ISSN: 3064-8270

Artificial Intelligence, Machine Learning, and Data Science Journal

Research Article

LEVERAGING DEEP LEARNING FOR DETECTING SUSPICIOUS ACTIVITY IN ONLINE FORUM DISCUSSIONS

Chidi Adebayo Olatunde and Adaeze Nnena Okafor

Department of Computer Science, Lead City University Ibadan, Nigeria

DOI: 10.5281/zenodo.14892634

Abstract

Nowadays, people are passionate about using the internet in their daily lives, and this has resulted in the rapidly increasing adoption and use of online forums. An online forum can be defined as a medium to share one's thoughts, feelings and emotions towards specific multimedia artefacts such as pictures, videos and paintings etc. However, the use of online forums has leads to the execution of many illegal activities such as trading of black market money online, distributing copyrighted movies, as well as using illegal words. Therefore, Law enforcement needs a system to effectively deal with this problem. An example of such system include the Network Intrusion Detection System (NIDS) that assists system administrators detect network security vulnerabilities in their organization. On the other, many challenges arise when developing a flexible and effective NIDS for unplanned and unpredictable attacks. This paper proposes a deep learning-based approach to develop a flexible and efficient NIDS to analyze suspicious and criminal activities occurring in online forums. The proposed system combined a variety of Deep learning techniques and Natural Language Processing (NLP) for suspicious keyword extraction as well as Support Vector Machine (SVM) for detection and classification of suspicious keywords. We present the performance of our approach and compare it with some previous works. The metrics to be compared include accuracy, precision, recall, and f-measure values. This research improves system performance and security compared to existing systems.

Keywords: Online forum; machine learning; deep learning; natural language processing; support vector machine

1. INTRODUCTION

The internet has become an effective and convenient communication channel for knowledge sharing, expression of opinions, products advertisement, and post textual data via the browser interface to share information[1]. Therefore, it is important to extract useful information from this plain text data to reveal the hidden data, and one of the most widely used technique for data extraction is machine learning. The goal of machine learning is to extract information from large datasets and transform it into an easy-to-understand format. As Internet technology continues to grow, it has led to the conduction of several lawful and unlawful activities. Several direct messages are discussed on Internet forums long before they are published in the traditional media. On the other hand, they also serve as an effective channel for unlawful activities such as copyrighted movie distribution, sending of threatening messages, and online gambling[1].

Similarly, hackers often use social media networks to discuss cyber-attacks, share strategies and tools, and identify potential victims for targeted attacks. Analysts examining these discussions can forward information about

ISSN: 3064-8270

Artificial Intelligence, Machine Learning, and Data Science Journal

Research Article

malicious activity to system administrators who can then detect, defend against, and recover from future attacks. For example, prior to the anticipated cyber-attacks on Israeli government websites by the hacking group Anonymous, government analysts were monitoring hackers on Facebook and in private chat rooms. As a result, system administrators were prepared to counter distributed denial-of-service attacks and defacement of government websites[2]. In addition, many facts prove that simply managing information on the internet through traditional management models is not enough. In this regard, web mining is a new research direction for information gathering and analysis on the explosive and unstructured Internet. Criminal web data always provides valuable and relevant information for legal management purposes. Furthermore, it is always very difficult to assess the various capabilities of a wide range of criminal web data and is therefore one of the most important tasks of legal management[2].

In recent decades, information technology has made great strides in hardware and software. This has had a huge impact on simplifying the communication processes of organizations, especially within the business environment. Originating in the United States, the Internet is based on "a network of connected computers, each connected to a set of other computers that communicate electronically with other computers around the world" (Henslowe, 1999, p.87). The evolution of the internet and the advent of social media led to the creation of social interactive platforms where individuals on the network create, share and share content6[7].

Due to the fact that forums are platforms for people from different geographic regions to express and share their opinions, and through marketing and communication, they influence many aspects of their lives. Therefore, monitoring suspicious discussions on forums is the best way to measure user loyalty. Some people use these discussion forums for illegal purposes by posting suspicious chats in the form of text, video or images and sharing them with other users[4].

2. RELATED WORKS

A study on Suspicious Pattern Detection (SPD) Algorithm for the identification of suspected cyber threat in instant chat messenger available on Social Networking Websites and Instant Messengers[5]. The proposed framework considered the Ontology based Information Extraction technique (OBIE) with a pre-defined knowledge base data mining approach of Association Rule Mining (ARM). The proposed concept involves three major steps: (a) word extraction from unstructured text (b) e-crime monitoring system program (c) SPD algorithm. The proposed concept was tested using the Global Terrorist Database (GTD). The proposed concept was compared with other Instant Messengers, Mobile Phone Apps and Social Networking Sites based on the ability to detect suspicious information during online chats. As per considered parameters, the proposed concept shows improved and efficient results[5].

Most of the data of online forums are stored in text format, therefore, a study conducted by 6 made use of only text format of suspected postings for the recognition of suspicious information available on the internet in the

form of user comments or views, identification of spammers, fraudsters, and other types of attackers that use the latest technology for the criminal activities. The study also proposed a framework that used an integrated approach of Support Vector Machine (SVM) and Particle Swarm Optimization (PSO). SVM is a statistical learning based data mining approach and PSO is swarm intelligence based concept considered to optimize the parameters of SVM[6]. The proposed framework showed improved results.

In the study by[7], a review was carried out on the use of Term Frequency–Inverse Document Frequency (TF-IDF) for document classification using logistic regression and linear support vector machine classifiers. Both classifiers were trained to rapidly require little computation to analyze a document, and provide an output score proportional to the probability that the input document contained cyber content. Results was in the form of the Detection Error Trade off (DET) curves that show how false- alarm and miss probabilities vary as the threshold on the classifier's output probability varies as plotted on normal deviate scales. The results also showed that Logistic regression classifier performs better than the keyword system and the logistic regression classifier passes through the performance target region, meaning it misses less than 10% of cyber documents with a false-alarm rate of less than 1%. data mining approach for the detection of suspicious activities on online forums[3]. The study was conducted using textual data from online forums to extract suspicious information. After the pre-processing steps of stop words removal and stemming process with Brute Force algorithm, the study also used the matching algorithm for suspicious keyword recognition. Finally, the study used the keyword spotting techniques, leaning based method and a combination of defined approaches for the overall recognition of suspicious human activity[3].

This research conducted by[4] proposed an approach to differentiate twitter data as either an actual information or rumour by extracting twitter data for some particular topic with the help of Hashtag functions. For the validation of concept for any particular information, data of some well-known news channels were evaluated and compared with the results of semantic and sentiment analysis of tweets. The study also proposed a prototype, namely, ―The Twitter Grapevine‖ to target the rumours specifically for Indian domains. The overall results were evaluated based on the accuracy analysis initially for digital India & facebook.org rumours and then for Kerala House & Beef Rumour topics. In these results, favourable and unfavourable result predictions have been evaluated. The study found that the accuracy of the results for the later experiments were much lower than the former one of 76.99%[4].

Social media offers a variety of avenues through which we can communicate with people. In fact, social media is known to have been used widely in educational field also. Over the last 30 years, the nature of communication has undergone a substantial change and it is still changing. Email has had a profound effect on the way people keep in touch. Communications are shorter and more frequent than when letters were the norm and response time has greatly diminished. Instant messaging has created another method of interaction, one where the length of messages is shorter and the style of the interaction is more conversational. Broadcast technologies like Twitter transform these short bursts of communication from oneon-one conversations to little news (or trivia) programs:

ISSN: 3064-8270

Artificial Intelligence, Machine Learning, and Data Science Journal

Research Article

which we can tune in whenever we want an update or have something to say. The traditional data mining techniques classifies the patterns in structured data for example, classification and prediction, association analysis, outlier analysis and cluster analysis. On the other hand, the newer techniques identify patterns from unstructured and structured data.

3. METHODOLOGY

The system proposed by this study will be implemented using a deep learning technique to monitor social media and discussion forums for suspicious feedback and comments.

This study made use of data collected from various online forums. These data were then transformed into to the CSV file format. On the other side of this method, users receive their account and access data from their website. The users need to log in to this system to start a discussion on any topic, and whenever any of the suspicious keywords are detected, the administrator will be notified, and the user will be warned about their activities.

An integrated approach of NLP and SVM was applied to detect suspicious topics in online forums. The support vector machine (SVM) is a statistical learning concept used as a classification and regression model to distinguish keywords based on suspicious activity from real information. Natural language processing (NLP) refers to the branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

The technique uses the SVM and NLP implemented using the Python programming language.

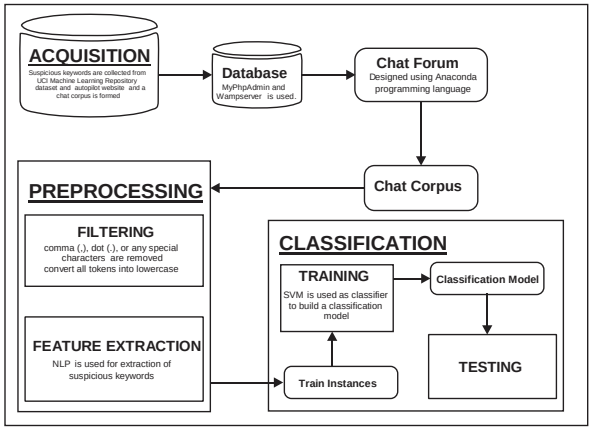

Figure 3.1: Diagram of The Model

ISSN: 3064-8270

Artificial Intelligence, Machine Learning, and Data Science Journal

Research Article

4. RESULTS AND DISCUSSION

The main aim of this research is to design a technique for monitoring suspicious discussions online forums. This application collects the postings and comments from the discussion sites and analyses those comments using machine learning techniques and algorithm.

4.1. Testing and Performance Evaluation

Testing is a technique for finding errors in source code by running it. In order to implement a system, it is necessary to test it first. A software should be thoroughly tested to identify any flaws. Testing is done in tandem with the development of the application, and the more the application is developed, the more testing is required. We can decrease the likelihood of an error or malfunction in the software by testing it. Many parameters and function implementations may change, causing the system to fail to perform as it should, even if the code compiles without problems[1].

We should additionally test the system to ensure that no functions were overlooked throughout the project's execution. Testing takes place once the project is completed. There are many different types of tests to guarantee that the system is completely functional and performing as intended. The unit testing and integration testing methods were used in this study.

Unit testing involves testing an individual component or unit. It is done as part of unit testing phase of software development life cycle and can be done in two phases. First, the correct names of the applications in the network was checked, then the two-way connection between the server and the client also checked. It is also possible to test top-down or bottom-up and then isolate the results through unit testing.

After the unit tests, the integration tests were conducted. The basic purpose of integration testing is to see if the modules can work well together, that is, to test the interfaces between them. We need to do a full test once the links between the modules are formed. Once the application is complete, system tests will be performed.

4.2. Evaluation Using System Usability Scale (SUS)

The new system was implemented with different group of users and was subjected to System Usability Scale (SUS) to determine its level of satisfaction, effectiveness and efficiency. The SUS consists of ten (10) standardized question based on Likert Scale where Strongly Disagree = 1, Disagree = 2, Agree = 3, Strongly Agree = 4.

SUS uses complex scoring system because it comprises of five (5) positive odd numbered questions and five (5) even negative numbered questions.

SUS score = (X + Y) * 2.5 where

X = Add up the total score of all odd numbered questions then subtract 5 while Y = Add up the total score of all even numbered questions then subtract from 25.

ISSN: 3064-8270

Artificial Intelligence, Machine Learning, and Data Science Journal

Research Article

Table 4.1Usability Score for User's Experience

Users	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS Score	NPS
1	4	1	4	1	3	1	4	1	4	2	82.5	Passive
2	4	1	4	1	4	1	4	1	4	1	87.5	Promoter
3	3	1	4	2	4	1	4	1	4	1	82.5	Promoter
4	4	1	3	1	4	2	4	2	4	1	80.0	Promoter
5	4	1	4	2	4	1	3	2	3	1	77.5	Passive
6	4	2	4	1	4	1	4	1	4	1	85.0	Promoter
7	4	2	4	2	4	2	3	2	4	1	75.0	Passive
8	4	1	4	1	4	1	4	1	4	1	87.5	Promoter
9	4	1	4	1	4	1	4	2	4	2	82.5	Promoter
10	4	1	4	2	4	2	4	1	4	1	82.5	Promoter
11	4	2	4	2	3	2	3	2	4	1	72.5	Passive
12	3	2	3	2	4	2	4	2	4	1	72.5	Passive
13	4	1	3	1	4	1	4	2	4	1	82.5	Passive
14	4	1	4	1	4	1	4	1	4	1	87.5	Promoter
15	4	1	3	2	4	2	3	2	3	1	72.5	Passive
16	3	1	4	2	4	1	4	1	4	1	82.5	Promoter
17	4	2	4	2	3	2	3	2	4	1	72.5	Passive
18	4	2	4	2	3	2	3	2	4	1	72.5	Passive
19	4	2	3	2	4	2	4	1	3	1	75.0	Passive
20	4	1	4	1	4	1	4	1	4	1	87.5	Promoter

Mean SUS Score = $\frac{\text{Sum of all SUS Scores}}{\text{Number of Users}}$

Sum of all SUS Scores for all Users = 82.5 + 87.5 + 82.5 + 80.0 + 77.5 + 85.0 + 75.0 + 87.5 + 82.5 + 82.5 + 72.5 + 72.5 + 82.5 + 87.5 + 72.5 + 82.5 + 72.5 + 72.5 + 75.0 + 87.5

$= \frac{1600}{20}$

The mean SUS Score = 80.0

Interpretation of Result

System Usability Scale (SUS) scores becomes meaningful by normalizing scores to produce percentile ranking. The mean SUS score (80.0) for the system was normalized into percentile ranking of 86%

ISSN: 3064-8270

Artificial Intelligence, Machine Learning, and Data Science Journal

Research Article

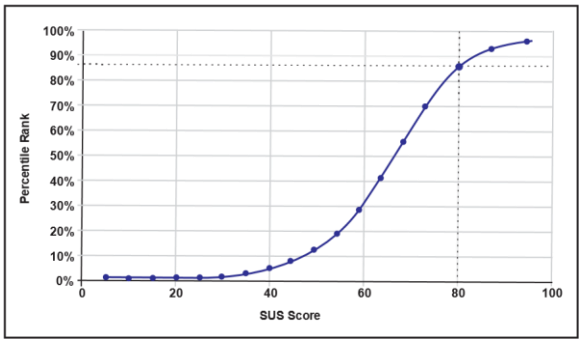

Figure 4.1: Percentile Ranking for Common SUS Scores.

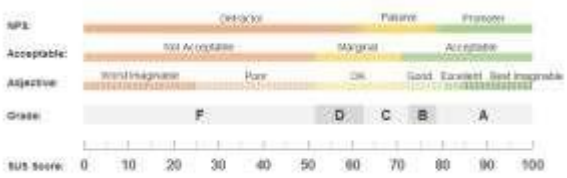

Figure 4.2: Percentiles, Grades, Adjectives, and NPS Categories to Describe Raw SUS Scores.

4.3. Comparative Evaluation Analysis

A comparative Analysis of the system has been carried out to establish the functionality of the study as described in table 4.2 while the bar chart in figure 4.3 Shows the graphical analysis of the system.

The study was compared with the Initial System and the result shows that this monitoring suspicious discussion system has an edge over the existing system.

Table 4.2 SUS Scores for Comparative Evaluation

	Chat Monitoring System	**Spam Detection System**	**Suspicious Keywords Detection System**
SUS Scores	80.0	76.2	71.2
2	87.5	70.5	70.5

ISSN: 3064-8270

Artificial Intelligence, Machine Learning, and Data Science Journal

Research Article

3	82.5	72.5	72.5
4	80.0	70.5	65.5
5	77.5	70.0	70.0
6	85.0	80.0	80.0
7	75.0	65.5	70.5
8	87.5	70.5	65.5
9	82.5	75.9	75.9
10	82.5	85.5	85.5
11	72.5	80.5	80.5
12	72.5	80.5	80.5
13	82.5	80.0	70.5
14	87.5	75.9	75.9
15	72.5	60.0	80.0
16	82.5	60.5	60.5
17	72.5	70.0	60.0
18	72.5	70.0	70.5
19	75.0	70.0	70.0
20	87.5	80.5	60.5

From the result in table 4.1 The raw and mean SUS score is 80.0 for the cybercrime monitoring system. It was normalized to percentile ranking 90. This indicate that the system was excellent and acceptable and the users were promoters. The users will not discourage others from using the proposed system while the mean SUS score for the initial system is 75.25 was normalized to percentile ranking 72. This indicate that the system is acceptable and the users were passive.

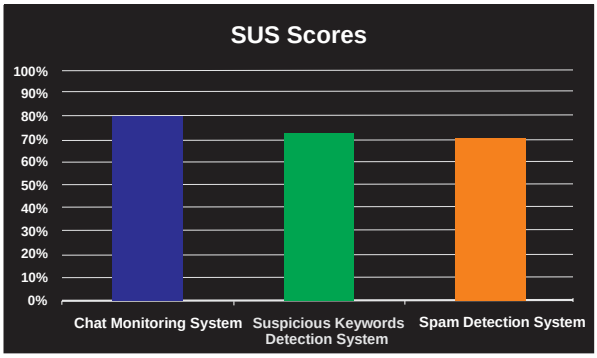

Figure 4.3: SUS Scores Bar Chat for Comparative Evaluation

5. CONTRIBUTION TO KNOWLEDGE

Having discovered the limitations of the existing systems, this research improves the performance and security of the existing systems as follows:

1. This research contributed theoretically to the existing body of knowledge through a more extensive understanding of emerging technologies, with potential for future research.
2. It also contributed essentially and experientially through developing a dependable and robust model.
3. This chat program aids in the reduction of terrorist actions by monitoring all talks in the discussion forums. This technology has the benefit of monitoring the entire chat without the user's awareness.
4. This system can make society more stable by reducing the number of crimes, and it will also provide security and protection for users.

6. Suggested Area for Further Research

In the future, the authors plan to train the model on larger datasets to improve overall performance. Suspicious text subdomains will also be considered to make the dataset more diverse. Furthermore, recurrent learning algorithms can be employed to capture the inherent sequential patterns of long texts.

References

F. Gandhi, D. Pansaniya & S. Naik, Ethical Hacking: Types of Hackers, Cyber Attacks and Security. *International Research Journal of Innovations in Engineering and Technology*, 6(1), 2022, p.28.

L.Á. Almeida, Data Model Classification Based On Machine Learning Techniques for Detection of Anomalous Traffic. *Cartagena de Indias*, 2019.

T. Dam, L.D. Klausner, D. Buhov & S. Schrittwieser, Large-Scale Analysis of PopUp Scam On Typosquatting Urls. *Proceedings of the 14th International Conference on Availability, Reliability and Security*, 2019, (pp. 1-9).

J.W. Messerschmidt, From Marx to Bonger: Socialist Writings On Women, Gender, And Crime. *Sociological Inquiry*, 58(4), 1988, pp.378-392.

J. Young, Working-Class Criminology. *Critical Criminology (Routledge Revivals)*, 2013, (pp. 79-110). Routledge.

H. Lee & K.S. Choi, Interrelationship Between Bitcoin, Ransomware, And Terrorist Activities: Criminal Opportunity Assessment Via Cyber-Routine Activities Theoretical Framework. *Victims & Offenders*, 16(3), 2021, pp.363-384.

R.A. Berk, Artificial Intelligence, Predictive Policing, And Risk Assessment for Law Enforcement. *Annual Review of Criminology*, 4, 2021, pp.209-237.