JOURNAL SYNTAX IDEA
p–ISSN: 2723-4339 e-ISSN: 2548-1398
Vol. 6, No. 10, Oktober 2024

COMPARISON OF DEEP LEARNING AND MACHINE LEARNING MODEL FOR PHISHING EMAIL CLASSIFICATION

Randy Himawan, Amalia Zahra
Bina Nusantara University, Indonesia
Email: randy.himawan@binus.ac.id

Abstract

Email is a medium in business communication that people use everyday and not only holds sensitive information but also identity for the user and organization. The uniqueness of information contained make phishing detection and remedy cannot be applied generally. Existing method applied in common on security software such as blacklisting keyword or email address is not enough to outpace evolving phishing method. Nowadays the implementation of machine learning and deep learning grow rapidly fast and the method used by machine learning and deep learning that can learn pattern of inputs and natural language processing making classification to detect email phishing promising, this study present proposed method of phishing mail classification by comparing the use of machine learning and deep learning model, The algorithm used in this paper are recurrent convolutional neural network and random forest with tf-idf and fasttext word embedding. The Dataset contain phishing and legitimate of an email gathered from a trading company in Indonesia. The dataset will be split into 70% for training and 30% for testing. As a result of this comparison, the random forest model with tf-idf word embedding achieve highest accuracy of 100% for the dataset used and the highest accuracy for recurrent convolutional neural network model with fasttext is 98.21%.

Keywords: Phishing, Recurrent Convolutional Neural Network, Random Forest, Term Frequency-Inverse Document Frequency, Fasttext

INTRODUCTION

Email phishing attack happened almost everytime, threaten individual, organization and business alike that can lead to financial loss and information breach. Most common attack happened is the business email compromised (BEC). Altough phishing is an old method that recorded happened in 1996 (Karim et al., 2023), it always evolving paralel with developing event and technologies. Attackers used various method to conduct the attack, most of the time they use conversation hijacking or sending bulk of email contain malicious aplication and url in order to gain financial fraud (Sheneamer, 2021; Thapa et al., 2023). Unlike malicious bulk email that security software can detect and mitigate, conversation hijacking is difficult to detect and observe by system and inexperienced user, attacker can register legitimate domain and some of them mimic legitimate sender domain (Khan et al., 2015; Sah & Parmar, 2017).

Most of email platform already implemented security measure to accomodate the problem (Fernandes et al., 2014), such as blacklisting certain keyword inside email and malicios domain but this solution is not enough due to user or organization might have different correspondence and different information context contained, in addition the sensivity contained might prevent organization and or user to share the email as a data to improve

How to cite: Randy Himawan, Amalia Zahra (2024) Comparison of Deep Learning and Machine Learning Model for Phishing Email Classification, *(06) 10,*

E-ISSN: 2684-883X

detection in current security software (Magdy et al., 2022; Somesha & Pais, 2022). The attacker usually use different email address or create new domain to repeat the attack.

From that issue, to better comprehend existing email phishing detection trend, literature studies performed. This paper demonstrate model comparison of previous works that achieve good acccuracy result to classify phishing email. The literature using recurrent convolutional neural network algorithm demonstrated by(Somesha & Pais, 2022) and the random forest algorithm with fasttext word embedding demostrated by (Lai et al., 2015). This experiment can be a base model to develop mail security system internally.

RESEARCH METHOD

Overview of the method or experiment step ilustrated on figure 1, Dataset gathered and cleaned, the next step we embed the defined feature with tf-idf or fasttext (Atawneh & Aljehani, 2023; Lian et al., 2015). The next step is classification, The classification process will be using machine learning and deep learning model. In the classifier, email subject and body processed with RCNN and RF alogrithm with TF-IDF and fasttext word embedding.

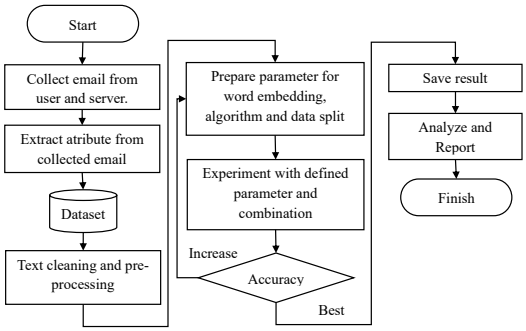

Figure 1 Experiment step

RCNN is a model developed to solve the limitations of Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) (Breiman, 2001). Looking to it's architecture RCNN is stack of recurrent convolutional layer (RCLs) that inserted with max-pooling layer when needed. The first step of RCNN model is a recurrent bi-directional that can produce less noise compared to other neural network models. With this structure, RCNN can extract more contextual information from inputs. The second step of it's sructure, RCNN applied max-pooling layer that automatically select which feature that has the most prominent role.

Random forest is a machine learning algortihm that developed by Breiman L (Karim et al., 2023). This model is a structured classifier tree like that independent and identical. Each tree contain random vector that give vote to the most popular class. Random forest used by many classification task due to the performance to process complex relation in data and pattern that not linear in multi dimension dataset (Rigatti, 2017).

Feature used in this experiment are subject and body content, the feature will be divide into 3 category such as subject only, subject with body, and body only. By separate the feature into this category we can see the most corelated feature that contribute to accuracy. The next step we process word embedding to the feature respectively using TF-IDF or fasttext, we used two word embedding as a comparison which embedding will produce highest accuracy. After word embedding the next step will be the classification using RCNN and RF algortihm.

The percentage data used for testing both model is 20% with shuffle and stratify configuration to get better balanced in data distribution for testing.

RESULT AND DISCUSSION

This section disclose the evaluation of the result obtain by using experiment step explained previously. The first experiment we use RCNN and TF-IDF word embedding, from this experiment we achieve result of 71.97% accuracy and 0.59 loss as displayed on figure 2. From the graph there is no significant increase of accuracy on bigger epoch (Ilie et al., 2021).

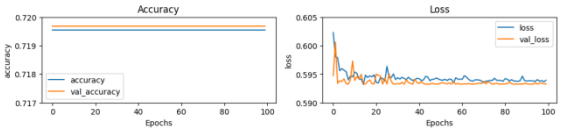

Figure 2 RCNN with TF-IDF graph result

The experiment using RCNN and subject only feature with Fasttext we got highest accuracy of 86.08% and 0.51 loss. By using fasttext there is an increase of accuracy and decreased loss than TF-IDF (Chawla et al., 2023). The using higher epoch also improve the validation accuracy score as seen on figure 3.

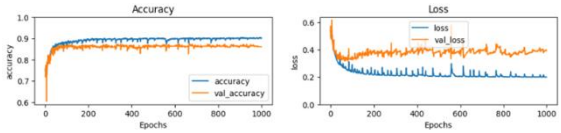

Figure 3 RCNN with fasttext graph result

Experiment using random forest and TF-IDF achieve 100% accuracy using subject as a feature and the fasttext embedding achieve 99.15% accuracy using the same feature. The detailed result of random forest experiment can be seen in table 1 below.

Table 1 Random forest experiment result

Algorithm	Embedding	Subject	Subject and Body	Body
RF	TF-IDF	100%	99.80%	95.63%
RF	Fasttext	99.15%	98.61%	94.04%

The experiment on the combination of three different feature, two word embedding and two algorithm with 2514 data can be summarized in table 1. the highest accuracy achieved by the combination of RF and TF-IDF model with subject only feature gaining 100% accuracy.

Table 2 All experiment accuracy result

Algorithm	Embedding	Subject	Subject and Body	Body
RCNN	TF-IDF	72.47%	71.97%	71.97%
RCNN	Fasttext	82.11%	96.62%	98.21%
RF	TF-IDF	100%	99.80%	95.63%
RF	Fasttext	99.15%	98.61%	94.04%

CONCLUSION

The model was experimented to classify phishing email in an business organization, once the three combination of features were determined from the mail dataset, then the classification process was applied using selected algorithm and word embedding to get the result. The model experimented gave the highest accuracy values of 100% by using random forest algorithm, TF-IDF word embedding and subject as a feature. The result for deep learning didn't achieve higher result due to supplied data might not enough to get the model trained better.

It may be a future work to study the latest phishing trend from different sources of dataset to gain better training and test result for the model, tuning the model parameter, and input support for mixed language and writings in email such as japanese, mandarin, arabic and other language.

BIBLIOGRAFI

Atawneh, S., & Aljehani, H. (2023). Phishing email detection model using deep learning. *Electronics*, *12*(20), 4261.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Chawla, S., Kaur, R., & Aggarwal, P. (2023). Text classification framework for short text based on TFIDF-FastText. *Multimedia Tools and Applications*, *82*(26), 40167–40180.

Fernandes, D. A. B., Soares, L. F. B., Gomes, J. V, Freire, M. M., & Inácio, P. R. M. (2014). Security issues in cloud environments: a survey. *International Journal of Information Security*, *13*, 113–170.

Ilie, V.-I., Truică, C.-O., Apostol, E.-S., & Paschke, A. (2021). Context-aware misinformation detection: A benchmark of deep learning architectures using word embeddings. *IEEE Access*, *9*, 162122–162146.

Karim, A., Shahroz, M., Mustofa, K., Belhaouari, S. B., & Joga, S. R. K. (2023). Phishing detection system through hybrid machine learning based on URL. *IEEE Access*, *11*, 36805–36822.

Khan, W. Z., Khan, M. K., Muhaya, F. T. Bin, Aalsalem, M. Y., & Chao, H.-C. (2015). A comprehensive study of email spam botnet detection. *IEEE Communications Surveys & Tutorials*, *17*(4), 2271–2295.

Lai, S., Xu, L., Liu, K., & Zhao, J. (2015). Recurrent convolutional neural networks for text classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, *29*(1).

Lian, C., Ruan, S., & Denœux, T. (2015). An evidential classifier based on feature selection and two-step classification strategy. *Pattern Recognition*, *48*(7), 2318–2327.

Magdy, S., Abouelseoud, Y., & Mikhail, M. (2022). Efficient spam and phishing emails

filtering based on deep learning. *Computer Networks*, *206*, 108826.
Rigatti, S. J. (2017). Random forest. *Journal of Insurance Medicine*, *47*(1), 31–39.
Sah, U. K., & Parmar, N. (2017). An approach for malicious spam detection in email with comparison of different classifiers. *International Research Journal of Engineering and Technology (IRJET)*, *4*(8), 2238–2242.
Sheneamer, A. (2021). Comparison of deep and traditional learning methods for email spam filtering. *International Journal of Advanced Computer Science and Applications*, *12*(1), 1–6.
Somesha, M., & Pais, A. R. (2022). Classification of phishing email using word embedding and machine learning techniques. *Journal of Cyber Security and Mobility*, 279–320.
Thapa, C., Tang, J. W., Abuadbba, A., Gao, Y., Camtepe, S., Nepal, S., Almashor, M., & Zheng, Y. (2023). Evaluation of federated learning in phishing email detection. *Sensors*, *23*(9), 4346.