

## The architecture social media and online newspaper credibility measurement for fake news detection

Rakhmat Arianto<sup>\*1</sup>, Harco Leslie Hendric Spits Warnars<sup>2</sup>, Edi Abdurachman<sup>3</sup>,  
Yaya Heryadi<sup>4</sup>, Ford Lumban Gaol<sup>5</sup>

<sup>1</sup>Informatics Engineering, Sekolah Tinggi Teknik PLN,  
Menara PLN, Lingkar Luar Barat Duri Kosambi, Cengkareng, Jakarta, Indonesia

<sup>1,2,3,4,5</sup>Computer Science, Bina Nusantara University

Kebon Jeruk Raya No. 27. Kebon Jeruk, Jakarta, Indonesia

\*Corresponding author, e-mail: arianto@sttpln.ac.id<sup>1</sup>, shendric@binus.edu<sup>2</sup>,  
edia@binus.edu<sup>3</sup>, yayaheryadi@binus.edu<sup>4</sup>

### Abstract

*Social media is one of the communication media favored by people in the world, especially in Indonesia. This is evidenced by the results of the APJII survey which shows that the majority of Indonesians use social media in their daily activities. One of the advantages of social media is the dissemination of information faster than conventional media so that the quality of information disseminated is lower than conventional media due to the process of disseminating information not through the filter process. By measuring the level of credibility of the online newspaper based on the time credibility, website credibility, and message credibility factors and measuring the level of credibility on social media based on the time credibility, Social Media Credibility, and Message Credibility factors with different levels of weight, it will produce a news likelihood level it's fake news or facts.*

**Keywords:** *credibility, fake news, measurement, online newspaper, social media*

**Copyright © 2019 Universitas Ahmad Dahlan. All rights reserved**

### 1. Introduction

The development of Internet technology in the world, especially in Indonesia, greatly influences the lifestyle of people who are in desperate need of internet for everyday life. This is reinforced by the results of a survey of internet users in Indonesia conducted by the Indonesian Internet Service Providers Association (APJII) in 2017 showing that internet users from year to year continue to increase to exceed 143.26 million people or 54.68% of the total number of Indonesians in the year 2017 which reached 262 million people. With details 87.13% use the internet to access social media [1]. The large number of internet users accessing social media is related to the benefits gained from social media, namely: (i) the majority of social media provides news more timely and requires lower costs to obtain information when compared to traditional news media such as newspapers or television. (ii) social media allows users to share information with other users, comment directly on the news presented, and discuss with other users related to the theme presented [2].

With the benefits of using social media, the quality of news disseminated through social media is lower than the news distributed through traditional media. This is influenced by one of the characteristics of social media where the message is delivered freely without having to go through the gatekeeper [3] so that social media users will easily make false news with certain objectives such as attacking political opponents, attacking business opponents, or even vilifying other groups. Likewise, the number of users is very large and has a low cost to access social media, so the news that has a negative meaning will very quickly spread to other social media users. The Indonesian Government through the Ministry of Communication and Information has anticipated the spread of false news by drafting law number 11 of 2008 or better known as the Electronic Information and Transaction act [4]. However, the existence of such legislation is deemed insufficient to reduce the spread of fake news on social media because some sites like turnbackhoax.id and hoaxornot.detik.com still keep updating their news about fake news spread on social media.

Fake news detection can be done in several ways that can be grouped into 3 categories, namely [5]: (i) Knowledge-based or commonly known as fact checking which is a way to detect fake news by searching for facts from the news using information retrieval method, semantic web, and linked open data (LOD). (ii) Context-based is the detection of false news based on the analysis of social networks where the spread of false news will be sought news sources that spread the first time. (iii) Stylebased, namely false news detection based on computational linguistics and natural linguistic processing or more specifically on the detection of lies from the news disseminated. All of these categories represent several factors in detecting fake news that have been published by the International Federation of Library Associations and Institution (IFLA). However, one of the factors that has not been included in the questionnaire is the factor of checking the date of publication.

The proposed architecture system consists of several stages, namely Keywords Extraction, News Scoring, Social Media Scoring. Keywords Extraction will function to get keywords from sentences or documents from the user. News Scoring serves to measure the level of fake news from a sentence or document from the user based on information obtained on news websites or online newspapers by paying attention to the time credibility, website credibility and message credibility factors. Furthermore, the Social Media Scoring stage will be carried out if the News Scoring results are still less than the specified weight, it will calculate the level of social media credibility that contains news from the user by considering the Time Credibility, Social Media Credibility, and Message Credibility factors with their respective weights. has been determined. It is hoped that with this architecture system, it can get a better level of credibility than previous research.

## 2. Related Works

IFLA through its blog provides information that to recognize a fake news can be done in several stages, namely: Consider the source (to understand its mission and purpose), Read beyond the headline (to understand the whole story), Check the authors (to see if they are real and credible), Assess the supporting sources (to ensure they support the claims), Check the date of publication (to see if the story is relevant and up to date), Ask if it is a joke (to determine if it is meant to be satire), Review your own biases (to see if they are affecting your judgement), Ask experts (to get confirmation from independent people with knowledge).

In the knowledge based category or what is commonly called fact-checking that uses information retrieval techniques exemplified in several research according to research [5] were research [6] that utilizes Text Runner [7] tools to obtain information from several online sources and information extraction and information indexing processes that will be compared to the inconsistency level based on questions entered by the user, then to improve the results of fake news detection, the [8] research proposes to check factual information obtained from several sources using a statistical model where the more news is discussed by several online sources, the information has more truth high. But in the study also still left a problem where the level of credibility and reputation of a website that contains the same information should be measured to filter which websites can be trusted or not. In research [9] which provides a literature review on the level of credibility and reputation of a website can be one of the inputs in completing previous research. Where one technique to analyze the level of credibility of a website can be determined based on the reputation of the website. The higher the reputation of a website, the higher the level of credibility of a website.

There are several studies that use the semantic website or LOD technique in the knowledge-based category, including the [10] research that builds a framework to test claim models by changing the parameters that have been determined and analyzing the changes in conclusions that occur due to changes in parameters. Whereas in the study [11] revealed the problems encountered in the fact checking method where finding the shortest path between concepts in a knowledge graph get less optimal measurement results so that a calculation metric is proposed to analyze the truth of the statement by considering the path length factors between concepts at issue. Whereas in the research conducted by Shi and Weninger [12] shows that the detection of fake news can be done by using the probability that the information in question must have a probability of connectedness with the facts that have been collected to obtain a higher level of credibility.

Measurement of credibility devoted to social media, research [13-15] show about measuring the level of credibility of a social media especially on Twitter. The measurement of the credibility of a social media can be measured by the owner of a social media account that interacts with others including how often the account is spreading information and how open about the identity of the account owner. Furthermore, research in [16] shown that the information shared on social media affects the perceptions of credibility from sources that share information. This is shown by conducting an experiment showing twitter pages to correspondents then correspondents specify the source they consider the page owner. The experimental results show that the Tweet reviewer affects the source credibility. Moreover, research [17], showed that to eliminate false news or spam on Twitter can be done using semi-supervised ranking models for scoring tweets according to their credibility on Tweets containing information relating to crisis events. The settlement of credibility issues with the information shared in social media especially Twitter was also shown in the study [18] in which two approaches were conducted to determine the credibility of information on Twitter (low, high and average). The first approach is to calculate the degree of similarity between information on Tweets to verified news and the second approach is based on the similarity with verified news sources in addition to a set of proposed features. Furthermore, CREDBANK was introduced in research [19] where CREDBANK is a collection of corpus derived from Twitter, topics, and events that have been determined by the level of credibility by human. The collection of corpus was obtained from streaming Twitter over a period of more than three months totaling over 1 billion tweets and has been automatically verified by humanity against 30 human annotators. Besides research in the field of journalism, RumorLens was introduced [20] which as a tool to help journalists to identify new rumors distributed via Twitter and show the audience that the information shared is a rumor and verification of the rumors which had been spread in other Tweets. So the user can determine whether the information from Twitter will be explored further, to be corrected from other Tweets, or not related to other Tweets.

In the study [21], research to detect information on Twitter is a rumor or not by performing five stages of the procedure by: identify Signal Tweets, identify Signal Clusters, detect Statements, Capture Non-Signal Tweets, and rank Candidate Rumor Clusters. In this study, identified a third of the top 50 clusters were judged to be rumors. The detection of hoaxes can be done as in research [22] where to detect identity, demographic background and linkage between documents can be detected based on the writing style of the source document. The contribution of this research is detecting stylistic deception in written documents. The test result using large feature set between regular documents with deceptive document shows 96.6% accuracy as (F-Measure). Moreover, research in [23] proposed a novel time-aware ranking model that leverages on multiple sources of crowd signals. The approach built on two major novelties. First, a unifying approach that given query  $q$ , mines and represents temporal evidence from multiple sources of crowd signals. The model could predict the temporal relevance of documents for query  $q$ . Second, a principled retrieval model that integrates temporal signals in a learning to rank framework, to rank results according to the predicted temporal relevance. Research in [24] it was shown that the limitations of fact-finding on web sources can be fragmented due to the scarcity of related resources so that in this study carried out the taking of a claimed article, and model the mutual interaction between: the stance (i.e., support or refute) of the sources, the language style of the articles, the reliability of the sources, and the claim's temporal footprint on the web. Extensive experiments demonstrate the viability of the method and its superiority over prior works. The result showed that the methods work well for early detection of emerging claims, as well as for claims with limited presence on the web and social media.

### **3. Proposed System Architecture**

In the fake news detector model proposed as seen in Figure 1, can be divided into 3 steps such as: Keyword Extractions, News scoring and Social Media Scoring. At the end, credibility score will be counted as justification as fake news or not.

#### **3.1. Keyword Extraction**

Keyword Extraction or commonly known as Preprocessing Text [25] is used to get important keywords or words from documents entered by the user. This stage is divided into 4 processes such as Tokenizing, Filtering, Stemming, and Tagging.

- a. Tokenizing is the process of separating words per word from sentences entered by the user. The process of separation can be done by separating words based on spaces or punctuation in the sentence.
- b. Filtering is a non-inclusive word-removal process that can be used as a keyword that is usually a person's pronoun or a hyphen. This process relies heavily on a collection of words called Stopword. If there is a word included in the Stopword, then the word is omitted.
- c. Stemming is a basic word search process on every word the result of filtering. The function of this process is to find the basic word of a word that has either a prefix, a suffix, an insertion, or a combination thereof. In previous research, special on Indonesian stemming can be done by using method of Sastrawi Stemming.
- d. Tagging [26] is a process of correction of the Stemming process where in the process of stemming there are some words that do not match the basic word that existed in the Big Indonesian Dictionary (KBBI) [27]. So it takes a basic Indonesian word database obtained from KBBI.

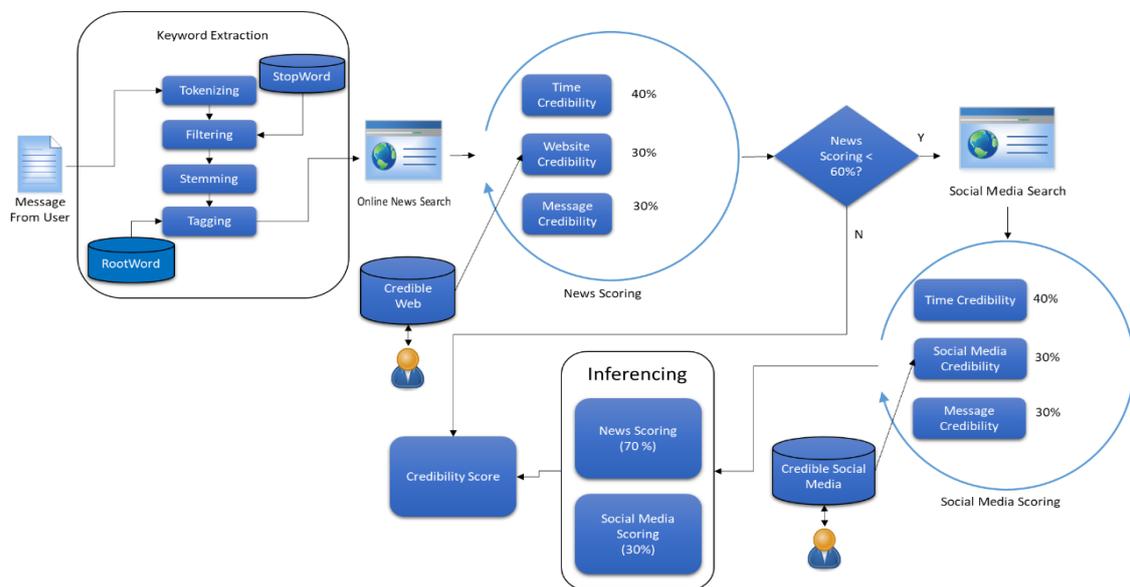


Figure 1. System architecture approach

The results of Keyword Extraction will be used at the stage of Online Search News that uses help from the Google Search API to find similar information on online news sites. Collection of news obtained from searching online news sites will be calculated similarity levels using the Document Similarity algorithm with Vector Space Model [28].

$$\cos(d_j, q) = \frac{d_j \cdot q}{\|d_j\| \|q\|} = \frac{\sum_{i=1}^N w_{i,j} w_{i,q}}{\sqrt{\sum_{i=1}^N w_{i,j}^2} \sqrt{\sum_{i=1}^N w_{i,q}^2}} \tag{1}$$

where:

$\cos(d_j, q)$  = similarity between document and query

$d_j$  = document

$q$  = query

whereas to find the weight vector for each document can be calculated by (2):

$$w_{t,d} = tf_{t,d} \cdot \log \frac{|D|}{|\{d' \in D | t \in d'\}|} \tag{2}$$

where:

$tf_{t,d}$  = term frequency of term t in document d

$\log \frac{|D|}{|\{d' \in D | t \in d'\}|}$  = inverse document frequency (a global parameter)  
 $|D|$  = the total number of documents in the document set  
 $|\{d' \in D | t \in d'\}|$  = the number of documents containing the term t.

### 3.2. News Scoring

News scoring (NS) as shown in Figure 1 will be counted with TC (Time Credibility), WC (Website Credibility) and MC (Message Credibility) and divided with N as Number of News Document, where the equation is shown in (1) where TC will be scored 0.4, WC is 0.3, and MC is 0.3 respectively. Time Credibility gets a higher weight because in the previous study did not pay attention to the factors of news publication being a determinant of false news detection, but the repetition of news publications was also included in one of the factors that caused false news according to IFLA.

Results from the online news search process will collect several news sites that contain information that has similarities to the queries entered by the user. Some of these sites will be taken when the news publication will be a factor in Time Credibility. The website name will be a factor in Website Credibility, and the contents of the news will be a factor in Message Credibility. If a news is more distant from the time of the smallest news publication, it will get a smaller Time Credibility value. If the Website that publishes the news has a high level of popularity, it will get a higher Website Credibility value. Whereas in Message Credibility, a news will get the highest value if it has a high level of similarity based on all news content on the query input from the user. Then the connectedness of the three factors can be formulated into:

$$NS = (TC * 0.4) + (WC * 0.3) + (MC * 0.3) \quad (3)$$

where

NS = News Scoring,  
 TC = Time Credibility  
 WC = Website Credibility  
 MC = Message Credibility

### 3.3. Social Media Scoring

Similar like News Scoring, then Social Media scoring (SMS) will be counted with 3 parameters such as TC (Time Credibility), SMC (Social Media Credibility) and MC (Message Credibility) and divided with N as Number of Social Media Document, where the equation is shown in (4) where TC will be scored 0.4, SMC is 0.3 and MC is 0.3 respectively. TC on SMS gets a higher weight because if there is a repetition of information dissemination, it will be directly considered false news.

This stage will be carried out if the value of News Scoring does not reach or less than 0.6 because the value of 0.6 becomes a representative value above the middle value between 0 and 1. However, this value will change depending on the results of the experiment to be carried out so that it can get a more accurate value. But if the NS value has reached more than 0.6, then the news can be said as fact news or have a high level of credibility. Social media used as a search site is Twitter and Facebook. Where in the search process used keywords entered by the user so that found some news that contains keywords from users on some social media profiles on Facebook or Twitter. Scoring is based on the time of publication (Time Credibility), the credibility of the account that publishes the news (Social Media Credibility), and the relevance of the content of the published news to the keywords of the user (Message Credibility).

$$SMS = (TC * 0.4) + (SMC * 0.3) + (MC * 0.3) \quad (4)$$

where,

SMS = Social Media Scoring  
 TC = Time Credibility  
 SMC = Social Media Credibility  
 MC = Message Credibility

### 3.4. Credibility Score

At this stage the conclusion of the scoring stage has been done in the previous stage. Where there are two conditions of conclusion scoring, namely: The conclusion is made when the scoring of News Scoring more than 0.6 and the conclusion is made when the scoring of News Scoring less than 0.6. Inferences made on the value of News Scoring more than 0.6 then the value of News Scoring results are considered as Credibility Score of the news entered by the user application. As shown in (5).

$$\text{If } (NS > 0.6) \text{ CS} = \text{NS} \quad (5)$$

where,

NS = News Scoring

CS = Credibility Score

The inference made on the News Scoring score of less than 60 then required a sum of scoring results from the News Scoring stage with the scoring result in Social Media Scoring stage and given weight on each scoring result as in (6).

$$\text{If } (NS < 0.6) \text{ CS} = (NS * 0.7) + (SMS * 0.3) \quad (6)$$

where

CS = Credibility Score

NS = News Scoring

SMS = Social Media Scoring

## 4. Conclusion

This research intends to propose a system architecture to detect hoax or fake news on Indonesian language news. Where the system architecture is built with several stages, namely: Keyword Extraction that serves to find the keyword of the text entered by the user, Scoring that serves to provide a credibility score of a news based on time, online news/social media, and message so it can be concluded the possibility from such news hoaxes or facts the proposed architectural system has represented several stages to avoid fake news by the IFLA, including: consider the source stage, check the author and supporting source represented in the Website Credibility process in News Scoring and Social Media Credibility in Social Media Scoring. Then at the Check the date stage it is represented in the Time Credibility process in the News Scoring and Social Media Scoring. Whereas in the Read Beyond and Check your biases stages in IFLA are represented in the Message Credibility process in News Scoring and Social Media Scoring. This research still necessary need to be tested with several datasets and optimization of the weighting values based on the tested dataset.

## Acknowledgement

The publication of this research was sponsored by PT. INDOINTERNET (Indonet), PT. TAMANSARI (Partner ACER Indonesia), Sekolah Tinggi Teknik PLN, and Bina Nusantara University for their sponsorship, supporting funds, and support so that this research and publication can be carried out.

## References

- [1] Penetration & Behavior of Indonesian Internet Users 2017 (in Indonesia: Penetrasi & Perilaku Pengguna Internet Indonesia 2017). APJII. 2017.
- [2] Shu K, Sliva A, Wang S, Tang J, Liu H. Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter*. 2017; 19(1): 22–36.
- [3] Gamble M. Gamble TK. *Communication Works*, Seventh Edition. Boston: McGraw-Hill College. 2002.
- [4] Government of the Republic of Indonesia (Pemerintah Republik Indonesia). *Law of the Republic of Indonesia Number 11 of 2008 concerning Information and Electronic Transactions (in Indonesia Undang-Undang Republik Indonesia Nomor 11 Tahun 2008 Tentang Informasi Dan Transaksi Elektronik)*. Ministry of Communication and Information Kementerian Komunikasi dan Informasi. 2008: 1–65.

- [5] Potthast M, Kiesel J, Reinartz J, Bevendorff J, Stein B. A Stylometric Inquiry into Hyperpartisan and Fake News. *ArXiv*. 2017.
- [6] Etzioni O, Banko M, Soderland S, Weld DS. Open information extraction from the web. *Communications of the ACM*. 2008; 51(12): 68.
- [7] Yates A, Cafarella M, Banko M, Etzioni O, M. Broadhead, S. Soderland. *TextRunner: Open Information Extraction on the Web*. Proceeding NAACL-Demonstrations '07 Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations. 2007; 42: 25–26.
- [8] Magdy A, Nayer W. *Web-based Statistical Fact Checking of Textual Documents*. SMUC '10 Proceedings of the 2<sup>nd</sup> international workshop on Search and mining user-generated contents. 2010: 103–109.
- [9] Ginsca AL, Popescu A, Lupu M. Credibility in Information Retrieval. *Foundations and Trends® in Information Retrieval*. 2015; 9(5): 355–475.
- [10] Wu Y, Agarwal PK, Li C, Yang J, Yu C. *Toward computational fact-checking*. Proceedings of the VLDB Endowment. 2014: 589–600.
- [11] Ciampaglia GL, Shiralkar P, Rocha LM, Bollen J, Menczer F, Flammini A. Computational fact checking from knowledge networks. *PLoS One*. 2015; 10.
- [12] Shi B, Wenginger T. *Fact Checking in Heterogeneous Information Networks*. WWW '16 Companion Proceedings of the 25<sup>th</sup> International Conference Companion on World Wide Web. 2016: 101–102.
- [13] Ledbetter AM, Redd SM. Celebrity Credibility on Social Media: A Conditional Process Analysis of Online Self-Disclosure Attitude as a Moderator of Posting Frequency and Parasocial Interaction. *Western Journal of Communication*. 2016: 1–18.
- [14] Lin X, Spence PR, Lachlan KA. Social media and credibility indicators: The effect of influence cues. *Computers in Human Behavior*. 2016; 63: 264–271.
- [15] Edwards C, Edwards A, Spence PR, Shelton AK. Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on Twitter. *Computers in Human Behavior*. 2014; 33: 372–376.
- [16] Westerman D, Spence PR, Van Der Heide B. Social Media as Information Source: Recency of Updates and Credibility of Information. *Journal of Computer-Mediated Communication*. 2014; 19(2): 171–183.
- [17] Gupta A, Kumaraguru P, Castillo C, Meier P. *Tweetcred: Real-time credibility assessment of content on twitter*. International Conference on Social Informatics. 2014: 228–243.
- [18] Al-Khalifa HS, Al-Eidan RM. An experimental system for measuring the credibility of news content in Twitter. *International Journal of Web Information Systems*. 2011; 7(2): 130–151.
- [19] Mitra T, Gilbert E. *CREDBANK: A Large-Scale Social Media Corpus With Associated Credibility Annotations*. ICWSM, 2015: 258–267.
- [20] Resnick P, Carton S, Park S, Shen Y, Zeffer N. *Rumorlens: A system for analyzing the impact of rumors and corrections in social media*. Proc. Computational Journalism Conference. 2014.
- [21] Zhao Z, Resnick P, Mei Q. *Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts*. Proceedings of the 24<sup>th</sup> International Conference on World Wide Web. 2015: 1395–1405.
- [22] Afroz S, Brennan M, Greenstadt R. *Detecting hoaxes, frauds, and deception in writing style online*. Proceedings-IEEE Symposium on Security & Privacy. 2012: 461–475.
- [23] Martins F, Magalhes J, Callan J. *Barbara Made the News: Mining the Behavior of Crowds for Time-Aware Learning to Rank*. Proceedings of the Ninth ACM International Conference on Web Search and Data Mining. 2016; (1): 667–676.
- [24] Popat K, Mukherjee S, Strötgen J, Weikum G. Where the Truth Lies: Explaining the Credibility of Emerging Claims on the Web and Social Media. *WWW (Companion Vol)*. 2017: 1003–1012.
- [25] Vijayarani S, Ilamathi J, Nithya M. Preprocessing Techniques for Text Mining - An Overview. *International Journal of Computer Science & Communication Networks*. 2015; 5(1): 7–16.
- [26] Mooney RJ. *Machine Learning Text Categorization*, CS 391L. Austin: University of Texas. 2006.
- [27] Tim Penyusun KB. *Large Indonesian Language Dictionary (in Indoensia: Kamus Besar Bahasa Indonesia (KBBI))*. Badan Pengembangan dan Pembinaan Bahasa). 2016.
- [28] Ruli R, Siregar A, Sinaga FA, Arianto R. Application for Determining Examining Lecturer in Thesis Examination Using Tf-Idf Method and Vector Space Model (in Indonesia: Aplikasi Penentuan Dosen Penguji Skripsi Menggunakan Metode Tf-Idf Dan Vector Space Model). *Computatio: Journal of Computer Science and Information Systems*. 2017; 1(2): 171–186.