International Journal of Informatics and Computation (IJICOM)
Vol. 4, No.1, August 2022
ISSN: 2685-8711, E-ISSN: 2714-5263 DOI : 10.35842/ijicom

Re-Fake: Fake Account Classification in OSN Using RNN

Romana Herlinda

Abstract
Online Social Network (OSN) is an application for enabling public communication and sharing information. However, the fake account in the OSN can spread false information with an unknown source. It is a challenging task to detect malicious accounts in a large OSN system. The existence of fake accounts or unknown accounts on OSN can be a severe issue in data privacy-preserving. Various communities have proposed many techniques to deal with fake accounts in OSN, including rule-based black-white technique until learning approaches. Therefore, in this study we propose a classification model using the RNN to detect fake accounts accurately and effectively. We conduct this study in several steps, including gathering datasets, pre-processing, extraction, training our models using RNN. Based on the experiment result, our proposed model can produce a higher accuracy than the conventional learning model.

Keywords:
Classification, Fake Accounts, Recurrent Neural Network, Deep Learning

1. Introduction

The rise of Online Social Network (OSN) has sparked the unprecedented circulation of false information among the public [1]. Fake accounts on OSN distribute fake news. Anonymous, fictitious, and other cryptic accounts are used by individuals who write, express themselves, utilize social media, and engage in other activities in cyberspace without revealing their true identities to others. The result of false information disseminated by fake accounts can mislead people who rely on OSN as a medium for obtaining information, and misinformation can occur among OSN users [2].

Growing OSN applications can also open a vast opportunity for certain people to commit fraud using fake accounts [3]. One of the frauds in OSN is misusing the identity of a person or agency, which will later be used to commit crimes such as buying and selling goods or doing business. This fraud can cause misunderstanding between users in buying and selling goods and doing business so that many people or users must be careful in making online transactions on OSN [4].

Spam comments have promotional purposes or are considered contextually irrelevant, often found on OSN [5]. As an example of spam comments on the public figure's Instagram OSN page, these spam comments can be in the form of posts with nothing to do with the posts and status on the OSN page concerned, such as sending unwanted information by users [6]. Spam activities are annoying because they can cause misleading information on the number of comments and disrupt the flow of discussion in status to have difficulty

Corresponding Author:
Romana Herlinda, Universitas Tanjungpura Pontianak, romanahherlinda@gmail.com

finding information. Moreover, the spam commenter on OSN is a fake account whose identity is unknown [7].

Cyberbullying is also one that deserves attention from the impact of fake accounts on OSN as a harassment action using technology [8]. The activity is usually carried out on OSN in the form of malicious comments, posting pictures, or videos meant to hurt or humiliate others. This has a huge impact on both the community and OSN users because it can cause panic that leads to death. In addition, cyberbullying can also cause mental decline for users who are victims [9].

To deal with malicious account problems, several papers proposed a learning approach to producing better accuracy. A paper discussed is the application of deep learning using GitSec in online developer communities such as GitHub, this is done because this community is open to the public, which can make it vulnerable to various types of malicious attacks carried out by fake accounts such as spam, fraud, and the spread of false information [10]. In addition, CNN and LSTM algorithms are also applied to classify fake accounts on OSN. Deep learning to OSN is carried out because information or news spread by anonymous accounts circulates very quickly, thus making users have difficulty finding or knowing the truth of news or information [11].

Therefore, to solve the above problem, we propose deep learning to build a classification model. It will be research to make it easier to detect fake accounts on OSN. The follows the study's contribution:

1. We construct a novel classification model to identify fake accounts on OSN. To measure the model performance, we conduct the training and testing process and calculate accuracy and loss.
2. We conduct evaluation metrics in detecting fake accounts on OSN and present a graph to prove the quality of the model. In this study, we utilize a fake account dataset to produce a classification model. The model processes the dataset as input and uses the model to make it easier for the public to detect fake accounts.
3. By using a learning algorithm, we present an efficient classification model using the deep learning method. Instead of using conventional methods, the proposed model can detect fake accounts on OSN accurately.

Organization: The remainder of this paper will write as follows: Part II delves further into related research. Part III outlines how defined this study’s issue. Part IV discusses the experimental setup, including feature learning approaches, data sets, and data pre-processing, while Part V gives the study's findings and extensive analysis. Finally, part VI summarizes the findings and identifies several unsolved issues in the research of fake account categorization.

2. Related Works

Nowadays, several articles proposed various studies, such as research using the Naïve Bayes algorithm obtained results with reasonable accuracy by simply pre-processing datasets using discrete techniques on selected features [12]. Another study discussed SVM algorithms using fewer features but can produce accurately [13]. A paper also explored bagging algorithms to produce an accurate result, with low errors [14].

Classification of fake accounts using DL algorithms gets accurate and significant results to create protection in OSN [15]. To detect fake accounts, other researchers also use SVM and neural networks. The SVM method achieved an F1 score of 86%, and neural networks achieved an F1 score of 95% [16]. Another study used the OWL and SWRL methods to classify fake accounts. In contrast, the results of this study were able to precisely identify the wrong account with accuracy (97%) and the results of spam or fake follower bots with an accuracy rate (94.9%) [17].

A paper also explored classification using the elbow method. Applying the component analysis principle to classify fake accounts gets results that show the accuracy of 99.6% success rate and 0% failure rate [18]. Another study also discussed the algorithm Sybil Walk that can produce more accurate results than the existing random walk-based method Sybil Walk achieved a false positive ratio of 1.3% and a negative rate of 17.3% [19]. In other studies, classification can detect and accurately identify fake accounts on OSN misusing for crimes [20].

Conventional approaches proposed the SVM and CNB methods which can produce the result with dataset Facebook. SVM shows 97% accuracy, and CNB shows 95% accuracy for identifying fake accounts based on BOW [21]. Another paper also proposed a Mandatory Unique Identification Model (CUIM) method to help improve security and effectively deal with fake account holders [22]. In another study, the classification of fake accounts used the SVM, RF, and NN methods. SVM shows higher prediction accuracy performance in classifying fake accounts than RF and NN [23].

Current papers explored fake account classification using CNN. The study can produce an accurate result with ROC 0.9500 - 0.9590 and AUC 0.9547 [24]. Another article presented CNN techniques based on matching pin purposes and board names. Obtained 886444 pins in 3920 accounts, 1503 fake accounts containing 345000 pins with a classification accuracy of 90.25% [25]. A paper that discussed the classification of fake accounts produces good accuracy in detecting and identifying fake accounts [26].

In another article, the classification model using CNN and RNN can classify fake accounts on Twitter with an accuracy 82% [27]. A study employed a hybrid deep learning model that links CNN and RNN to build a fake account classification model, and it can achieve significant results [28]. Using the learning method to establish a protection scheme is one of the most intensive research trends in network security. It opens wide opportunities to overcome the constraints of traditional machine learning methods. In traditional machine learning algorithms, the features are extracted by humans [29].

Therefore, we propose a new model to solve false account detection in OSN by training large features using the RNN algorithm. Thus, not only to detect fake accounts but also to assist users in detecting fake accounts accurately. Thus, this model might be used as material for further research on the classification of fake accounts.

3. Proposed Method

This section will provide a formal statement of the experiment issue and some ideas discussed in this article.

A. Problem Definitions

Our research focuses on detecting fake accounts on OSN. To use fake accounts on OSN, we construct a fake account classification model using the RNN algorithm. First, the data is divided into two classes, and the information represents the feature (a) and bias (b). The classification process does not use data on functions that have parameters. Then the function will calculate the weight of each feature in the vector by multiplying it by the parameter. Thus, equation (1) can rewrite as equation (2), where a is the i-element of vector a, which has a range ∞.

Table 1: Contains the variables and information of the RNN algorithm

Variables	Information
x	Input
s	Hidden Layer
y	Output
x_t	Input at the t-time
x_{t-1}	Input at the previous time
s_t	Hidden State at the t-time
s_{t-1}	Hidden State at a previous time
y_t	Output at the t-time

y_{t-1}	Output at a previous time
f	Iteration relationship with activation function
w	Parameter matrix and vector
a	Feature
b	Bias

$$f(a) = a.w + b \tag{1}$$

$$f(a) = a_1 w_1 + a_1 w_2 + \cdots + a_N w_N + b \tag{2}$$

In our study, we used functional regression for the classification of false accounts in OSN. Because this regression function will produce a constant value, the threshold is used, or a specific value limit will be provided. Such as, $f(a) > threshold$ if put into the first class and vice versa $f(a) \leq threshold$ put into the second class The threshold approach is performed by changing the process to -1, and 1 as output (equation 4), where -1 represents the input classified into the first class and the value 1 indicates the input classed into the second class, using the sign function (Equation 3).

$$sgn(a) = \begin{cases} -1 \; if \; a \; < 0 \\ 0 \; if \; a = 0 \\ 1 \; if \; a \; > 0 \end{cases} \tag{3}$$

$$Output = sgn(f(a)) \tag{4}$$

B. Proposed Method

This study utilizes RNN algorithms to build deep learning classification models to detect malicious accounts in the OSN. To construct our model, we implement the RNN algorithm to make a deep learning classification model. RNN modeling is an effective way to solve fake accounts on OSN because the ability to process it is called repeatedly with the results that it can handle input and output variables of varying length [30].

In this study, the method adopts RNN to store information from the past is by looping in its architecture, which automatically keeps information from the past stored. Repetition or looping on RNN is taking the input value x and then entering it into the RNN, which contains the value of the hidden layer, which will update every time the RNN reads a new input and output. In computation using RNN, there is f function f will depend on weight w. The weight of w will receive the new state value from the hidden layer minus 1, which becomes input in the x_t state and stored in s_t (hidden state). This process is carried out using equation 5.

$$s_t = f_w(s_{t-1}, x_t) \tag{5}$$

The value x is inserted into the activation function f and the same w weight in each calculation. Simply put, a w_{xs} the weight matrix is multiplied by the x_t input and another w_{ss}weight matrix multiplied against the value of the previously hidden layer or s_{t-1}. Both matrices are added. If there are nonlinear data, then the summation of the two matrices is multiplied by snow, as in equation 6.

$$s_t = \tanh(W_{ss}s_{t-1} + W_{xs}x_t) \quad (6)$$

RNN architecture generates several y_t at all times because there is another weight matrix w of hidden layer w_s thus changing some of the y values seen in equation 7.

$$y_t = W_{sy}s_t \quad (7)$$

4. Experimental Setup

1. Main Idea

   In the current years, deep learning has obtained good performance and wide application in computer vision. It is widely implemented to establish a protection scheme for computer networks, including the adoption of the RNN algorithm in security area research. The main idea of this study is to build a classification model by using the RNN algorithm to classify fake accounts on the OSN based on followers, name, and date of registered accounts. Using the RNN algorithm to build a classification model can achieve significant and accurate results because the ability to process it is called repeatedly [31].

2. Dataset

   In this study, we utilize the fake account dataset obtained from the github.com site. It has several features, followers, names, and recorded dates. We divide the dataset into two parts in the training and testing process, namely data training and data testing [32]. To conduct this study, we collect fake account datasets adopt in this study amounted to 2,818 samples used to research building deep learning classification models to make it easier to detect fake accounts. To build our model, we separate 80% as training data and 20% as testing data. In table 2 contains the number of training and testing datasets.

   Table 2 describes our experiment dataset to train and test the model

Dataset Label	OSN features	
	Training (80%)	Testing (20%)
Fake Account	314	78
Account	314	79

3. Data Pre-Processing

   In this study, we conduct pre-processing by using the vectorization process to change the form of data that was previously unstructured into structured. Then in the pre-processing stage, the dataset used is the fake account dataset, totaling 785 data, which will be vectorized using the tokenizer method, with feature normalization and feature selection. Finally, the tokenizer method is used in the pre-processing process to facilitate the RNN algorithm in receiving data input [33].

4. Classification Method

   In the first stage, we collect datasets with fake and benign labels before the training process. Then enter the pre-processing stage, which is the stage where the dataset vectorization will be carried out. This vectorization stage is done to facilitate building a classification model (training) using the RNN algorithm.

   We separate the dataset into two parts, 80% training, and 20% testing, to training our model. First, we must train our model to build an optimal fake account

classification model using a training dataset. Then, the model will be tested using dataset testing to determine the model's performance. Finally, to optimize the training scores, we also tune some hyperparameters to get the optimal classification model.

5. Result & Analysis

1. Classification Test

To classify fake accounts, we test a classification model with different learning parameters. In well-known, we examined several training optimizers, one of which is Adam. We also apply the concept of local variables to compute a weight loss function with various learning speeds.

In this experiment, our model produces accuracy by adjusting various hyperparameters for the highest performance. In the training and testing process, we adjusted the epoch = 50, the batch size = 64, and the learning speed 0.2. In the trials that have been done, the model can classify with an accuracy rate of 81.0%.
Table 3 describes the performance of the RNN, especially in training and testing.

Hyperparameter	Optimizer	Training Accuracy	Testing Accuracy
Epoch = 50	Adam	0.8678	0. 9363
lr = 0.0002	RMSProp	0.8742	0.9108
batch size 64 learning speed 0.2	SGD	0.8806	0. 8854

Table 4 Results of the classification with various optimizer functions.

Hyperparameter	Optimizer	Training Loss	Testing Loss
Epoch = 50	Adam	0.6807	0.6781
lr = 0.0002	RMSProp	0.3792	0.30
batch size 64 learning speed 0.2	SGD	0.3659	38 0.3563

2. Evaluation Metrics

The experiment calculates Recall, Precision, F1 Score, and Confusion Matrix to evaluate the performance model (CF) in the evaluation metrics process. This study uses Recall to identify the fraction of the data set that falls into the dangerous category and labels it. The Precision defines the degree of the data set, which is classified as dangerous but excludes malicious data. Finally, we created a balance between Precision and Recalled to locate all accounts with malicious links while tolerating poor Precision if the test results are unimportant. As a result, an F1 score is required to represent the best combination of accuracy and Recall. The dataset's Recall, Precision, and F1 Score are shown in Table 5.

Table 5. Dataset Recall, Precision, and F1 Score

Classification Model	Precision	Recall	F1-score
False	0.26	0.22	0.24
True	0.88	0.90	0.89
RNN			0.81
Macro avg	0.57	0.56	0.56
Wighted avg	0.88	0.81	0.80

3. Confusion Matrix

The proposed model can obtain the best accuracy value to represent the maximum detection rate in categorizing benign and fake accounts, based on CM calculations in Figure 1. Our approach, in particular, delivers not only great accuracy but also improved graphics performancet.

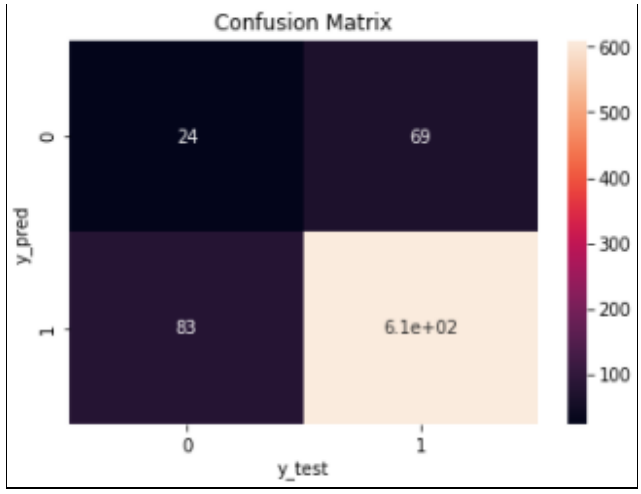

Fig 1: Confusion Matrix RNN algorithm

This paper also presents a tabular Confusion Matrix that depicts the model's performance on the known test data. The confusion matrix, in particular, contains information on True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) (FN). This is very helpful since the categorization results are often insufficiently represented in a single number.

Fig. 1 shows the evaluation of the metric with the confusion matrix (CM) using the CNN algorithm. We can see the number predicted by our classifier from the confusion matrix, separately for the two classes. In the confusion matrix, the proposed model obtained TP = 24, TN = 692, FP = 69, and FN = 83.

6. Conclusion

The detection of fake accounts on OSN is a crucial step in categorizing them. The current approach suggests employing traditional machine learning to tackle the problem. However, this is both expensive and time-consuming. This study creates a fake account categorization model using the RNN approach to improve graph performance. When applying the proposed model, we find that the RNN technique produces better accuracy and less loss. As a result, using the RNN approach to train a classification model with extraordinary hardware processing capabilities might be a viable choice.

Based on experiments, the classification of fake accounts using RNN can achieve high accuracy with small losses. The RNN accuracy value that we got is 0.81 while the F1-score value is 0.80, Recall is 0.81, and Precision is 0.88. Our proposed model can not only produce higher accuracy but also improve graphics performance. As a result, we suggest that the RNN-based classification model might be a potential approach for identifying fake OSN accounts.

Further research on the classification of fake accounts could benefit from subsequent models, which could be integrated with new approaches such as the CNN algorithm in the future that could be combined with new techniques such as creating new regulators to train networks.

Acknowledgment

This paper is conducted in the Department of Informatics, Respati University of Yogyakarta, Indonesia.

References

[1] Wilson Ceron, Mathias Felipe & Marcos G. Quiles, “Fake news agenda in the era of COVID-19: Identifying trends through fact-checking content.” Online Social Networks and Media, Vol. 21, pp 100-116, 2021.

[2] Sy.Yuliani, Shahrin Sahib, et al., “Hoax News Classification using Machine Learning Algorithms”, International Journal of Engineering and Advanced Technology, Vol.9, No.2, pp 2249-8958, 2019.

[3] Tahereh Pourhabibi et al. l, “Fraud detection: A systematic literature review of graph-based anomaly detection approaches.”, Decision Support Systems, Vol.133, No.113303, 2020.

[4] Arun Vishwanath, “Habitual Facebook Use and its Impact on Getting Deceived on Social Media”, Journal of Computer-Mediated Communication Vol. 20, No. 1, pp 83-98, 2015

[5] Shreyas Aiyara, Nisha P Shetty, “N-Gram Assisted Youtube Spam Comment Detection”, International Conference on Computational Intelligence and Data Science (ICCIDS), Vol.132, pp 174-182, 2018.

[6] Zulfikar Aloma, Barbara Carminatib & Elena Ferrari, “A deep learning model for Twitter spam detection”, Vol.18, No. 100079, 2020.

[7] Zakia Zaman, Sadia Sharmin, “Spam Detection in Social Media Employing Machine Learning Tool for Text Mining.”, IEEE International Conference on Signal-Image Technology and Internet Based Sys, Vol.978, No.1, pp 4283-5386, 2018.

[8] W.Akram, R.Kumar, “A Study on Positive and Negative Effects of Social Media on Society”, International Journal of Computer Sciences and Engineering, Vol.5, No.10, pp 2347-2693, 2017.

[9] Batol et all, “*Cyberbullying Detection:A Survey On Mltilingual Techniques .”*

IEEE European Modelling Symposium, pp 2473-3539, 2017.
[10] Qingyuan Gong, et al." Detecting Malicious Accounts in Online Developer Communities Using Deep Learning.", In The 28th ACM International Conference on Information and Knowledge Management (CIKM '19), 2019.
[11] Muhammad Umer, et al." Fake News Stance Detection Using Deep Learning Architecture (CNN-LSTM)", IEEE, Vol.8, 2020.
[12] Ersahin Buket et all, "Twitter Fake Account Detection", IEEE International Conference on Computer Science and Engineering, pp 388–392, 2017.
[13] Sarah Khaled, Hoda M. O. Mokhtar & Neamat El-Tazi, "Detecting Fake Accounts on Social Media.", IEEE International Conference on Big Data, 2018.
[14] Saeid Sheikhi, "An Efficient Method for Detection of Fake Accounts on the Instagram Platform.", Revue D Intelligence Artificielle, Vol.34, No.4, pp 429-436, 2020.
[15] Putra Wanda, Marselina Endah Hiswati , Huang J. Jie, "DeepOSN: Bringing deep learning as malicious detection scheme in online social network.", Journal of Information Security and Applications, 52, 2020.
[16] Fatih Cagatay Akyon , Esat Kalfaoglu, "Instagram Fake and Automated Account Detection Insagram Sahte ve Otomatik Hesap Kullanımı Tespiti.", IEEE, Vol.978, No.1, pp 7281-2868, 2019.
[17] Mohammed Jabardi "Twitter Fake Account Detection and Classification using Ontological Engineering and Semantic Web Rule Language.", Karbala International Journal of Modern Science, Vol. 6, No. 4, pp 404-413, 2020.
[18] Mohammadreza Mohammadrezaei , Mohammad Ebrahim Shiri & Amir Masoud Rahmani, "Detection of fake accounts in social networks based on One Class Classification.", The ISC Int'l Journal of Information Security, Vol.11, No.2, pp 1–12, 2019.
[19] Mohammadreza Mohammadrezaei , Mohammad Ebrahim Shiri & Amir Masoud Rahmani, "Detection of fake accounts in social networks based on One Class Classification.", The ISC Int'l Journal of Information Security, Vol.11, No.2, pp 1–12, 2019.
[20] Disha Agarwal, Atrakesh Pandey, "Determining Fake Accounts on Facebook", International Journal of Management, Technology And Engineering, Vol. IX, No.IV, pp 2249-7455, 2019.
[21] Putra Wanda, Huang Jin Jie, "DeepProfile: Finding fake profile in online social network using dynamic CNN.", Journal of Information Security and Applications, Vol. 52, 2020.
[22] Enas Elgeldawi et all, "Detection And Characterization Of Fake Accounts On The Pinterest Social Network.", International Journal of Computer Networking, Wireless and Mobile Communications, Vol. 4, No. 3, 2250-1568, 2278-9448, 2014.
[23] Jia Jinyuan, Wang Binghui & Gong Neil Zhenqiang, "Random Walk Based Fake Account Detection in Online Social Networks." IEEE International Conference on Dependable Systems and Networks, pp 273–284, 2017.
[24] Rohit Raturi, "Machine Learning Implementation for Identifying Fake Accounts in Social Network", International Journal of Pure and Applied Mathematics, Vol. 118, No. 20, pp 4785-4797, 2018.
[25] Saumya Batham et all, "CUIM: An Approach to deal with Fake Accounts on facebook", International Conference on Emerging Research in Computing Information Communication and Applications (ERCICA-13), Elsevier, 2014.
[26] dr.K.Sreenivasa Rao, dr.G.Sreeram & dr. B.Deevena Raju, "Detecting Fake Account On Social Media Using Machine Learning Algorithms", International Journal Of Control And Automation, Vol. 13, No. 1, pp 95-100, 2020.
[27] Ajao, et al, "Fake News Identification on Twitter with Hybrid CNN and RNN Models.", Proceedings of the 9th International Conference on Social Media and Society - SMSociety '18, pp 226–230, 2018.
[28] Osama et al," Fake news detection: A hybrid CNN-RNN based deep learning approach", International Journal of Information Management Data Insights, Vol.1, No.1, 2021.
[29] Imamverdiyev, Yadigar N. and Fargana J. Abdullayeva. "Deep Learning in Cybersecurity: Challenges and Approaches." IJCWT vol.10, no.2, pp.82-105. 2020
[30] Alex Sherstinsky," Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network", Physica D: Nonlinear Phenomena, Vol.404, No.132306, pp 0167-2789, 2020.
[31] Buber, Ebubekir and Diri, Banu, "Web Page Classification Using RNN.", Procedia Computer Science, Vol.158, pp 62-72, 2019.

[32] Hongyu Liu, et all, “CNN and RNN based payload classification methods for attack detection”, Accepted Manuscript, 2018.
[33] Huang, Ji, et all,” Detecting Domain Generation Algorithms With Convolutional Neural Language Models”, IEEE International Conference On Trust, pp 1360-1367, 2018.