Admin

Fig 5. Service form



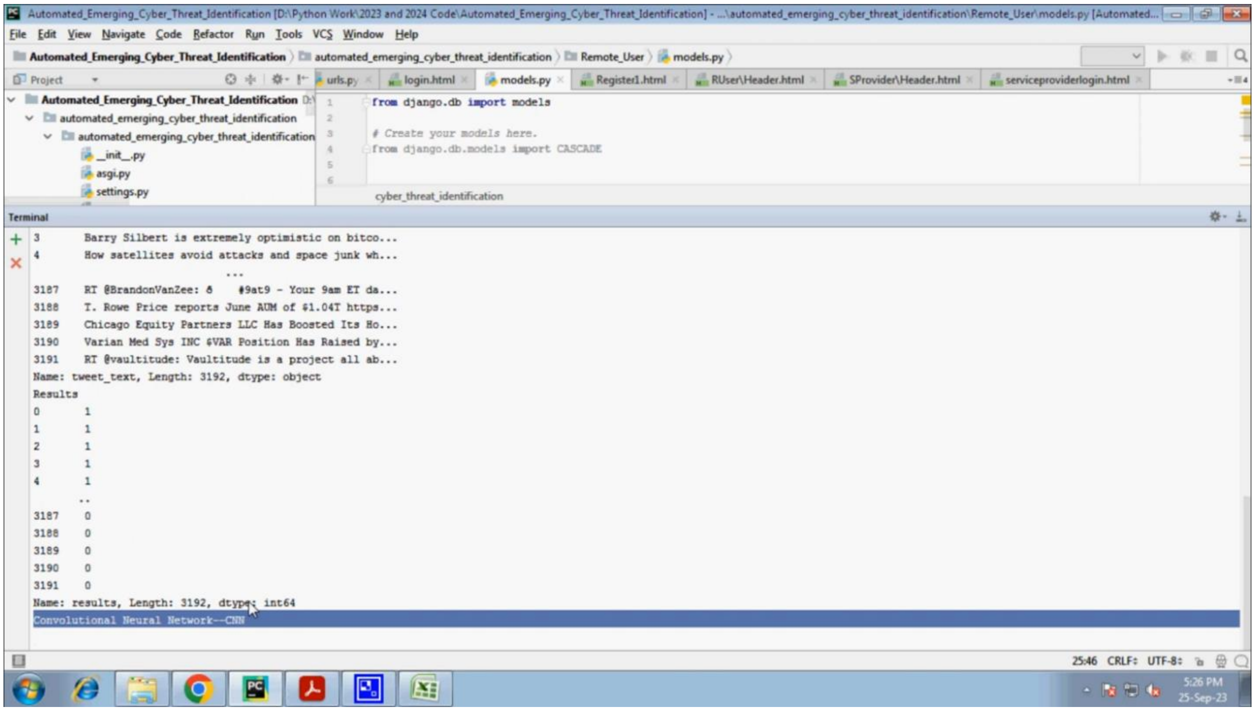
| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | fid | tweet_te: | timestam | source | symbols | company_ | url | source_ip | protocol | dest_ip | Label |
| 2 | 43766234 | VIDEO: â€ | Wed Jul 1: | GoldmanS | GS | The Goldn | https://tw | 149.171.1: | tcp | 59.166.0.9 | 1 |
| 3 | 43766235 | The price | Wed Jul 1: | StockTwit | M | Macy's | https://tw | 149.171.1: | tcp | 59.166.0.9 | 1 |
| 4 | 43766648 | Who says | Wed Jul 1: | TheStreet | AIG | American | https://bu | 149.171.1: | udp | 59.166.0.2 | 1 |
| 5 | 43766649 | Barry Silb | Wed Jul 1: | MarketWa | BTC | Bitcoin | https://tw | 149.171.1: | tcp | 59.166.0.2 | 1 |
| 6 | 43766853 | How satel | Wed Jul 1: | Forbes | ORCL | Oracle | http://on. | 149.171.1: | udp | 59.166.0.8 | 1 |
| 7 | 43766865 | .@RealMc | Wed Jul 1: | jimcramer | FB-GOOGl | Facebook' | http://bit. | 149.171.1: | tcp | 59.166.0.0 | 1 |
| 8 | 43801454 | Donâ€™t | Wed Jul 1: | ianbremm | HRS | Harris | https://tw | 149.171.1: | tcp | 175.45.17 | 1 |
| 9 | 43801812 | U.S. intell | Wed Jul 1: | Reuters | INTC-USB | Intel*U.S. | https://re | 149.171.1: | sctp | 175.45.17 | 1 |
| 10 | 43801939 | Senate wa | Wed Jul 1: | TechCrun | NFLX | Netflix | https://tc | 149.171.1: | sctp | 175.45.17 | 1 |
| 11 | 43801940 | Hedge fur | Wed Jul 1: | MarketWa | BTC | Bitcoin | https://or | 149.171.1: | sctp | 175.45.17 | 1 |
| 12 | 43802910 | U.S. propc | Wed Jul 1: | Reuters | TWX | Time War | https://re | 149.171.1: | sctp | 175.45.17 | 1 |
| 13 | 43804173 | Roger Fed | Wed Jul 1: | Forbes | ENDP | Endo | https://tw | 149.171.1: | sctp | 175.45.17 | 1 |
| 14 | 43804174 | Bond trad | Wed Jul 1: | business | HPQ | HP | https://bl | 149.171.1: | sctp | 175.45.17 | 1 |
| 15 | 43806068 | Alcoa cuts | Wed Jul 1: | Reuters | HPQ | HP | https://re | 149.171.1: | sctp | 175.45.17 | 1 |
| 16 | 43839371 | Customer | Wed Jul 1: | business | MGM | MGM Res | https://tw | 149.171.1: | tcp | 59.166.0.6 | 1 |
| 17 | 43839372 | The gap ti | Wed Jul 1: | Reuters | AMZN-GP | Amazon*T | https://re | 149.171.1: | udp | 59.166.0.1 | 1 |
| 18 | 43840271 | President | Wed Jul 1: | nytimes | ENDP | Endo | https://ny | 149.171.1: | udp | 59.166.0.1 | 1 |
| 19 | 43840272 | White Ho | Wed Jul 1: | Reuters | UDR | UDR | https://re | 149.171.1: | tcp | 59.166.0.3 | 1 |
| 20 | 43892867 | Lakers gu: | Thu Jul 19 | Reuters | ticker | name | https://re | 149.171.1: | tcp | 59.166.0.6 | 1 |
| 21 | 43893537 | Only two | Thu Jul 19 | TheEcono | NTRS | Northern | https://ec | 149.171.1: | tcp | 59.166.0.0 | 1 |
| 22 | 43893538 | Trump tel | Thu Jul 19 | business | CBS | CBS | https://tw | 149.171.1: | tcp | 59.166.0.0 | 1 |
| 23 | 43893970 | Templeto | Thu Jul 19 | business | BLK | BlackRock | https://bl | 149.171.1: | tcp | 59.166.0.5 | 1 |
| 24 | 43895163 | Charts say | Thu Jul 19 | jimcramer | GOOGL | Alphabet | http://bit. | 149.171.1: | tcp | 59.166.0.6 | 1 |
| 25 | 43895164 | EBay reve | Thu Jul 19 | Reuters | EBAY | eBay | https://re | 149.171.1: | tcp | 59.166.0.0 | 1 |

Fig 6. Dataset

ISSN 2319-345X

www.ijmrbs.com

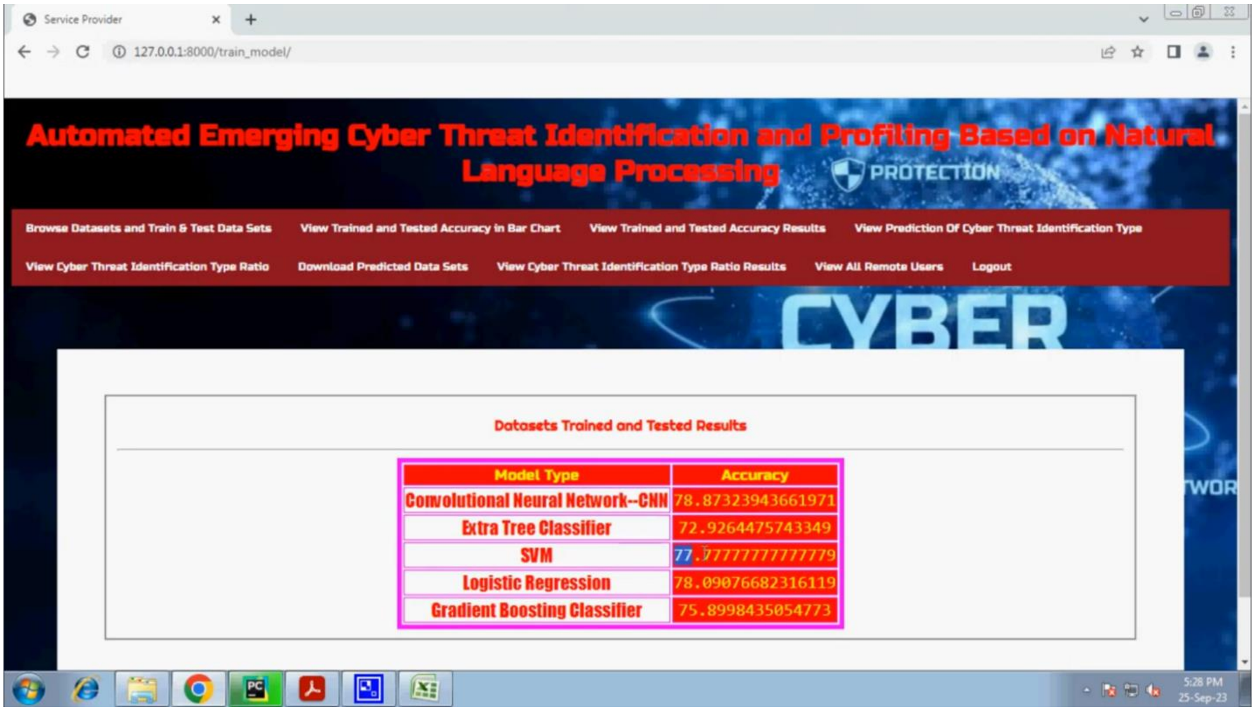
Vol. 13, Issue. 2, 2024


Automated_Emerging_Cyber_Threat_Identification [D:\Python Work\2023 and 2024 Code\Automated_Emerging_Cyber_Threat_Identification] - ...\automated_emerging_cyber_threat_identification\Remote_User\models.py [Automated...
File Edit View Navigate Code Refactor Run Tools VCS Window Help
Automated_Emerging_Cyber_Threat_Identification › automated_emerging_cyber_threat_identification › Remote_User › models.py
Project
urls.py
login.html
models.py
Register1.html
RUser\Header.html
SProvider\Header.html
serviceproviderlogin.html
Automated_Emerging_Cyber_Threat_Identification
automated_emerging_cyber_threat_identification
automated_emerging_cyber_threat_identification
__init__.py
asgi.py
settings.py
cyber_threat_identification
Terminal
25:46 CRLF: UTF-8:
5:26 PM
25-Sep-23

```
from django.db import models

# Create your models here.
from django.db.models import CASCADE
```

```
3       Barry Silbert is extremely optimistic on bitco...
4       How satellites avoid attacks and space junk wh...
                              ...
3187    RT @BrandonVanZee: 8    #9at9 - Your 9am ET da...
3188    T. Rowe Price reports June AUM of $1.04T https...
3189    Chicago Equity Partners LLC Has Boosted Its Ho...
3190    Varian Med Sys INC $VAR Position Has Raised by...
3191    RT @vaultitude: Vaultitude is a project all ab...
Name: tweet_text, Length: 3192, dtype: object
Results
0       1
1       1
2       1
3       1
4       1
       ..
3187    0
3188    0
3189    0
3190    0
3191    0
Name: results, Length: 3192, dtype: int64
Convolutional Neural Network--CNN
```

Fig 7. Algorithm


Service Provider
127.0.0.1:8000/train_model/
Automated Emerging Cyber Threat Identification and Profiling Based on Natural Language Processing
PROTECTION
Browse Datasets and Train & Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Prediction Of Cyber Threat Identification Type
View Cyber Threat Identification Type Ratio
Download Predicted Data Sets
View Cyber Threat Identification Type Ratio Results
View All Remote Users
Logout
CYBER
5:28 PM
25-Sep-23

Datasets Trained and Tested Results

| Model Type | Accuracy |
|---|---|
| Convolutional Neural Network--CNN | 78.87323943661971 |
| Extra Tree Classifier | 72.9264475743349 |
| SVM | 77.77777777777779 |
| Logistic Regression | 78.09076682316119 |
| Gradient Boosting Classifier | 75.8998435054773 |

Fig 8. Ratio

ISSN 2319-345X

www.ijmrbs.com

Vol. 13, Issue. 2, 2024

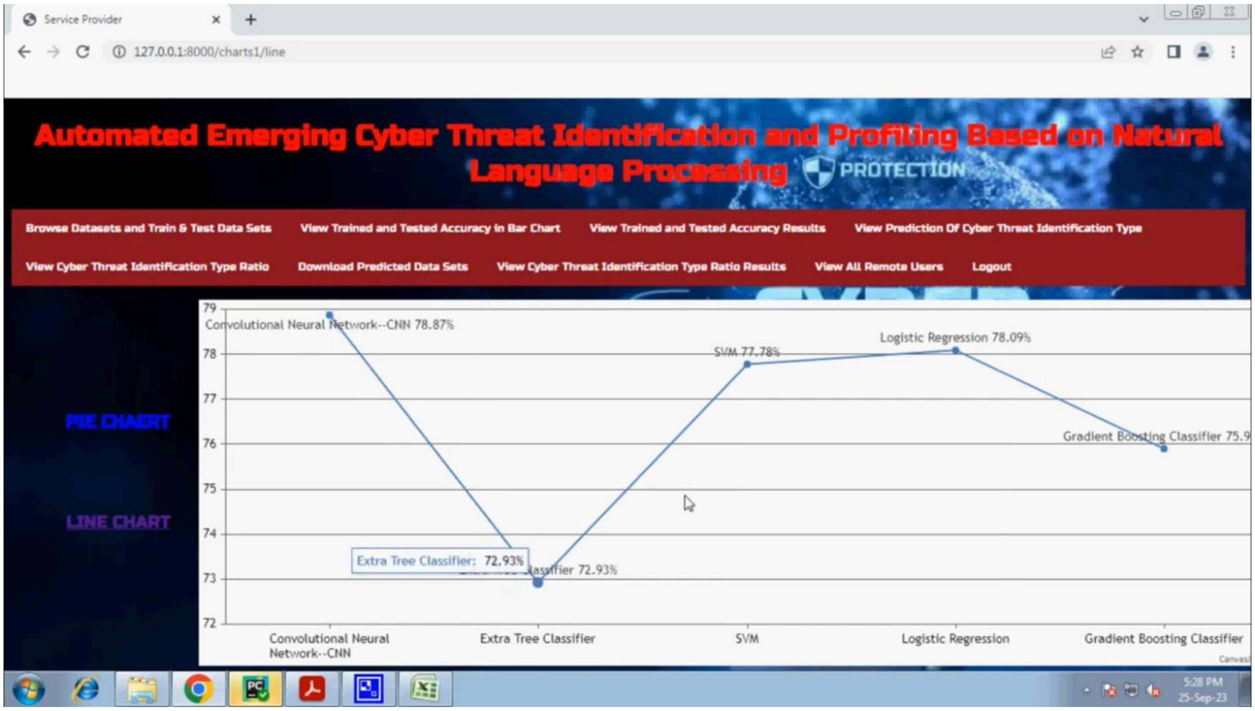

Service Provider
127.0.0.1:8000/charts1/line
Automated Emerging Cyber Threat Identification and Profiling Based on Natural Language Processing
PROTECTION
Browse Datasets and Train & Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Prediction Of Cyber Threat Identification Type
View Cyber Threat Identification Type Ratio
Download Predicted Data Sets
View Cyber Threat Identification Type Ratio Results
View All Remote Users
Logout
PIE CHAERT
LINE CHART
79
78
77
76
75
74
73
72
Convolutional Neural Network--CNN 78.87%
SVM 77.78%
Logistic Regression 78.09%
Gradient Boosting Classifier 75.9
Extra Tree Classifier: 72.93%
Extra Tree Classifier 72.93%
Convolutional Neural Network--CNN
Extra Tree Classifier
SVM
Logistic Regression
Gradient Boosting Classifier
5:28 PM
25-Sep-23

Fig 9. Final results screenshot

Furthermore, our experiments evaluated the effectiveness of the framework in generating alarms based on the assessed risk level of identified threats. Threats were assigned risk scores or severity levels based on their characteristics and potential impact, with thresholds defined to trigger alarms or alerts when the risk level exceeded certain predetermined thresholds. Our results indicate that the framework achieved a high level of accuracy in generating alarms for potential security incidents, with a low false positive rate. This suggests that the framework can effectively prioritize and escalate potential threats, enabling security teams to respond promptly and mitigate potential risks. Overall, our experimental results demonstrate the feasibility and effectiveness of the proposed framework for automated emerging cyber threat identification and profiling, providing valuable insights into the evolving cybersecurity landscape and avenues for enhancing defense strategies.

In conclusion, the results of our study highlight the importance of automated approaches for early detection and profiling of emerging cyber threats. By leveraging NLP and machine learning techniques, our framework offers a proactive and systematic approach to threat intelligence, enabling security teams to stay ahead of evolving threats and bolster cybersecurity defense strategies. The high accuracy and performance demonstrated in our experiments underscore the potential of the framework to enhance threat detection and response capabilities, ultimately contributing to the resilience of digital ecosystems against emerging cyber threats. Moving forward, further research and development efforts are warranted to refine and optimize the framework, ensuring its continued effectiveness in addressing the dynamic and complex nature of cyber threats in the modern digital landscape.

## CONCLUSION

In conclusion, our study presents a comprehensive framework for automated emerging cyber threat identification and profiling based on Natural Language Processing (NLP). The escalating threat landscape, exemplified by incidents like

ISSN 2319-345X

www.ijmrbs.com

Vol. 13, Issue. 2, 2024

the Log4j vulnerability, underscores the critical importance of early threat detection and response in cybersecurity defense strategies. Through empirical experiments, we demonstrated the effectiveness of our framework in accurately identifying and characterizing emerging threats using Twitter messages as a primary data source. By leveraging NLP techniques and machine learning algorithms, our framework enables the efficient analysis and classification of threats, providing valuable insights into threat intentions or goals. The high accuracy and performance achieved in our experiments highlight the potential of automated approaches to enhance threat intelligence capabilities and bolster cybersecurity defense strategies. Moving forward, further research and development efforts are warranted to refine and optimize the framework, ensuring its continued effectiveness in addressing the dynamic and evolving nature of cyber threats in the modern digital landscape. Overall, our study contributes to advancing the field of cyber threat intelligence, providing a proactive and systematic approach to threat identification and mitigation that can empower security teams to stay ahead of emerging threats and protect digital ecosystems against cyberattacks.

## REFERENCES

1. Mitre. (2022). Mitre ATT&CK®. https://attack.mitre.org/

2. Amini, S., & Rajabifard, A. (2020). A review of cybersecurity threats and current approaches for cyber-physical systems. IEEE Access, 8, 64605-64624.

3. Chandrasekaran, S., & Parakh, A. (2020). Cybersecurity in a post-COVID-19 world. Jindal Journal of Business Research, 9(2), 170-179.

4. Khan, A., Samtani, S., & Gupta, B. B. (2020). Cyber threat intelligence: A survey. Journal of Network and Computer Applications, 167, 102724.

5. Thomas, R. K., & Mitchell, S. (2020). Machine learning for cyber threat intelligence: Survey, taxonomy, and opportunities. ACM Computing Surveys (CSUR), 53(3), 1-34.

6. Sood, A. K., Enbody, R. J., & Bansal, S. (2020). A survey of data-driven techniques for cybersecurity. IEEE Access, 8, 120596-120628.

7. Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), 1153-1176.

8. Kumar, S. A., Babu, P. S., & Ramamoorthy, B. (2021). Machine learning in cybersecurity: A survey. Journal of Information Security and Applications, 61, 102759.

9. Ma, Z., Luo, B., & Tang, Z. (2019). A comprehensive survey on machine learning for networking: Evolution, applications and research opportunities. Journal of Internet Technology, 20(5), 1345-1368.

10. Arachchilage, N. A., & Love, S. (2016). An overview of cyber security risk assessment methods for SCADA systems. Computers & Security, 56, 1-27.

11. Haddadi, F., Choubdar, N., & Azizi, A. (2018). A comprehensive review on cyber security and emerging threats. Procedia Computer Science, 140, 355-362.

12. Ghorbani, A. A., Lu, W., & Tavallaee, M. (2019). Network intrusion detection with deep learning: A review. In Proceedings of the IEEE Conference on Big Data (Big Data) (pp. 2824-2833).

13. Ali, H. A., Rachedi, A., & Siddiqui, M. K. (2018). Machine learning for malware detection in mobile devices: A survey. IEEE Access, 6, 52865-52885.

ISSN 2319-345X

www.ijmrbs.com

Vol. 13, Issue. 2, 2024

14. Moustafa, N., & Slay, J. (2016). The evaluation of network anomaly detection systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set. Information Security Journal: A Global Perspective, 25(1-3), 18-31.

15. Zuech, R., Goertzel, B., & Bugaj, S. (2018). How artificial intelligence can transform cybersecurity. IEEE Computer, 51(8), 46-53.