International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

A Novel Technique for Authorship Verification of Hijacked Online Social Networks User Accounts

Astha Gupta[1], Mahesh Parmar[2]

[1]Computer Science and Engineering, Madhav Institute of Technology and Science, Gwalior
asthagupta919[at]gmail.com

[2]Computer Science and Engineering, Madhav Institute of Technology and Science, Gwalior
maheshparmar[at]mitsgwalior.in

Abstract: ***The Web has a huge amount of data accessible for internet users, and a large amount of data is also produced, thanks to the development and expansion of web technology. The Internet has become an online learning platform to exchange ideas and share views. Social networking services like Twitter, Facebook, and Google+ quickly acquire popularity since they enable users to exchange opinions on issues, talk with other groups or post messages worldwide. The expanded usage of Online Social Network (OSN) has become necessary to appear to grow Authorship Verification (AV), OSN is the environment in which users can connect with other users to discuss ideas of any topics then expand data and information. AV is considered as a resource of researches and information in different ways, as is the case Sentiment Analysis (SA). In this paper, the proposed technique is compared with the previous feature extraction technique which was inefficient in providing better results comprised of the Tweets API dataset. Twitter is a popular website for social networking users posting and interacting with "tweets". The new model is henceforth capable to provide better accuracy.***

Keywords: Online social networks, hijacking attacks, Hijacked Social Media Accounts, feature extraction, machine learning, classification, LSTM.

1. Introduction

A significant number of Internet-connected individuals worldwide currently use social media (SM) and social networking sites (SNS). With the advantage of the near-immediate connection to possibly billions of other individuals, the temptation might be to connect to social media as easily and as fast as possible. Advantages are frequently inconvenient and current exposure of infringements on privacy, ID theft, and the risks of over-sharing should make social media users aware of how much information they reveal to sign up for and utilize the website's service. Facebook, Twitter, Snapchat, and Instagram are presently among the numerous SNSs accessible to customers. Each site includes the requirements for the user to read if a new user registers a site account. Usually, these agreements cannot be read or comprehended in full but many people accept these conditions and proceed to input sensitive information to establish an account. [1].

Online social networks (OSNs), for millions of Internet users, have become the mainstream cultural phenomena. Combining user-built profiles with communication mechanisms allows users to pseudo-permanently "in contact" OSNs utilize social connections in the real world and mingle our offline and online lives even more. Facebook is the third most frequentlyvisited site on the Internet [2] with 1.94 billion monthly active users since 2017. Over 313 million months active users who send tweets in more than 40 languages claim Twitter, a social microblogging site. [3].

AV is of great importance not only to gather facts in legal proceedings but also to identify deceptive intent & false news in e-commerce &SM. Social networking sites in particular have been an omnipresent way to communicate. Unfortunately, the dissemination of material from unspecified sites on these websites is also infamous. The compromises of such checking accounts are enticing and inspiring to the intruders since, without a doubt, the user network interacts positively with the demands of the hacker, in particular the account that takes five days. As a consequence, the use of misinformation, bribery, and identity theft can be a victim of criminal predators. In terms of authorship testing by computer, huge steps were taken with the introduction of Machine Learning (ML) techniques. Even so, the study of the texts on social media remains difficult. The verification of authorship is one of the most difficult tasks for categorizing the texts in a style. Given a collection of documents, the question is if the latter is either by that author, all of them by the same author, and some unknown author's documents. AV is considered as a resource of researches and information in different ways, as is the case Sentiment Analysis (SA).The basis of sentiment analysis and opinion mining is these social networking platforms. Studying the views can lead to decision-making or marketing advantages. Since the hypothesis of a human brain can be conveniently estimated by examined his updates on these web pages. Now, so the ML algorithm can be used to extract the inference of the public opinion, such as LR, &SVM, the algorithm of XG Boost & NBS algorithms [4].

There are various social network attacks, which make the users published data at risk. In my literature review, few attacks have been chosen for review and survey to have a look at various social network attacks. For this, firstly several attacks have been grouped in social networks. The groups are classified into five different categories namely Network Structural Attacks, Privacy Attacks, Social Media Attacks, Social Networking Sites Attacks, and Modern Attacks. The attacker takes full control of the user's communication. An example of an airplane, where the hijacker takes control of a flight and masquerades himself, and establishes a connection between two entities. If the

attacker gets successful in cracking the password of the user account, then they can hijack the account. Session Hijacking Attack in which the attacker target to exploit the computer system session to get information of user's computer data. In Cookie Hijacking the attacker exploits the computer session. This cookie is used to establish a temporary session that can be easily taken by a malicious attacker. The attacker tries to capture the HyperText Transport Protocol (HTTP) headers. These headers contain the session cookies, the attacker copies the HTTP session to attain the retrieval of targeted people's accounts to get private data and info of users [5-7].

The remaining work is summarized as follows: Section II summarizes related works; Section IIIintroducesa thorough summary of proposed methodology; Section IV providesdetailed simulation results achieved, & Section VI concludes the paper.

2. Literature Review

N. R. Fatahillah et al. (2017) Investigation is necessary to categorize tweets including positive and negative orutterances using a naïve Bayes classification algorithm. The findings of this research are integrated into a Twitter-classifying system. The system is developed utilizing the js Node technology with NB classification as the classification computation technique. The greatest accuracy produced by NBC systems is 93 percent based on the tests conducted [8].

E. V. Altay and B. Alatas (2018) In this research the use of technologies for NLP & approaches for ML was used to detect cyberbullying using the following: Bayesian logistic regression, random forest algorithms, multilayer sensor, J48 algorithms & SVMs. From our understanding, the achievements of these algorithms are first evaluated with the actual data using various metrics in experimental tests. [9].

K. Indira et al. (2019) A fundamental image of location prediction using tweets is explored in the suggested framework. In particular, the location of tweets is predicted from tweets. When describing tweet contents and circumstances, how problems depend on these text inputs is essential. In this study, they estimate the location of users using ML methods such as NB, SVM & DT from the tweet content. [10].

Ş. Genç and E. Surer (2019) In this research, they try to identify clickbaits on Twitter postings in Turkish news. To this end, the news headlines of Limon Haber1 and Spoiler Haber2's Twitter accounts were gathered for clickbait data &Evrensel Newspapers3 and Diken Newspaper4's Twitter accounts for nonclickbait data. Extensive experiments on headlines indicate that our model achieves clickbait identification using an ANN with an accuracy of 0.91 with an F1 score of 0.91, the highest score in the Turkish data set. [11].

B. Boenninghoff et al. (2019) In this study, suggest a novel topology for neural network similitudes that considerably improves the effectiveness of the verification task of the author using such difficult data sets. This problem has been subject to numerous effective technological methods, most of which are based on conventional linguistic characteristics such as n-grams. For some kinds of written materials such as books & novels, these algorithms provide excellent results. Forensic authorship checks for social media however are considerably more difficult since communications with a wide range of various genres and themes tend to be very brief. There has been little success with conventional techniques based on characteristics such as n-gram [12].

M. M. M. Hlaing & N. S. M. Kham (2020) In this paper, propose a method for the detection of fake social media news including both news social media and online contexts. They utilize the synonymous extraction technique and three classifiers based on a multidimensional dataset. Testing findings demonstrate the efficiency in defining news authenticity in online news media [13].

M. R. Alam et al. (2020) Used the Sentiment Analysis to evaluate the position's persuasion. They developed a model to categorize the Bangla posts in several sectors utilizing techniques SVM, KNN, DT, RF, LR. They used the algorithm that gives the best dependable performance in classifying the social media post in English [14]

A. Kesarwani (2020) Developers need to recognize fake or authentic news more effectively. The unique characteristic of identifying fake media news makes current detection algorithms unsuitable or suitable. Secondary data is thereafter important to examine. Secondary data may include social activities on social media for users. In this study, therefore, with the support of the K-Nearest neighbor classification, they propose a simple method for identifying fake news on social media. Approximately 79% of the categorization accuracy was evaluated on Facebook news posts. [15].

3. Research Methodology

a) Problem Identification

Numerous methods have been developed, including using machine learning methods, to resolve the issue of authorship verification. These methods have helped identify the distinguishing manner a character communicates or writes. Authorship verification as the name suggests itself, verifying whether a tweet was posted by the original author or their bots. In this, we mainly focus on the stylometry of the author and the bot. Stylometry works on the assumption that every author has a specific style of writing and it has some specific feature. The problem in this is that can we make the model learn about the stylometry of the author. We did it using NLP techniques to extract features then concatenate them to form one and then pass them to the model to learn from it.

b) Proposed Methodology

In the proposed methodology, the data is collected from Twitter API. In previous work various machine learning algorithms were applied in which feature extraction techniques were also applied, but due to some limitations in these techniques, the result achieved were not upto the mark, therefore a new technique of word embedding has been introduced followed by Ngram which is further classified using classification model LSTM. LSTM revolutionized the two areas of ML & neurocomputing. The extraction stage consists of two steps: first, choose the best characteristics,

then extract the n-gram depending on the chosen features. Selecting excellent features seeks to enhance the quality of n-gram extraction and decrease the computational complexity and noise.

1) Preprocessing

Preprocessing is the overall term for all the transformation of the data, including centering, normalization, rotation, shifting, shear, etc., before being transformed into the model. In preprocessing, the data has been cleaned by removing punctuations, hashtags, and emojis to obtain the text only.

2) Feature extraction

ML algorithms learn of a predefined set of training data characteristics to generate test data output. But the fundamental issue with language processing is that ML algorithms cannot operate directly on the raw text. We require certain methods of extraction to transform the text into a matrix (or vector) of characteristics. FE is a kind of reduction in dimensionality where a high number of picture pixels are successfully shown so that interesting image components are effectively recorded. Some of the most common extraction techniques include:

Natural language processing (NLP): It is an area of languages, computer science, & AI which focuses on computer-human language interactions, particularly on how computers are programmed for the processing and analysis of vast quantities of natural language data. The outcome is a computer able to "understand" the contents of texts, such as the language's contextual complexities. The system can correctly extract information and insights from the papers, classify and arrange the documents themselves.

Therefore, by analyzing **N-grams** (mainly bigrams) we may solve this issue instead of individual words (i.e. unigrams). This may maintain local word ordering. If we take into consideration all possible bigrams in the reviews provided, we can always eliminate **N-grams** with high frequency, since they are present in nearly all texts. These high-frequency N-grams are usually referred to as articles, determiners, etc.

Word embedding is one of the vector space models representing information. Word embedding retains contexts & word connections so that related words are more correctly detected. **Word embedding** includes several different implementations like **word2vec**, GloVe, FastText, and so on. Word2vec is one of the most common words embedding implementations that Google developed in 2013. It discusses word embedding using two-layer, shallow NNs to detect contextual significance. **Word2vec** is excellent at grouping related words and generating very precise estimates about the significance of words based on contexts.

3) Long Short-Term Memory (LSTM)

LSTM networks are a kind of RNN that may be dependent on the order in sequence issues. LSTMs are a complicated field of deep learning. It may be difficult to understand what LSTMs are and how bidirectional and sequence-to-sequence terminology connect to the area. LSTM has been used primarily to represent long-term connections. The GRU is given in this section as a modification of the LSTM cell before additional LSTM network designs are described. GRU was designing time series to provide a method to enhance the capacity to prevent long-term dependency by improving short-term information integration.

LSTM cells can add or remove cell state information by using various gates within a cell.Gates allow data to enter the cell state or prevents it from accessing the cell state with the aid of the multiplication and sigmoid NN layer.

A sigmoid layer generates a no.among 0 & 1 that determines how much information is to be allowed through the gate. Output value near 0 would not let something through a gate, while information is allowed by a value close to 1.

c) Proposed Algorithm

Step 1. Collecting the dataset, and separating it in train and test CSV manually.
Step 2. Perform EDA on the dataset to analyze the dataset then fining the dataset by removing punctuation, hashtags, emojis, etc.., then again visualize the data.
Step 3. Extracting features like Syntactic Features and semantic features.
Step 4. Merging the feature and it into one feature pack.
Step 5. Creating a hybrid model of LSTM and GRU model for the verification task, passing the dataset to the model. Let it train for a while.
Step 6. After training the model calculating the performance of the model by applying it to the test set.

d) Proposed Flowchart

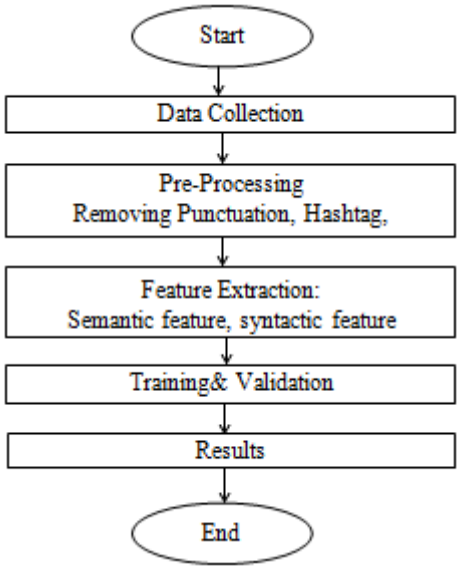

Figure 1: Flow Chart of proposed Methodology

4. Results and Discussion

The proposed methodology in this research has been implemented in python 3.0 using the Tweets API dataset. The result is evaluated using various performance parameters which are accuracy, precision, recall, and F-measure.

A. Performance Parameters

True Negatives (TN) -The values estimated as negative (meaning the real class value is 'no' and forecasted class value is 'not') are perfectly accurate. E.g., when a real class shows that this passenger hasn't survived &the forecasted class tells you the same. The real class occurs when the predicted class, a false positive as well as a false negative occur in opposition.

True Positives (TP) -The real class value and the predicted class value are both correct. Consider, for instance, the difference between actual class value, which is "has survived this passenger," and predictive class value, which tells you, "This passenger is likely to be the same one next time."

False Negatives (FN) –Real class is yes, whereas forecasted class is not. In other words, for example, when we see how valuable each passenger class is, it may tell us that passengers have survived or that passengers are likely to die.

False Positives (FP) –In the situation when the class is "No" and the forecasted class is "Yes," In other words, if the class real says that this passenger has not lived, but the class forecast predicts that he will, this passenger has died.

Accuracy -It is a fundamental accuracy indicator that is proportional to the total measurements. Symmetrical datasets provide better statistical accuracy since the proportion of false negatives and false positives are almost equal.

$$\text{Accuracy} = TP+TN/TP+FP+FN+TN$$

Precision - It is the ratio of positive observations properly predicted to total positive observations forecast.

$$\text{Precision} = TP/TP+FP$$

Recall (Sensitivity)- It is a ratio of positive comments that were accurately predicted against all actual class observations - yes.

$$\text{Recall} = TP/TP+FN$$

F1 score - It is recall & precision weighted average. This score, therefore, takes into account both false negatives & false positives.

$$\text{F1 Score} = 2*(\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$$

Table 1 and figure 2 represent the comparison values of the base and proposed results. The proposed results used four parameters like Accuracy, Precision, Recall, and F-Score. These parameters formulas give better results in comparison to previous results.

Table 1: Comparison table of base and propose results

Result	Base	Propose
Accuracy	81.11	98.65
Precision	80.66	98.74
Recall	80.32	98.78
F-Score	81.65	98.76

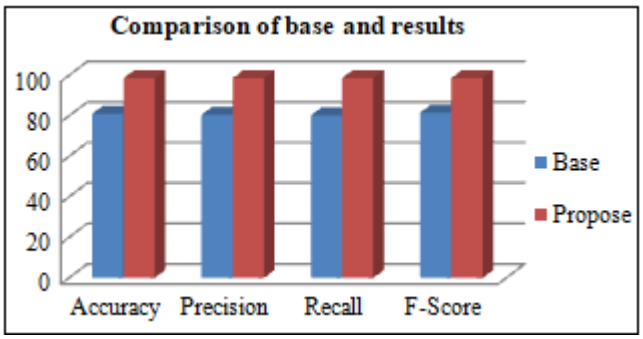

Figure 2: Comparison of base and proposed technique

5. Conclusion

It is now possible for anybody on the globe to communicate their thoughts and ideas via micro-blogging sites such as Twitter, Facebook, or blogs so on. In this context, a new, human-compromised mechanism for the verification of authorship for hacked social media accounts is introduced. Major textual features are extracted from a Twitter-based dataset. The above-proposed model is constructed in Python 3.0 in which the feature extraction algorithm Word2vec is applied followed by the Ngram model. Further LSTM model is applied for classification purposes which were compared by the previous work in which Bag of words was used for feature extraction but had some limitations which were overcome by the new proposed model.

References

[1] Aljohani, M., Nisbet, A., &Blincoe, K. (2016). A survey of social media users privacy settings & information disclosure. In Johnstone, M. (Ed.). (2016). The Proceedings of 14th Australian Information Security Management Conference, 5-6 December 2016, Edith Cowan University, Perth, Western Australia. (pp.67-75).

[2] Zephoria, The Top 20 Valuable Facebook Statistics, Zephoria2017, URL: https: //zephoria.com/top-15-valuable-facebook-statistics/.

[3] Twitter, Twitter Usage/Company Facts, Twitter2017, URL: https://about. twitter.com/company

[4] NektariaPotha and EfstathiosStamatatos," A Profile-Based Method for Authorship Verification", SETN 2014, LNAI 8445, pp. 313–326, 2014. © Springer International Publishing Switzerland 2014

[5] Rafeef Kareem, "Fake Profiles Types of Online Social Networks: A Survey", International Journal of Engineering & Technology, 7 (4.19) (2018) 919-925

[6] R. Ganguli, A. Mehta, and S. Sen, "A Survey on Machine Learning Methodologies in Social Network Analysis," 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2020, pp. 484-489, DOI: 10.1109/ICRITO48877.2020.9197984

[7] Mr. M. Sathish Kumar#1 and Dr. B.Indrani," A Study on Web Hijacking Techniques and Browser Attacks", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 5 (2018) pp. 2614-2618

[8] N. R. Fatahillah, P. Suryati, and C. Haryawan, "Implementation of Naive Bayes classifier algorithm

on social media (Twitter) to the teaching of Indonesian hate speech," 2017 International Conference on Sustainable Information Engineering and Technology (SIET), 2017, pp. 128-131, DOI: 10.1109/SIET.2017.8304122.

[9] E. V. Altay and B. Alatas, "Detection of Cyberbullying in Social Networks Using Machine Learning Methods," 2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT), 2018, pp. 87-91, DOI: 10.1109/IBIGDELFT.2018.8625321.

[10] K. Indira, E. Brumancia, P. S. Kumar and S. P. T. Reddy, "Location prediction on Twitter using machine learning Techniques," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019, pp. 700-703, DOI: 10.1109/ICOEI.2019.8862768.

[11] Ş. Genç and E. Surer, "Detecting “Clickbait” News on Social Media Using Machine Learning Algorithms," 2019 27th Signal Processing and Communications Applications Conference (SIU), 2019, pp. 1-4, DOI: 10.1109/SIU.2019.8806257.

[12] B. Boenninghoff, R. M. Nickel, S. Zeiler, and D. Kolossa, "Similarity Learning for Authorship Verification in Social Media," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2019, pp. 2457-2461, DOI: 10.1109/ICASSP.2019.8683405.

[13] M. M. M. Hlaing and N. S. M. Kham, "Defining News Authenticity on Social Media Using Machine Learning Approach," 2020 IEEE Conference on Computer Applications(ICCA), 2020, pp. 1-6, DOI: 10.1109/ICCA49400.2020.9022837.

[14] M. R. Alam, A. Akter, M. A. Shafin, M. M. Hasan and A. Mahmud, "Social Media Content Categorization Using Supervised Based Machine Learning Methods and Natural Language Processing in Bangla Language," 2020 11th International Conference on Electrical and Computer Engineering (ICECE), 2020, pp. 270-273, DOI: 10.1109/ICECE51571.2020.9393095.

[15] A. Kesarwani, S. S. Chauhan and A. R. Nair, "Fake News Detection on Social Media using K-Nearest Neighbor Classifier," 2020 International Conference on Advances in Computing and Communication Engineering (ICACCE), 2020, pp. 1-4, DOI: 10.1109/ICACCE49060.2020.9154997.