Cyber Threat's detection using Machine Learning Algorithms

Wendyam Carine Tapsoba, Didier Bassole, Rodrique Kafando, Abdoul Kader Kabore, Aminata Sabané, Tégawendé F. Bissyandé

Cyber Threat's detection using Machine Learning Algorithms

Wendyam Carine Tapsoba[1], Didier Bassole[2], Rodrique Kafando[1], Abdoul Kader Kabore[1], Aminata Sabané[2], Tégawendé F. Bissyandé[1 2]

[1]Centre d'Excellence CITADEL, Université Virtuelle du Burkina Faso
[2]Département d'Informatique, UFR/SEA, Université Joseph KI-ZERBO, Burkina Faso

*E-mail : carinetapsoba1697@gmail.com

Abstract

Cybercrime presents significant challenges to the security of data and computer systems. The security of data and computer systems has become an escalating concern with the growing utilization of the Internet in Burkina Faso. This article addresses these challenges by proposing a novel approach employing machine learning algorithms, including Support Vector Machine (SVM), Random Forest (RF), and Long Short-Term Memory (LSTM), to enhance the detection and classification of cyber threats. We utilized three datasets containing text and URLs to explore various cyber threats such as phishing, cyberbullying, and online scams. The performance of our model, when compared to previous research, demonstrates promising results. Notably, the RF algorithm exhibited exceptional accuracy, achieving rates of 99.99% for cyberbullying, 99.4% for phishing, and 99.94% for online scams.

Keywords

Machine Learning, cyberthreat, detection

I INTRODUCTION

The advent of the digital age has brought opportunities, but also cybercrime, an ever-evolving threat. Despite a low rate of digital usage in Africa, Kaspersky counted 28 million cyberattacks between January and August 2020 [8]. Financial losses in Burkina Faso in 2022 were 1.1 billion CFA francs, and the cumulative figure for the first two quarters of 2023 was 597.8 million CFA francs[1]. These losses impact on business credibility and privacy. Privacy protection is becoming crucial, with attacks such as phishing and identity theft. Social networks amplify these risks.

To counter these threats, the integration of artificial intelligence is emerging as an innovative solution. Cybersecurity requires automated techniques to detect cyberthreats, particularly on social networks. So how can machine learning be harnessed to improve the detection of cyberthreats and deal with their diversity?

The aim is to use machine learning to detect and classify cyberthreats such as phishing, online scams and cyberstalking. The aim is to propose an approach, evaluate its performance, and implement an intelligent platform for detecting and classifying cyberthreats.

[1]BCLCC statistics

In the present paper, the output of each training session was obtained on the basis of data consisting essentially of textual information on jobs, tweets and URLs characteristics from Kaggle. Having undergone a certain amount of pre-processing according to their format, these were subjected to several algorithms in order to form models capable of detecting a phishing URL, a scam ad and an intimidating or harassing message. In the remainder of this paper, we will review related literature, present our solution in the methodology section, discuss the results and conclude with an outlook.

II LITERATURE REVIEW

2.1 Concepts and definitions

The National Institute of Standards and Technology (NIST) defines computer security as "the protection of information from unauthorized disclosure, modification, destruction and disruption, while providing continuous availability, integrity and confidentiality.

According to the glossary of the French Information Systems Security Agency (ANSSI) [15]:

- cybersecurity covers all information system security techniques, and is based on the fight against cybercrime and the implementation of cyberdefense.
- cybercrime is defined as any act that contravenes international treaties or national laws, using networks or information systems as a means of committing an offence or crime, or targeting them.
- cyberthreat is the coordinated set of actions carried out in cyberspace that target information or the systems that process it, undermining its availability, integrity or confidentiality.

In the digital realm, where opportunities abound, there is also a rising tide of online criminal activities utilizing various cyberthreat techniques. These include phishing, where personal or financial information is deceitfully obtained; cyberstalking, involving aggressive electronic harassment; Distributed Denial of Service (DDoS) attacks that overload and incapacitate websites or online services; and fake job scams that deceive job seekers with bogus employment offers. Each of these cybercrimes leverages technology to exploit, harass, or defraud victims, underscoring the need for robust cybersecurity measures.

2.2 Related works

As the schemes employed by computer criminals become increasingly elusive, researchers are developing new techniques to spot and counter these attacks. The main focus is on Machine Learning and Deep Learning algorithms. ML, DL and TL algorithms represent a more effective and accurate means of detecting and preventing cyberattacks. In particular, these algorithms have been used to identify specific categories of cyberattack, such as phishing, malware, distributed denial-of-service (DDoS) attacks and social engineering attacks [2, 5–7, 9, 10]. Several studies have looked at the detection of cybercrime using machine learning methods. These studies are classified according to the approach used to detect the various threats.

- Behavioral approach

The behavior-based approach focuses on monitoring the behavior of users or online systems. Machine learning models are used to recognize activities that deviate from established norms,

such as unauthorized access attempts or unusual data movements. Researchers have explored this approach to cybercrime detection.

Cagri et al. [1] in 2018, used machine learning approaches such as SVM, decision trees and random forests to automatically detect suspicious accounts on the Twitter platform. Certain behavioral, profile and content features extracted from the tweets were then applied to a certain approach to identify the anonymous account. However, the authors only focused on tweet-based suspicious account detection when implementing this model.

In 2022, Ahammad et al. [5] tackle the problem of phishing, a form of cybercrime aimed at stealing personal data via malicious websites. It highlights the challenges of detecting malicious URLs. The author proposes a solution using machine learning algorithms such as Random Forests, Decision Trees, Light GBM, Logistic Regression, and Support Vector Machine (SVM) to analyze the behavior and characteristics of suspicious URLs. (They achieve an accuracy of 86% with LGBM).

Hung Le et al.[16] have developed a malicious URL detection system. To this end, they present URLNet, an end-to-end deep learning framework designed to acquire non-linear URL integration for direct detection of malicious URLs from the URL itself. They adopt the principles of Convolutional Neural Networks (CNN) to the characters and words present in the URL, to learn URL integration within a joint optimization framework. This approach enables the model to capture diverse semantic information, which was not possible with pre-existing models. In addition, they propose advanced embedding techniques to solve the rare word problem frequently encountered in this task.

- Anomaly detection approach

Anomaly detection is based on the creation of a model of normal system behavior. Any significant deviation from this behavior is then considered a possible cyberattack. This can include monitoring network activity, traffic patterns, unusual connections, etc. Here are a few examples of how different researchers have approached the subject with this approach.

Ganesan & Mayilvahanan [14] have developed a methodological approach aimed at identifying unexpected patterns. They examined cybercrime-related data from a database containing a variety of data fields extracted from web pages on the Internet. These data fields encompassed aspects such as cyberbullying, scams, theft, identity theft, defamation and harassment. Their aim was to introduce a model to categorize computer crimes as violent or non-violent, and to classify them into different categories such as cyberterrorism, cyberstalking, pornography, cyberbullying and cyberfraud. The machine learning algorithms used in this model are SVM, DT, ANN, NB.

Vinayakumar Ravi et al.[11] proposed a new automated approach using deep learning to detect randomly generated domain names (DGAs) and DNS homograph attacks, without the need for reverse engineering. They claim that current DGA detection methods are laborious, time-consuming and error-prone. Evaluation of the model on real data sets shows significant effectiveness, with an accuracy of 0.99. Furthermore, the model is resistant to common adversarial attacks, underlining the importance of developing robust detection models in the face of adversarial learning. Algorithms used for this model include BiLSTM, GRU, CNN and RNN.

In the work by Zolanvari et al. [3], the authors compared several traditional ML (Machine Learning) algorithms, including K-Nearest Neighbors (KNN), Random Forest (RF), Decision

Tree (DT), Logistic Regression (LR), Artificial Neural Network (ANN), Naive Bayes (NB) and Support Vector Machine (SVM) to detect cyber-attacks in a water storage system. Their evaluation showed that the RF algorithm is the best model with a recall of 0.9744, ANN is the fifth-best algorithm with a recall of 0.8718 and LR had a recall of 0.4744. They also reported that ANN could not detect 12.82% of attacks, but considered 0.03% of normal samples as an attack due to the unbalanced nature of the data. Based on the results, the authors considered many attack samples as normal samples without labeling as many normal samples as attacks for LR, SVM and KNN.

Considering the research mentioned above, it has been observed that, faced with the variety and abundance of data generated by computer networks and systems, conventional approaches to IT security are now proving insufficient to provide adequate protection for systems and users. Machine Learning and Deep Learning methods represent a more effective and accurate means of detecting and preventing cyber threats.

III METHODOLOGY

This study is proposed to improve the detection and classification of cybercrime using ML and DL methods.

3.1 Datasets presentation

In the context of this study, we used three (03) different datasets, which we present in table 1. Due to the lack of data within the BCLCC, we predominantly relied on publicly available data.

Table 1: summary of dataset

Dataset	Source	Data types	Features	lines
Phishing Legitimate full[2]	Kaggle	Urls	50	10000
Fake Job Posting Prediction[3]	Kaggle	Textual	18	18000
Suspicious Communication[4]	Kaggle	Textual	2	20001

Dataset 1
The dataset contains more than 800,000 URLs, of which 52% are legitimate and the remaining 47% are phishing domains. Legitimate URLs are URLs that do not lead to any infecting website, and do not attempt to inject any malware into the user's computer. The dataset contains two columns, URL and status, where the status column represents values coded as 0 and 1, where 0 represents phishing domains and 1 represents legitimate domains.The two categories are almost equivalent, so there is no imbalance in the classes.

Dataset 2
Real / Fake Job Posting Prediction [13]: Prediction of fake job postings; This dataset contains 18,000 job descriptions, of which around 800 are false.It contains 18 columns, 17 of which are reserved for features and the last column is for the target variable. The data consists mainly of textual job information.

Dataset3
On April 15, 2020, UNICEF issued a warning in response to the increased risk of cyberbullying during the COVID-19 pandemic, due to widespread school closures, increased screen time and fewer face-to-face social interactions. The statistics on cyberbullying are alarming: 36.5% of

middle and high school students have felt cyberbullied, and 87% have observed instances of cyberbullying, the effects of which range from lower academic performance to depression and suicidal thoughts. In light of all this, this dataset contains over 47,000 tweets tagged according to cyberbullying class: age; ethnic origin; gender; religion; other type of cyberbullying and no cyberbullying. The data was balanced to contain around 8,000 tweets for each class.

3.2 data pre-processing

Data pre-processing aims to improve data quality, reduce noise, eliminate outliers, normalize data, and organize it so that it can be easily exploited for analysis, modeling or visualization tasks. Here are some of the common operations performed in our method: synthetic Minority Oversampling Technique (SMOTE) [4], tokenization [12], removing stop words[17] and feature selection.

3.3 Architecture

Our structure is divided into five distinct sections:

- the first step in our architecture is to collect data from various sources) such as social networks (twitter), web pages and others. This data provides an overview of activities on the web, and serves as the basis for detecting cyber threats.
- in the second stage, the data undergoes a pre-processing process. This includes data cleaning to eliminate outliers and errors, normalization to make the data comparable, and management of missing values. This step is essential to ensure data quality prior to analysis.
- the third step involves extracting features from the pre-processed data. These features are potential indicators of malicious activity, such as IP addresses, connection patterns, the frequency of forged links in the status bar, the length of a URL, the number of dots in a URL, the prerequisites in a job advert, its description and the salary claim. Careful selection of these characteristics is crucial to the performance of the detection system.
- to detect cyberthreats, we use machine learning and deep learning algorithms (RF, LSTM and SVM). These models analyze extracted features and identify anomalous or malicious activity. They are trained using historical data that includes examples of cybercriminal activity as well as normal activity. Model training requires rigorous validation to ensure its effectiveness.
- once the models have been successfully trained, the architecture can be deployed in production to monitor in real time, and take into account other relevant data. Our models identify potential cyber-threats (phishing, online scams and cyber-stalking) by comparing real-time data with trained models.

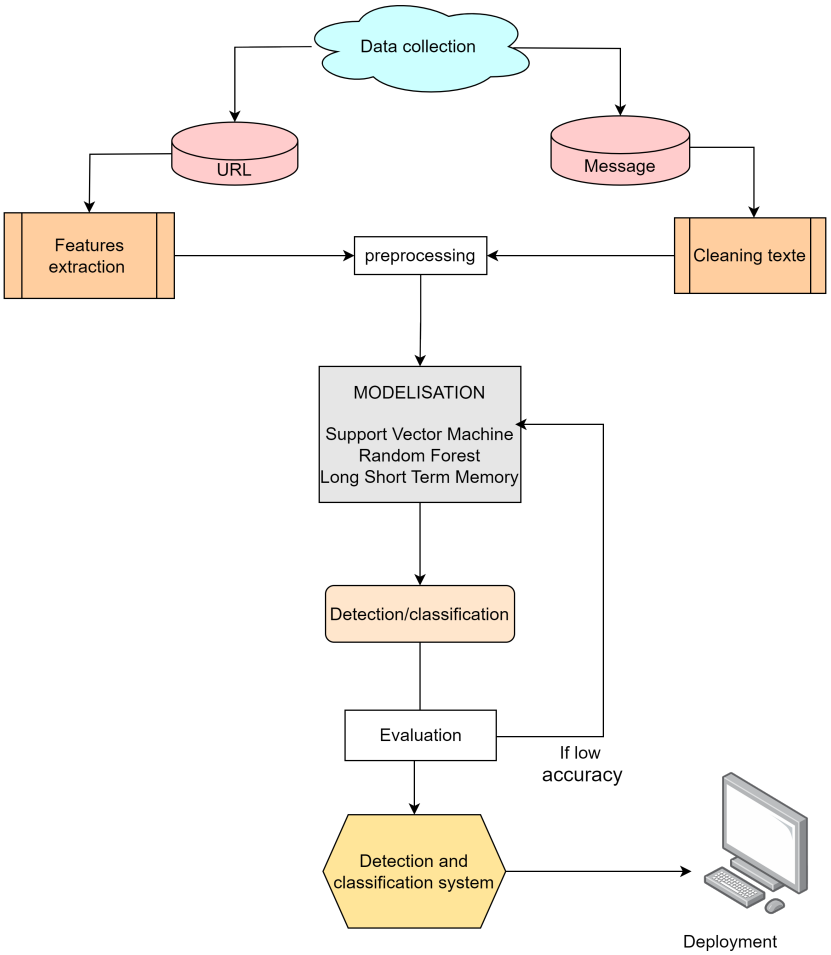

Figure 1: Architecture

We used several tools, mainly from the open source community. The models were implemented using the Python programming language and the Jupyter Notebooks editor. Our working environment is Google Colab, a cloud platform offering access to shared resources, including graphics processing units (GPUs) and random access memory (RAM), operated through Jupyter Notebook. For data processing and model implementation, we employed a set of libraries and tools such as Tensorflow, Pandas, Keras, Scikit-learn. To implement our detection system, we used streamlit, an open-source tool that enables developers to rapidly create interactive web applications using the Python programming language.

IV RESULTS

To evaluate these results, we consider the following metrics: accuracy and F1 score, which are very good indicators among the others. Accuracy reflects the accuracy of the model, while the F1 score provides a single score that combines the notions of precision and recall in a single indicator. They therefore summarize the level of the model.
For our three models, we have fairly high values, i.e. 99.99% for accuracy and 99.94% for the F1 score, giving a solid assessment of the models. An accuracy of 99.99% indicates that the model has correctly classified the vast majority of samples, and an F1 score of 99.94% suggests an excellent balance between the model's ability to avoid false positives and false negatives.

As a reminder, we trained three models capable of classifying three cyberthreats, namely phishing, scamming and cyberstalking. For phishing, our results are presented in the table below 2.

Table 2: Algorithm performance for phishing attacks

	SVM	RF	LSTM
Accuracy	85.75%	98.4%	94.9%
Precision	82.89%	98.51%	94.26%
Recall	90.51%	98.2%	95.75%
F1 Score	86.53%	98.41%	95%

The results for online scam are shown in table 3.

Table 3: Algorithm performance for online scams

	SVM	RF	LSTM
Accuracy	99.28%	99.94%	99.4%
Precision	98.98%	99.88%	98.8%
Recall	99.57%	1	1
F1 Score	99.27%	99.94%	99.4%

In table 4 we have the results of cyberbullying.

Table 4: Algorithm performance for cyberbullying n

	SVM	RF	LSTM
Accuracy	59.09%	**99.99**%	99.35%
Precision	41%	94%	91%
Recall	24%	93%	91%
F1 Score	17%	93%	91%

As previous works do not use the same data as we do, we decided to consider those employing ML techniques similar to ours to conduct our comparative analysis, which is summarized in the table 5.

Compared with previous work, we have been able to improve the detection and classification of cyberthreats with an evolution rate of 2.12%. What's more, our approach innovatively integrates a detection system, representing a significant advance in the field of computer security.

Table 5: Comparison table

Author	Algorithmes	Accuracy	system
[5]	SVM RF LR	SVM: 86%	absent
[3]	SVM RF DT	RF: 97.44%	absent
[14]	SVM DT NB	-	absent
[1]	SVM RF DT	RF: 97.87%	present
Our works	**SVM RF LSTM**	**RF : 99.99%**	**present**

4.1 Conclusion and future work

This in-depth study focused on the detection and classification of cyberattacks using machine learning and deep learning approaches. Three predominantly text-based datasets, including tweets, job advertisements and URLs, were subjected to a thorough processing process. These datasets were then trained using various algorithms, including SVM, LSTM and RF. We were also able to set up an intelligent cyberthreat detection tool.

We made a comparison with the state of the art, which allowed us to see the considerable contribution of our work, while recognizing opportunities for improvement.

The inherent limitations of our research are diverse, including the challenge of the limited selection of cyberthreats considered, given the diversity and profusion of current threats. This limitation may restrict the generalizability of our results to the full spectrum of cybercriminal activity.

It's important to stress that cybersecurity is a dynamic field, and we need to remain vigilant in the face of constant changes in attacker techniques. So, to improve the overall performance of our system, we'll consider exploring ways of expanding our dataset, both in terms of threat diversity and temporal characteristics and to explore new machine learning techniques.

ACKNOWLEDGEMENT

This work was conducted as part of the Artificial Intelligence for Development in Africa (AI4D Africa) program, with the financial support of Canada's International Development Research Centre (IDRC) and the Swedish International Development Cooperation Agency (Sida).

References

[1] Ç. B. Aslan, R. B. Sağlam, and S. Li. "Automatic Detection of Cyber Security Related Accounts on Online Social Networks: Twitter as an example". In: *Proceedings of the 9th International Conference on Social Media and Society*. SMSociety '18: International Conference on Social Media and Society. Copenhagen Denmark: ACM, July 18, 2018, pages 236–240. ISBN: 978-1-4503-6334-1.

[2] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, and S. Venkatraman. "Robust Intelligent Malware Detection Using Deep Learning". In: *IEEE Access* 7 (2019), pages 46717–46738. ISSN: 2169-3536.

[3] M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan, and R. Jain. "Machine Learning-Based Network Vulnerability Analysis of Industrial Internet of Things". In: *IEEE Internet of Things Journal* 6.4 (Aug. 2019), pages 6822–6834. ISSN: 2327-4662, 2372-2541.

[4] R. Kassel. *Gestion des problèmes de Classification déséquilibrée - Partie II*. Formation Data Science | DataScientest.com. Mar. 25, 2020. URL: https://datascientest.com/comment-gerer-les-problemes-de-classification-desequilibree-partie-ii (visited on 01/28/2024).

[5] S. H. Ahammad, S. D. Kale, G. D. Upadhye, S. D. Pande, E. V. Babu, A. V. Dhumane, and M. D. K. J. Bahadur. "Phishing URL detection using machine learning methods". In: *Advances in Engineering Software* 173 (Nov. 2022), page 103288. ISSN: 09659978.

[6] M. Ahsan, K. E. Nygard, R. Gomes, M. M. Chowdhury, N. Rifat, and J. F. Connolly. "Cybersecurity Threats and Their Mitigation Approaches Using Machine Learning—A Review". In: *Journal of Cybersecurity and Privacy* 2.3 (July 10, 2022), pages 527–555. ISSN: 2624-800X.

[7] N. E. H. B. CHAABENE. “Detection d’utilisateurs violents et de ´ menaces dans les reseaux sociaux”. THESE. Paris, FRANCE: Institut Polytechnique de Paris, Jan. 31, 2022. 185 pages.

[8] *Cybersécurité: l’Afrique a perdu 10% de son PIB dans les cyberattaques en 2021*. La Tribune. May 4, 2022. URL: https://afrique.latribune.fr/africa-tech/2022-05-04/cybersecurite-l-afrique-a-perdu-10-de-son-pib-dans-les-cyberattaques-en-2021-916363.html.

[9] L. Elluri, V. Mandalapu, P. Vyas, and N. Roy. *Recent Advancements in Machine Learning For Cybercrime Prediction*. Oct. 9, 2023. arXiv: 2304.04819[cs].

[10] G. Mumtaz, S. Akram, M. W. Iqbal, M. U. Ashraf, K. A. Almarhabi, A. M. Alghamdi, and A. A. Bahaddad. “Classification and Prediction of Significant Cyber Incidents (SCI) Using Data Mining and Machine Learning (DM-ML)”. In: *IEEE Access* 11 (2023), pages 94486–94496. ISSN: 2169-3536.

[11] V. Ravi, M. Alazab, S. Srinivasan, A. Arunachalam, and K. P. Soman. “Adversarial Defense: DGA-Based Botnets and DNS Homographs Detection Through Integrated Deep Learning”. In: *IEEE Transactions on Engineering Management* 70.1 (Jan. 2023), pages 249–266. ISSN: 0018-9391, 1558-0040.

[12] *API de Tokenisation et Lemmatisation*. URL: https://nlpcloud.com/fr/nlp-tokenization-api.html (visited on 01/25/2024).

[13] *Employment Scam Aegean Dataset*. URL: http://emscad.samos.aegean.gr/ (visited on 01/25/2024).

[14] M. Ganesan and P. Mayilvahanan. “Cyber Crime Analysis in Social Media Using Data Mining Technique”. In: ().

[15] *Glossaire | ANSSI*. URL: https://cyber.gouv.fr/glossaire (visited on 01/10/2024).

[16] H. Le, Q. Pham, and D. Sahoo. “URLNet : Apprentissage d’une représentation d’URL avec Deep Learning pour Détection d’URL malveillantes”. In: (), page 13.

[17] *Nettoyez et normalisez les données*. OpenClassrooms. URL: https://openclassrooms.com/fr/courses/4470541-analysez-vos-donnees-textuelles/4854971-nettoyez-et-normalisez-les-donnees (visited on 01/28/2024).