International Journal of Engineering & Technology, 7 (3.2) (2018) 778-782

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Review Study of Hoax Email Characteristic

SY.Yuliani[1*], Shahrin Sahib[2], Mohd Faizal Abdollah[2], Mohammed Nasser Al-Mhiqani[2], Aldy Rialdy Atmadja[3]

[1]Department of Informatics, Universitas Widyatama, Bandung, Indonesia
[2]Information Security and Networking Research Group (InFORSNET), Faculty of Information Communication Technology, Universiti Teknikal Malaysia Melaka
[3]Informatics Department, Sekolah Tinggi Teknologi Garut, Garut, Indonesia
**Corresponding author E-mail: sy.yuliani@widyatma.ac.id*

Abstract

Hoax on email is one form of attack in the cyber world where an email account will be sent with fake news that has many goals to take advantage or raise the rating of sales of a product. A Hoax can affect many people by damaging the credibility of the image of a person or group. The phenomenon of this hoax would cause anxiety in the community and even more bad effects because of the potential for the wrong power of the news or information. In this paper we review the Hoax detection systems, Types of Hoax, and machine learning models that has been used to detect the Hoax. This work serves as a basis for further studies on Hoax detection systems.

Keywords: *Hoax, Fake news, Hoax detection, Hoax detection systems*

1. Introduction

Email is a tool to exchange messages and information, delivery of messages to be faster and effective and very cheap by using the internet network media. The use of email continues to grow even as other interpersonal communication methods, such as short messages, social chat and other social media. A side from being an interpersonal communication tool email is also a supporting factor in the business world, as well as to consumers, especially online sales. Based on the results of a statistical report sourced from one of the market research technology company Radicati Group, reported in 2017 the number of daily business email users reached 269 billion, which will be expected to grow and grow to reach 319.6 billion by the end of 2021[1].

Furthermore, email accounts around the world will be expected to increase slightly faster than the number of email users worldwide. This is because since everyone will have more than one email account for privacy purposes, they will differentiate email for personal and email for the public. This rapid development is due to all forms of online communication, online transactions and banking require the user to have an email address.

Hoax on email is one form of attack in the cyber world where an email account will be sent with fake news that has many goals to take advantage or raise the rating of sales of a product, which would be good news hoax via email will be made the recipient panic, then the recipient will be led to forward the email containing the hoax, the trick usually use the media group or mailing list to facilitate its spread[2].

According to research conducted by Chen, Yong and Ishak, the definition of hoax is misleading and malicious information because it can alter human perception by conveying false information as important news as true news[3,4]. A Hoax can affect many people by damaging the credibility of the image of a person or group. The phenomenon of this hoax would cause anxiety in the community and even more bad effects because of the potential for the wrong power of a news or information. In this case, the public should be more observant in sorting the information and must first be ensured accuracy before sharing or sharing information.

In this paper, the sections are divided into five sections. The first section describes the background of the research. Section 2 talks about another research that related to the detection system of a hoax. In addition, section 3 includes the literature study and the methods used to detect hoaxes.

2. Related Studies

Many kinds of research focused on the study of detecting hoax. Chen, et al describe hoax studies on his research. The problem is that misleading information is always a distortion of draft growth. Some hoaxes are made in such a way that they can be private data provided they are required for official purposes, The research proposed developing a hoax detection system by incorporating text matching method using Levenshtein Distance measure, The proposed model is used to identify text-based hoax emails. Sensitivity and specificity are used to evaluate the accuracy of the system in identifying hoax emails[3].

In 2017, Eugenio Tacchini, et al found the problem of social networking sites (SNSs) that had been about the hoaxing of information sharing by users very freely and in this study about posts can be classified as hoaxes or non-hoaxes with high accuracy [5].

Another research, Elyashar found the problem with this research is how to divide the dataset into topic categories and authenticity in online discussions, the process used to retrieve their post accounts to train traditional ML groupings, and manual labels for label accounts. The research proposed an approach for the detection of

email hoax, identifying link prediction-based features that are found useful for account classification[6].
Sirajudeen et al had found a problem with spreading a fake online news that has been identified as one of the main concerns of online abuse. This Research proposed an evaluation of the effectiveness of algorithms, able to detect and filter to reasonable degree of accuracy what constitute an online fake news, multi-layered evaluations technique to be built as an app where all information read online is associated with a tag, given a description of the facts about the contain[7].

Table **1** shows the comparison of previous research related to detect hoax information in social media, news and email.

Table 1: Shows the similarities and differences from study literature

Previous work	2017, Eugenio Tacchini, et al	2017, Aviad Elyashar, et al	2014, Yoke Yie Chen et al	2017, Sirajudeen, et al
Technique	Logistic regression and harmonic BLC algorithm	Support Vector Mesin dan Stochastic Gradient Descent	Levenshtein Distance	Support Vector Machine and Stochastic Gradient Descent
Problem	The need for hoax detection, because of the large amount of information dissemination	The online news not only produces the right facts	Update hoax database to prevent hoaxes existence	Spreading the fake news using microbloggins sites
Contribution	Classified Facebook post as hoaxes and non-hoaxes	Method for detecting fake news based on machine learning, detecting fake and legitimate account	Developing hoax detection system based on text maching algorithm	Conceptual framework for detecting a fake news
Form	Text	Text	Text	Text
Hoax Detection	Yes	Yes	Yes	Yes
Dataset	Facebook posts	News in online social media	Email	Online news
Accuracy	Logistic regression (average) : 79.4% Harmonic BLC (average) : 99.1%	Best accuracy using XG Boost : 89 %	Value of sensivity : 71 % Value of specificity : 80 %	Not described

3. Detection Email Hoax

3.1 Email

Email is a method of sending a letter through a computer network or the Internet, Email serves as a means to send letters or messages through the network Internet. Emails began to be used in the 1960s. At that time the Internet has not been formed, there is only a collection of 'mainframe' which is formed as a network. Beginning in the 1980s, electronic mail is used for the general public. Nowadays, many postal companies in various countries are declining their income because the society is no longer for postal services. Peoples prefer to choose an email then send a letter with the post office. The structure of email is described in Fig 1. The email is contained a header, message and attachment.

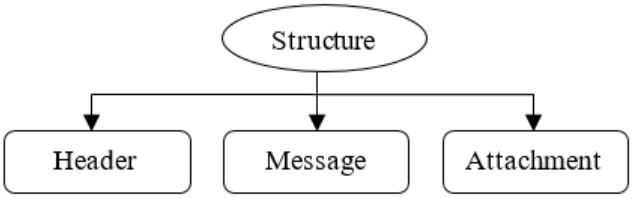

Fig 1. Structure Email

A. Header
Header, a set of lines containing information about the transmission of messages, such as the sender's address, the recipient's address, or the stamp that indicates the time when the message was sent by the intermediary server for the transport agent, acting as the electronic mail sorting office.
B. Message
The message, which consists of the following two elements, Header is a set of lines describing the message settings, such as sender, recipient, date, etc
C. Attachment
Attachments on emails containing files or documents you want to include in emails can be text, images, sounds and video.

3.2 Hoax

A hoax is an information that is not true but is made as if it is true. Based on Zannetou, Sirivianos, Blackburn, Kourtelis research, a hoax is known as biased/inaccurate news. It means a fake news is an attempt to deceive or outsmart the reader or listener to believe in something, where the false news knows that the news is not true [8]. The word "Hoax" is believed to have existed since hundreds of years ago, derived from the word "hocus" in the Hocus Pocus spell which originally is Latin hoc est corpus, meaning this is the body [2][9]. This spell is used by magicians to declare that something is true, but the fact is not necessarily true. Hoax usually in the form of warning emails, false advice or news - news that usually ends with an appeal to spread widely. A hoax can be distributed anywhere and anytime, via email, social media Facebook, Twitter, WhatsApp, line, SMS and other media.

Table 2. Sample of Email Hoaxes [10]

Example Query	Example Response (in Croatian)
To: hoax@cert.hr From: tomislav.petkovic@inet.hr Subject: Test DEAR TOMISLAV, My name is MUSTAPHA NDOH, a citizen of BOUAKI REPUBLIC OF COTE D IVOIRE. I got your contact from the Chamber of Commerce here during my search for an international business relationship. I am residing in Abidjan the capital city of Cote dIvoire. In fact I worked with the ROBERT GUEI as one of his personal confidance.During the renewed political crisis of which the he was brutally mudered by the government forces who claimed it to be the rebels who has been causing problems in my country. It was then that I moved the total sum of US $25,000,000.00 (Twenty Five Million United States Dollars) which was from the sales of cofee and cocoa and gold…	To: tomislav.petkovic@inet.hr From: hoax@cert.hr Subject: Odgovor na vasu prijavu Postovani korisnice, Ovo je automatski generirana poruka. E-mail koji ste nam proslijedili prepoznat je gotovo sigurno kao hoax "Nigerian Scam-en-ver8". Za vise informacija, pogledajte na http://www.cert.hr/hoax.html A sadrzaj uzorka po kojem je prepoznat ovaj hoax mozete pogledati na: http://www.cert.hr/hoax_indetail.php?hid=3632 Nadamo se da Vam ovaj Hoax nije prouzrocio bilo kakve neugodnosti. Molimo Vas da i ubuduce bilo kakve sumnjive poruke dobijene elektronickom postom proslijedite nama. Pozdrav, CARNet CERT www.cert.hr

Based on content-based typology, email hoaxes can be separated into five types. There are virus hoaxes, give away email hoaxes, charity email hoaxes, urban legends email hoaxes and hoaxed hoaxes[11]. The type of e wed in Fig 2.

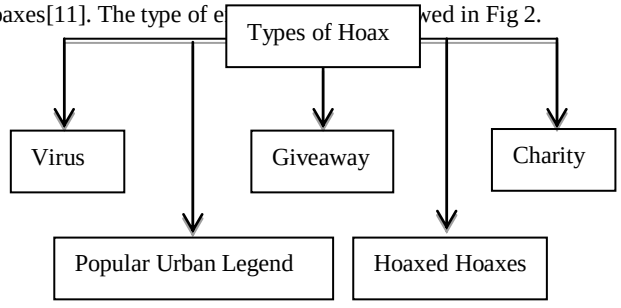

Fig 2. Type of Email Hoaxes

A. Virus hoaxes

The Hoax Virus works Quite simply, it can be assumed that this initial Hoax Email is evolving in line with hacker culture and concerns about online security. Some of the contents of the Email Hoaxes virus are usually by giving a warning email containing information about a security threat that is not actually there. The description conveyed contains computer [4] virus effects made very sensible (file deletion, reproduction via email address book address) to a fantastic (hardware destruction, install of malware via plaintext message). Very often, they contain references to software or Internet security firms, which act as credibility markers.

B. Giveaway email hoaxes

Giveaway email hoaxes are one of email hoaxes type. It spreads by sending email to peoples. If the user performs as instructed within a certain period of time the user will be given the form of goods or money for free. This is monitored premises using email tracking. Giveaway Email Hoaxes are the easiest subgroup to analyze in terms of. The uniqueness of the behavior, this email contains news if the scam is successful and trusted. Fig 3 depicts the example of giveaway email hoaxes[11].

Hello Everyone,

And thank you for signing up for my Beta Email Tracking Application or (BETA) for short. My name is Bill Gates. Here at Microsoft we have just compiled an e-mail tracing program that tracks everyone to whom this message is forwarded to. It does this through an unique IP (Internet Protocol) address log book database.

We are experimenting with this and need your help. Forward this to everyone you know and if it reaches 1000 people everyone on the list you will receive $1000 and a copy of Windows98 at my expense. Enjoy.

Note: Duplicate entries will not be counted. You will be notified by email with further instructions once this email has reached 1000 people. Windows98 will not be shipped until it has been released to the general public.

Your friend, Bill Gates & The Microsoft Development Team

Fig 3. Giveaway Email Hoaxes

C. Charity Email Hoaxes

Charity Email Hoaxes was as 'sympathy hoaxes', it contains emails that disseminate hoax information for fundraising, or donations through accounts, eg a sick child but his or her parents cannot afford to pay maintenance fees by selling the child's name to get as many donations in a bank account as an example of a hoax charity email.

Example 11. The 'Jessica Mydek' charity EH (ID: 119). LITTLE JESSICA MYDEK IS SEVEN YEARS OLD AND IS SUFFERING FROM AN ACUTE AND VERY RARE CASE OF CEREBRAL CARCINOMA. THIS CONDITION CAUSES SEVERE MALIGNANT BRAIN TUMORS AND IS A TERMINAL ILLNESS. THE DOCTORS HAVE GIVEN HER SIX MONTHS TO LIVE. AS PART OF HER DYING WISH, SHE WANTED TO START A CHAIN LETTER TO INFORM PEOPLE OF THIS CONDITION AND TO SEND PEOPLE THE MESSAGE TO LIVE LIFE TO THE FULLEST AND ENJOY EVERY MOMENT, A CHANCE THAT SHE WILL NEVER HAVE.FURTHERMORE, THE AMERICAN CANCER SOCIETY AND SEVERAL CORPORATE SPONSORS HAVE AGREED TO DONATE THREE CENTS TOWARD CONTINUING CANCER RESEARCH FOR EVERY NEW PERSON THAT GETS FORWARDED THIS MESSAGE. PLEASE GIVE JESSICA AND ALL CANCER VICTIMS A CHANCE. IF THERE ARE ANY QUESTIONS, SEND THEM TO THE AMERICAN CANCER SOCIETY AT ACS@AOL.COM

Fig 4. Charity Email Hoaxes

D. Urban legends Email Hoaxes

Urban Legend is a contemporary myth or legend that is the whole. Many of these emails are disseminated in the form of information relating to mystery, horror, humour, or even a moral story. Urban legend does not always mean the same lie as the story spread by word of mouth. Urban legends are also often exaggerated to become more sensational, following an example of urban legends email hoaxes can be seen in Fig 5. Urban Legend Email Hoaxes

Jakarta - Recently a poster solicitation in social media. The poster invites Muslims to hold action 212 Volume 2. The horrendous of this poster also joined the Commander of the TNI General Gatot Nurmantyo. The poster seemed to wake Gatot Nurmantyo participate in the action. At the moment, the action posters 212 include photos

Fig 5. Urban Legend Email Hoaxes

E. Hoaxed Hoaxes

Hoaxed hoaxes is one of the last category that essentially defined by a pragmatic property. The hoaxed hoaxes form is shown in the email clearly on an email and display as a parody. Figure 6 showed the example of 'BedTimes' hoaxed hoax.

If you receive an email entitled "Bedtimes" delete it IMMEDIATELY. Do not open it. Apparently this one is pretty nasty. It will not only erase everything on your hard drive, but it will also delete anything on disks within 20 feet of your computer.

Fig 6. Example of Hoaxed Hoaxes Text

3.3 Characteristic of Hoax and Fake Email

Every people are ever sent the email by an unknown person. In this condition, it can affect personal email and network overloading. So, it can be known as viruses or hoax information in the system, when it is detected by antivirus. There are thousands of email viruses and hoaxes email that included in an email[12]. As an example, hoaxes email contain bad grammar and misspelled words. Fig 7. Example of Email Viruses shows an email that contains fake information.

"Hiiii How are youuuuuuuu? look to bill caricature it's vvvery verrrry fffunny i promise you will love it? ok buy

Fig 7. Example of Email Viruses

There are several characteristics to detect email that containing hoaxes [12,13]. First, the email messages are forwarded by friend or colleague. In this emails usually, are not contain an attachment. In addition, it has not trusted organization and contains a link to the bogus site. Second, the email requests personal information such as name, email, phone or bank account. It can be dangerous when people send email without crosschecking and making validation. Another condition, the email also send with a non-specific greeting such as "dear customer". On the other hand, most of the official email send an email with calling the proper name or full name. Finally, antivirus give a warning when the email is incoming, send it to a trusted site and viruses are checked in antivirus database. The analysis process of viruses and hoaxes email characteristics hopefully makes it simple to create a system that can recognize hoax emails.

3.4 Machine Learning

Machine Learning is an area of artificial intelligence associated with the development of techniques that can be programmed and learned from past data. Broadly speaking, there are two approaches seen from the way learning is done. The approach is

done by supervised and unsupervised learning. Supervised learning is one method of learning with the practice and coach. Many techniques in the pattern recognition are included in this category, for example, regression, discriminant analysis (LDA), Artificial Neural Networks (ANN) and Support Vector Machine (SVM). Meanwhile, unsupervised learning is a method without including training and coaches. Thus, in a group of data without a specific label or class. This unsupervised learning technique should classify data against some desired classes.
Hoaxes filtering are the type of binary classification. This research focused on detecting legitimate e-mails as a negative (-) and positive (+) instances. Machine Learning is a domain of computer science that can be developed with computer systems to improve the performance in a task based on experience or data. Many kinds of research for e-mail classification uses statistical approaches or machine learning techniques to build a model or classifier specifically for the task of filtering spam from a user email stream.

A. Naive Bayes (NB)
Naive Bayes classification technique is a technique that is still popular to date and arguably the most famous statistic spam multiplier. This technique is called 'naive' because it ignores the possibility of dependence or input variation and reduces multivariate problems to a group of uni-variate problems. This technique uses a probabilistic approach to infer a data. Naive Bayes does not require complicated iterative parameter estimation schemes and easy to build, interpret, applied in large datasets and highly effective [14][15].

B. Support vector machines (SVMs)
Support vector machines (SVMs) are one of the best-supervised learning algorithms. SVMs have become one of the most popular classifiers in the Machine Learning researches because they provide superior generalization performance, require less training data, and can solve high-dimensional data with the use of kernels [14]. Support vector machines give the result by mapping the vectors as a feature into a linear or non-linear feature space through a kernel function. The feature space generates a hyperplane. This hyperplane called an optimal separating hyperplane (OSH) which splits the positive and the negative samples with maximum margin. The hyperplane is used for non-linear decision boundary.

C. Clustering Techniques
Clustering is a task for making a group of data based on similarity. Clustering techniques have been widely used in a research for variety domain. Hoax email have applied unsupervised learning for clustering. Whistel and Clarke partition e-mail datasets into ham and spam clusters. The research focused on a novel investigation of email spam clustering[16]. The study showed significant result by using clustering approach and give better result than semi-supervised approaches. So, it can be proven that clustering can be a powerful tool for e-mail spam filtering.

3.5 Methods for Mitigating E-mail Hoax

Although there are 'social' methods such as never respond to spam, never forward chain-letters to fight spam, they have had a narrow effect on spam so far is seen by the number of spam messages received daily by users. Technical measures seem to be the most effective in countering spam. Prior to machine learning techniques, many different technical measures were employed for spam filtering, like - rule-based spam filtering, white lists, black lists, challenge-response (C/R) systems, spam filtering, honeypots, OCR filters, and many others, each with its own merits and drawbacks. Black-lists, white-lists, challenge-response (C/R) systems, etc. are origin-based techniques used by reputation-based filters. In figure Fig **8:** Methods for Mitigating Email Hoaxes are showed the basic methods to mitigating email hoaxes, there are three methods: heuristic filters, blacklisting and whitelisting.

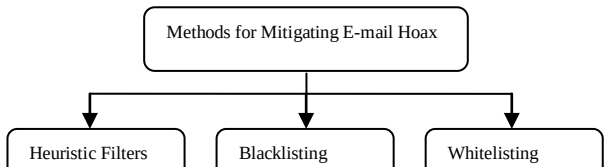

Fig 8: Methods for Mitigating Email Hoaxes

A. Heuristic Filters
Knowledge engineering can be used for spam filtering. Knowledge approach are based on coded rules or heuristics[17]. A content based heuristics filter analyzes the words like 'spammy', 'lottery' words. They were designed a patterns system that observed in messages[18]. Detection of spam are implemented with the use of keyword spotting rules. The RIPPER algorithm can achieve good performance, comparing weighting traditional method like TF-IDF weighting. The drawback of heuristic filters is that maintaining an effective set of rules is a time consuming affair, moreover the rules have to keep update with the newest trends in spam.

B. Blacklisting
Blacklisting e-mail addresses or IP addresses can be done from the user or server level. The message is automatically blocked at the SMTP connection phase when receive an e-mail from untrusted sources. This method becomes a simple lookup in the blacklist every time; hence the computational cost is low. Black-lists include Real-time Blackhole ListS (RBL) and Domain Name System Black-lists. Common black-list databases include proxies or open relays, networks or individual addresses guilty of sending spam. Google blacklists and SpamHaus 7 are the products to prevent spam by using blacklists method. Blacklist techniques through effective, suffer from many drawbacks. A legitimate address may be blacklisted by the filter erroneously or arbitrarily. Innocent users can get victimized and entire domains (e.g. Hotmail) can get blocked when e-mail IDs or IP addresses are used by spammers without the owner's consent.

C. Whitelisting
Whitelisting is another method of blacklisting. An e-mail whitelist is a list of pre-approved or trusted contacts, domains, or IP addresses that are able to communicate to a mail user. But, all e-mails can be blocked by this method. This restrictive method may introduce an extremely high false positive rate instead of reducing it. Such a method may be good for instant messaging environments but is not a good choice as it prohibits establishing new contacts through e-mail. Moreover if spammers somehow got their hands on the whitelist, it would be easy to evade the filter using spoofed addresses, or using well-known whitelisted mailing lists. This method requires a lot of maintenance but provides moderate filtering rate. It can be employed together with other anti-spam techniques[19].

3.6 Text Mining

Text mining is the process of retrieving information from the text. Information is usually obtained through forecasting patterns and trends in statistical pattern learning. Text mining is parsing, along with the addition of some linguistic features of derivatives and the removal of some of them, and subsequent insertion into the database, determining the poles in the structured data, and finally evaluating and interpreting the output; text mining usually refers to some combination of relevance, novelty, and interestingness. The text mining are included text categorization, text clustering, concept/entity extraction, granular taxonomy production,

sentiment analysis, document conclusions, and entity relationship modeling, learning relationships between entities.

3.7 Extraction

The effectiveness and success of email hoax filters depends on feature extraction. Feature extraction is a step for defining and creating those features to make the classifier perform better. There are many steps in feature extraction from an e-mail [20]:

a. Lexical Analysis (Tokenization)

Tokenization is a step which start with splitting sentence into words to see the terms that relevant with hoax terms. Headers, attachments, and HTML tags are stripped, leaving behind just the e-mail body and subject line text. IP addresses and domain names can also be considered as tokens.

b. Stop-word Removal

Stop-word removal is a method for removing frequently non-informative words, e.g. 'a', 'an', 'the', 'are', 'is', and etc. Obscure texts or symbols may also be deleted in subsequent steps. Stop-word removal makes the candidate terms more efficient and reduces the feature space considerably[21].

c. Stemming

Word-stemming is a process of converting words to base forms, by removing words that contains tenses, gerund forms, , plurals, prefixes and also suffixes. Stemming is closely related to lemmatization which while reducing a word considers the part of speech and the context of the word. The advantages of stemming and lemmatization are for reducing feature space dimension and becoming better in classifier accuracy.

d. Representation

Representation involves the conversion of an email message into a specific or structured format as needed by the machine learning algorithm being employed.

4. Conclusion

Hoax's email here is a false news spread over email, it is designed to deceive an email recipients, usually for business, or personal use. For this reason, this paper discusses the characteristics of email hoax. Based on the characteristics of this hoax email can be concluded as follows that Hoax Email is, Asynchronous one message sent via email, communication from the sender is connected to a number of recipients in the social network, Contains false information, with no extratextual consequences, and directives for dissemination. This framework will be the basis and benchmark for further analysis to build hoax email data sets

References

[1] The Radicati Group Inc., Email Statistics Report, 2017-2021, 44 (2017) 4.

[2] A.B. Prasetijo, R.R. Isnanto, D. Eridani, Y.A.A. Soetrisno, M. Arfan, A. Sofwan, Hoax detection system on Indonesian news sites based on text classification using SVM and SGD, 2017 4th Int. Conf. Inf. Technol. Comput. Electr. Eng. (2017) 45–49. doi:10.1109/ICITACEE.2017.8257673.

[3] Y.Y. Chen, S.-P. Yong, A. Ishak, Email Hoax Detection System Using Levenshtein Distance Method, J. Comput. 9 (2014) 441–446. doi:10.4304/jcp.9.2.441-446.

[4] N. Kurniasih, N. Kurniasih, L.A. Abdillah, I.K. Sudarsana, I.W.L. Yogantara, I.N.T. Astawa, R.F. Nanuru, A. Miagina, J.O. Sabarua, M. Jamil, J. Tandisalla, E. Duan, F.G.J. Rupilele, M.D. Utama, M. Laisila, A.S. Ahmar, R. Rahim, *Prototype Application Hate Speech Detection Website Using String Matching and Searching Algorithm*, Int. J. Eng. Technol. 7 (2018) 62–64. doi:10.14419/ijet.v7i2.5.13952.

[5] E. Tacchini, G. Ballarin, M.L. Della Vedova, S. Moret, L. de Alfaro, Some like it Hoax: Automated fake news detection in social networks, CEUR Workshop Proc. (2017). doi:10.1257/jep.31.2.211.

[6] A. Elyashar, J. Bendahan, R. Puzis, Is the News Deceptive? Fake News Detection using Topic Authenticity, (2017) 16–21.

[7] S.M. Sirajudeen, N.F.A. Azmi, A.I. Abubakar, Online fake news detection algorithm, J. Theor. Appl. Inf. Technol. 95 (2017) 4114–4122.

[8] S. Zannettou, M. Sirivianos, J. Blackburn, N. Kourtellis, The Web of False Information: Rumors, Fake News, Hoaxes, Clickbait, and Various Other Shenanigans, (2018) 1–26.

[9] M. Vuković, K. Pripužić, H. Belani, An intelligent automatic hoax detection system, Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 5711 LNAI (2009) 318–325. doi:10.1007/978-3-642-04595-0_39.

[10] T. Petković, T. Petković, Z. Kostanjčar, P. Pale, E-Mail System for Automatic Hoax Recognition, (2005).

[11] T. Heyd, Email hoaxes: Form, function, genre ecology., Email Hoaxes Form, Funct. Genre Ecol. (2008).

[12] S.M. University, Characteristics of Viruses and Virus Hoaxes, (n.d.).

[13] Scam emails, (n.d.).

[14] X. Wu, V. Kumar, Q.J. Ross, J. Ghosh, Q. Yang, H. Motoda, G.J. McLachlan, A. Ng, B. Liu, P.S. Yu, Z.H. Zhou, M. Steinbach, D.J. Hand, D. Steinberg, Top 10 algorithms in data mining, 2008. doi:10.1007/s10115-007-0114-2.

[15] W.B. Zulfikar, N. Lukman, Perbandingan Naive Bayes Classifier Dengan Nearest Neighbor Untuk Identifikasi Penyakit Mata, J. Online Inform. 1 (2016) 82–86. doi:10.15575/join.v1i2.33.

[16] J.S. Whissell, C.L. a Clarke, Clustering for Semi-Supervised Spam Filtering Categories and Subject Descriptors, Proc. 8th Annu. Collab. Electron. Messag. Anti-Abuse Spam Conf. ACM. (2011) 125–134.

[17] E.P. Sanz, J.M.G. Hidalgo, J.C.C. Pérez, Email Spam Filtering, Adv. Comput. 74 (2008) 45–114. doi:10.1016/S0065-2458(08)00603-7.

[18] T.S. Guzella, W.M. Caminhas, A review of machine learning approaches to Spam filtering, Expert Syst. Appl. 36 (2009) 10206–10222. doi:10.1016/j.eswa.2009.02.037.

[19] E. Michelakis, I. Androutsopoulos, G. Paliouras, G. Sakkis, Filtron : A Learning-Based Anti-Spam Filter . Filtron : A Learning-Based Anti-Spam Filter, in: CEAS 2004 - First Conf. Email Anti-Spam, 2004.

[20] A. Bhowmick, S.M. Hazarika, Machine Learning for E-mail Spam Filtering: Review,Techniques and Trends, (2016).

[21] A.R. Atmadja, A. Purwarianti, Comparison on the rule based method and statistical based method on emotion classification for Indonesian Twitter text, 2015 Int. Conf. Inf. Technol. Syst. Innov. ICITSI 2015 - Proc. (2016). doi:10.1109/ICITSI.2015.7437692.