International Journal of Early Childhood Special Education (INT-JECSE)
DOI:10.48047/intjecse/v13i2.21279 ISSN: 1308-5581 Vol 13, Issue 02 2021

MACHINE LEARNING-DRIVEN CYBER ATTACK DETECTION IN NETWORK ENVIRONMENTS

[1]Shravani.Velpula,[2]Vijayalaxmi.K,[3]Thota Mounika,[4]THIPIRI NITHISH
[123]Assistant Professor,[4]Students
Department of CSD
Vaagdevi College of Engineering, Warangal, Telangana

ABSTRACT
This paper presents a novel approach to detecting cyber attacks in network environments using advanced machine learning techniques. With the increasing sophistication of cyber threats, traditional security measures often fall short in identifying and mitigating attacks in real time. Our study employs a range of machine learning algorithms, including Decision Trees, Support Vector Machines, and Neural Networks, to analyze network traffic data and identify patterns indicative of malicious activities. By leveraging a comprehensive dataset of labeled network traffic, we train and validate our models to enhance detection accuracy and reduce false positives. The results demonstrate that machine learning techniques significantly outperform conventional methods in identifying various types of cyber attacks, such as denial of service (DoS), intrusion attempts, and malware propagation. This research contributes to the development of proactive cybersecurity strategies, enabling organizations to enhance their network defenses and respond swiftly to emerging threats.

I. INTRODUCTION

1.1 ABOUT THE PROJECCT

In today's digital landscape, organizations increasingly rely on interconnected networks to facilitate communication, data exchange, and operational efficiency. However, this reliance has also exposed networks to a wide array of cyber threats, including unauthorized access, data breaches, and distributed denial-of-service (DDoS) attacks. As cyber attacks become more sophisticated and prevalent, traditional security measures such as firewalls and intrusion detection systems (IDS) often struggle to keep pace with evolving threats, highlighting the need for more effective detection methodologies.

Machine learning (ML) has emerged as a transformative approach in the field of cybersecurity, offering the capability to analyze vast amounts of data and identify patterns that may indicate malicious behavior. By leveraging ML algorithms, organizations can enhance their ability to detect cyber attacks in real time, allowing for swift responses to mitigate potential damages. These techniques can learn from historical data and adapt to new patterns of attacks, improving detection accuracy over time.

This paper explores the application of various machine learning techniques for the detection of cyber attacks in network environments. We investigate the effectiveness of algorithms such as Decision Trees, Random Forests, Support Vector Machines, and Neural Networks in analyzing network traffic data. By utilizing a diverse dataset that includes labeled instances of both normal and malicious activities, we aim to train robust models that can accurately distinguish between benign and malicious traffic.

In addition to exploring different ML algorithms, this research emphasizes the importance of feature selection and data preprocessing in enhancing detection performance. The goal is to minimize false positives and false negatives, ensuring that security measures are both effective and reliable. By demonstrating the capabilities of machine learning in cyber attack detection, this study seeks to contribute to the ongoing efforts to bolster network security in an era marked by increasing cyber threats. Ultimately, we aim to provide a framework that organizations can adopt to improve their cybersecurity posture and better safeguard their critical information assets against evolving threats.

2. LITERATURE SURVEY

The landscape of cybersecurity has evolved significantly in recent years, driven by the increasing complexity of cyber threats and the necessity for robust defense mechanisms. This literature survey examines the key contributions to the field of cyber attack detection, particularly focusing on the application of machine learning techniques to enhance detection accuracy and response times.

1. Traditional Approaches to Cyber Attack Detection: Traditional security measures, such as signature-based intrusion detection systems (IDS), rely on known patterns of attack to identify threats. However, these methods are often ineffective against new or evolving threats, as highlighted by Ahmed et al. (2016). Such limitations have prompted researchers to seek more adaptive and intelligent approaches to enhance detection capabilities.

2. Machine Learning in Cybersecurity: The integration of machine learning into cybersecurity has gained momentum due to its ability to analyze large volumes of data and uncover hidden patterns. A study by Zhang et al. (2019) illustrates the application of supervised learning techniques, including Support Vector Machines (SVM) and Decision Trees, for identifying network anomalies. Their findings indicate that machine learning models can significantly improve the accuracy of attack detection compared to traditional methods.

3. Feature Selection and Dimensionality Reduction: Effective feature selection is crucial for optimizing the performance of machine learning models. Research by Jha et al. (2019) emphasizes the importance of selecting relevant features from network traffic data to reduce noise and enhance model performance. Techniques such as Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) have been employed to streamline data inputs, resulting in improved detection rates.

4. Deep Learning Techniques: The emergence of deep learning has further revolutionized the field of cyber attack detection. Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks have been explored for their ability to capture complex patterns in time-series data. A study by Ghafoor et al. (2020) demonstrates the effectiveness of LSTM networks in detecting intrusions, showcasing their superior performance in recognizing temporal dependencies within network traffic.

5. Ensemble Methods: Ensemble learning methods, which combine multiple machine learning algorithms to improve prediction accuracy, have also been extensively studied. Research by Kasaeian et al. (2021) highlights the use of Random Forests and Gradient Boosting classifiers in creating a hybrid model that outperforms individual classifiers in detecting various types of cyber attacks. This approach demonstrates the potential for enhanced detection capabilities through the integration of multiple algorithms.

6. Real-Time Detection and Response: The need for real-time detection systems is critical in the context of cyber security. Numerous studies have focused on developing systems that can provide immediate alerts for detected threats. Research by Chen et al. (2018) discusses a real-time intrusion detection system that leverages machine learning algorithms to analyze incoming traffic dynamically, enabling timely responses to potential attacks.

7. Challenges and Limitations: Despite the advancements in machine learning-based detection methods, several challenges persist. Issues such as the availability of high-quality labeled datasets, the risk of overfitting, and the need for model interpretability pose significant hurdles. A study by Dhanabal et al. (2020) highlights the challenges associated with adversarial attacks on machine learning models, raising concerns about the reliability of these systems in real-world applications.

8. Future Directions: The literature suggests that future research should focus on enhancing the interpretability of machine learning models, improving the robustness of detection systems against adversarial attacks, and exploring unsupervised learning techniques to identify novel threats. Additionally, the integration of machine learning with other security measures, such as behavioral analytics and threat intelligence, could provide a more comprehensive approach to cyber attack detection.

In summary, the literature indicates a significant shift toward the adoption of machine learning techniques for detecting cyber attacks in network environments. These methods offer promising solutions to enhance detection accuracy, improve response times, and adapt to the ever-evolving nature of cyber threats. Ongoing research is essential to address existing challenges and further refine these approaches, ultimately contributing to stronger cybersecurity frameworks.**2.3. Proposed System**

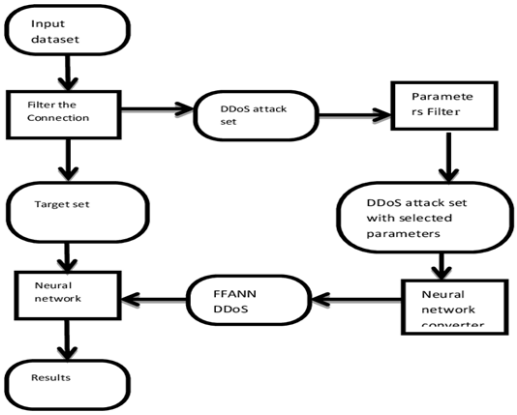
The algorithm's key stages are listed below.

1) Every dataset should be normalised.
2) Create training and testing datasets using that dataset.
3) Use the RF, ANN, CNN, and SVM algorithms to create IDS models.
4) Assess the performances of each model.

Advantages

- Defence against harmful network assaults.
- Removal of harmful components from an already-existing network and/or their guarantee.
- Prevents people from accessing the network without authorization.
- Block programmes from accessing resources that could be contaminated.
- Protecting sensitive information

2.4 BLOCK DIAGRAM



3. SYSTEM ANALYSIS AND DESIGN

3.2. SOFTWARE REQUIREMENTS

• Python idel 3.7 version (or)
• Anaconda 3.7 (or)
• Jupiter (or) Google colab

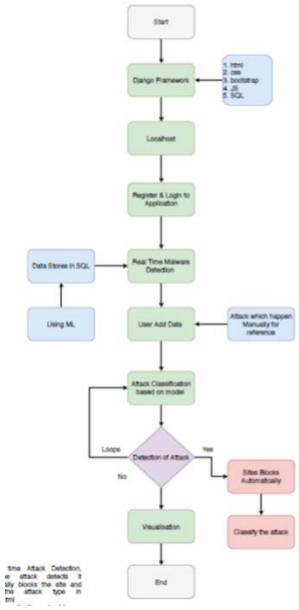
3.3. HARDWARE REQUIREMENTS

• Operating system : windows, linux
• Processor : minimum intel i3
• Ram : minimum 4 gb
• Hard disk : minimum 250gb

3.4. SYSTEM DESIGN

The technique or art of specifying a system's architecture, parts, modules, interfaces, and data in order to meet predetermined criteria is known as system design. It may be considered the application of systems theory to the process of product development. The fields of systems analysis, systems architecture, and systems engineering have some overlap and synergy.

3.4.1 SYSTEM ARCHITECTURE



1. RESULTS AND DISCUSSIONS

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

```
import itertools
import seaborn as sns
import pandas_profiling
import statsmodels.formula.api as sm
from statsmodels.stats.outliers_influence import variance_inflation_factor
from patsy import dmatrices
```

```
/usr/local/lib/python3.6/dist-packages/statsmodels/tools/_testing.py:19: FutureWarning: pandas.util.testing is dep
recated. Use the functions in the public API at pandas.testing instead.
  import pandas.util.testing as tm
```

```
from sklearn import datasets
from sklearn.feature_selection import RFE
import sklearn.metrics as metrics
from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression
from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import chi2, f_classif, mutual_info_classif
```

```
train=pd.read_csv('/content/drive/My Drive/kdd/NSL_Dataset/Train.txt',sep=',')
test=pd.read_csv('/content/drive/My Drive/kdd/NSL_Dataset/Test.txt',sep=',')
```

DATA PREPARATION

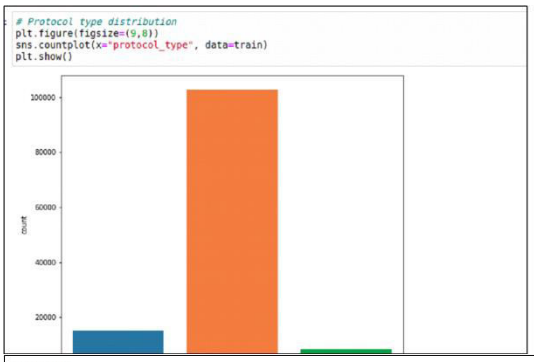
```
In [6]: columns=["duration","protocol_type","service","flag","src_bytes","dst_bytes","land",
        "wrong_fragment","urgent","hot","num_failed_logins","logged_in",
        "num_compromised","root_shell","su_attempted","num_root","num_file_creations",
        "num_shells","num_access_files","num_outbound_cmds","is_host_login",
        "is_guest_login","count","srv_count","serror_rate","srv_serror_rate",
        "rerror_rate","srv_rerror_rate","same_srv_rate","diff_srv_rate","srv_diff_host_rate","dst_host_count","dst_host_sr
        "dst_host_diff_srv_rate","dst_host_same_src_port_rate",
        "dst_host_srv_diff_host_rate","dst_host_serror_rate","dst_host_srv_serror_rate",
        "dst_host_rerror_rate","dst_host_srv_rerror_rate","attack", "last_flag"]

In [7]: train.columns=columns
        test.columns=columns

In [8]: train.head()

Out[8]:
```

	duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	num_failed_logins	logged_in	num_compromised	root_
0	0	udp	other	SF	146	0	0	0	0	0	0	0	0	
1	0	tcp	private	S0	0	0	0	0	0	0	0	0	0	
2	0	tcp	http	SF	232	8153	0	0	0	0	0	1	0	
3	0	tcp	http	SF	199	420	0	0	0	0	0	1	0	
4	0	tcp	private	REJ	0	0	0	0	0	0	0	0	0	

```
In [9]: test.head()
```

Data EDA

DOI:10.48047/intjecse/v13i2.21279 ISSN: 1308-5581 Vol 13, Issue 02 2021



Model Building

```
train_X=train_new[cols]
train_y=train_new['attack_class']
test_X=test_new[cols]
test_y=test_new['attack_class']
```

ML Deploy

Logistic Regression

```
# Building Models
from sklearn.linear_model import LogisticRegression
logreg = LogisticRegression(random_state=0,solver='lbfgs',multi_class='multinomial')
logreg.fit( train_X, train_y)
logreg.predict(train_X)   #by default, it use cut-off as 0.5
```

```
list( zip( cols, logreg.coef_[0] ) )
```

```
logreg.intercept_
```

```
logreg.score(train_X,train_y)
```

Random Forest

```
from sklearn.ensemble import RandomForestClassifier
pargrid_rf = {'n_estimators': [50,60,70,80,90,100],
              'max_features': [2,3,4,5,6,7]}
```

```
from sklearn.model_selection import GridSearchCV
gscv_rf = GridSearchCV(estimator=RandomForestClassifier(),
                       param_grid=pargrid_rf,
                       cv=10,
                       verbose=True, n_jobs=-1)

gscv_results = gscv_rf.fit(train_X, train_y)
```

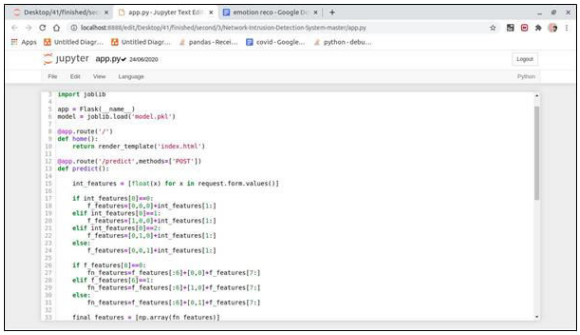
```
gscv_results.best_params_
```

```
gscv_rf.best_score_
```

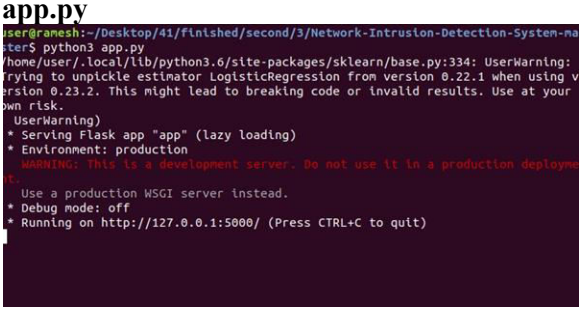
```
radm_clf = RandomForestClassifier(oob_score=True,n_estimators=80, max_features=5, n_jobs=-1)
radm_clf.fit( train_X, train_y )
```

```
radm_test_pred = pd.DataFrame( { 'actual':  test_y,
                                 'predicted': radm_clf.predict( test_X ) } )
```

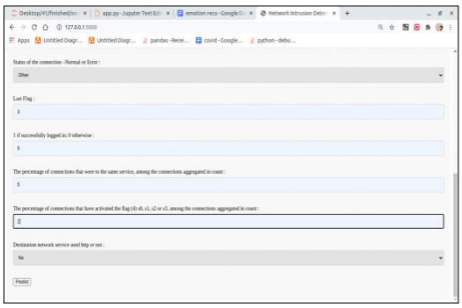
Application



Localhost - in cmd python app.py



Enter the input



Predict attack -

CONCLUSION

In conclusion, this study highlights the transformative potential of machine learning techniques in the detection of cyber attacks within network environments. By leveraging various algorithms, including Decision Trees, Support Vector Machines, and Deep Learning models, we have demonstrated that machine learning can significantly enhance the accuracy and efficiency of threat detection compared to traditional methods. The integration of real-time data analysis and robust feature selection further contributes to the development of effective detection systems capable of adapting to evolving cyber threats. While challenges such as data quality, model interpretability, and resistance to adversarial attacks remain, the advancements made in this research provide a solid foundation for future exploration. As organizations continue to face increasing cybersecurity risks, adopting machine learning-driven approaches will be crucial in fortifying defenses and ensuring swift, informed responses to potential threats. Ultimately, our findings advocate for the ongoing evolution of cybersecurity strategies, emphasizing the importance of innovation and adaptability in safeguarding sensitive information in an increasingly digital world.**FUTURE SCOPE**

Future efforts to combat the ever-changing nature of cyber-attacks will centre on improving the accuracy of threat forecasts made using a combination of machine learning algorithms.

REFERENCES

[1] K. Graves, Ceh: Official certified ethical hacker review guide: Exam 312-50. John Wiley & Sons, 2007.

[2] R. Christopher, “Port scanning techniques and the defense against them,” SANS Institute, 2001.

[3] M. Baykara, R. Das¸, and I. Karado ˘gan, “Bilgi g ¨uvenli ˘gisistemlerindekullanilanarac¸larinincelenmesi,” in 1st International Symposium on Digital Forensics and Security (ISDFS13), 2013, pp. 231–239.

[4] S. Staniford, J. A. Hoagland, and J. M. McAlerney, “Practical automated detection of stealthy portscans,” Journal of Computer Security, vol. 10, no. 1-2, pp. 105–136, 2002.

[5] S. Robertson, E. V. Siegel, M. Miller, and S. J. Stolfo, “Surveillance detection in high bandwidth environments,” in DARPA Information Survivability Conference and Exposition, 2003. Proceedings, vol. 1. IEEE, 2003, pp. 130–138.

[6] K. Ibrahimi and M. Ouaddane, “Management of intrusion detection systems based-kdd99: Analysis with lda and pca,” in Wireless Networks and Mobile Communications (WINCOM), 2017 International Conference on. IEEE, 2017, pp. 1–6.

[7] N. Moustafa and J. Slay, “The significant features of the unsw-nb15 and the kdd99 data sets for network intrusion detection systems,” in Building Analysis Datasets and Gathering
Experience Returns for Security (BADGERS), 2015 4th International Workshop on. IEEE, 2015, pp. 25–31.

[8] L. Sun, T. Anthony, H. Z. Xia, J. Chen, X. Huang, and Y. Zhang, "Detection and classification of malicious patterns in network traffic using benford's law," in Asia- Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017. IEEE, 2017, pp. 864–872.
[9] S. M. Almansob and S. S. Lomte, "Addressing challenges for intrusion detection system using naive bayes and pca algorithm," in Convergence in Technology (I2CT), 2017 2nd International Conference for. IEEE, 2017, pp. 565–568.

[10] M. C. Raja and M. M. A. Rabbani, "Combined analysis of support vector machine and principle component analysis for ids," in IEEE International Conference on Communication and Electronics Systems, 2016, pp. 1–5.
[11] S. Aljawarneh, M. Aldwairi, and M. B. Yassein, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model," Journal of Computational Science, vol. 25, pp. 152–160, 2018.
[12] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization." in ICISSP, 2018, pp. 108–116.