Register 6 (1) 2020 pp. 1-10
Jurnal Ilmiah Teknologi Sistem Informasi (Scientific Journal of Information System Technology)

ISSN 2502-3357 (Online)
ISSN 2503-0477 (Print)

Available online to www.journal.unipdu.ac.id

Unipdu

S2-Accredited – SK No. 34/E/KPT/2018
Journal page is available to www.journal.unipdu.ac.id:8080/index.php/register

Clickbait detection: A literature review of the methods used

Nurrida Aini Zuhroh [a], *Nur Aini Rakhmawati* [b]

[a,b] *Information Systems Department, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*

email:[a] *nurrida.18052@mhs.its.ac.id,* [b] *nur.aini@is.its.ac.id*

ARTICLE INFO

Article history:
Received 22 May 2019
Revised 5 July 2019
Accepted 22 July 2019
Published 29 October 2019

Keywords:
clikbait
deep learning
literature review
machine learning
word embedding

IEEE style in citing this article:
N. A. Zuhroh and N. A. Rakhmawati, "Clickbait detection: A literature review of the methods used," *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 6, no. 1, pp. 1-10, 2020.

ABSTRACT

Online news portals are currently one of the fastest sources of information used by people. Its impact is due to the credibility of the news produced by actors from the media industry, which is sometimes questioned. However, one of the problems associated with this medium used to obtain information is clickbait. This technique aims to attract users to click hyperbolic headlines with content that often disappoints the reader. This study was, therefore, conducted to determine: 1) existing dataset available. 2) The method used in clickbait detection which consists of data preprocessing, analysis of features, and classification. 3) Difference steps from the method used.

1. Introduction

Currently, digital media has replaced print media, with an increase in the number of news portals that provides a variety of information. The growth of online media has a negative impact, such as clickbait, which refers to the use of excessive or sensational headlines that aim to attract traffic and clicks to increase site revenue. It is usually written in misleading and provocative sentences [1]. A title is an essential element in the news, which gives an initial impression and influences the user's perception [2]. Lowenstein proposed a theory which states that clickbait is formed by knowledge gaps created by one's curiosity in certain matters. This gap is capable of affecting one's emotions [3].

In 2016, Potthast et al. conducted an initial study in detecting clickbait titles using machine learning methods. The study used data sets obtained from Twitter, which consisted of 2992 tweets, with 767 identified as clickbait [4]. In addition, Chakraborty et al. analyzed 15,000 clickbait and non-clickbait titles using Stanford CoreNLP [5] to determine several characteristics that distinguish between two categories with the first difference in the sentence structure. Clickbait titles tend to have longer sentence structure consisting of hyperbolic words and slang, with 62% used as titles and containing one to 40 words [6]. Rony et al. used two datasets, headlines dataset provided by Chakraborty et al. [6] and media corpus obtained from the Facebook Graph API. The research aimed to analyze the impact of clickbait, which showed that it gets more user responses [7].

A recent study on the detection of clickbait was conducted by Dong et al. [8]. The study measured the similarity/consistency between titles and news. It used the Bidirectional Gated Recurrent Unit (BiGRU) algorithm as a classification method to determine the correlations between headlines and news content with obtained accuracy values above 85% in two different datasets. Its study in Indonesian

N. A. Zuhroh & N. A. Rakhmawati./Register 6 (1) 2020 pp. 1-10
ISSN 2502-3357 (Online)
ISSN 2503-0477 (Print)

articles is still very limited. However, Maulidi & Palandi [9] conducted research using the Naive Bayes classification algorithm (NB). Furthermore, Yavi [10] used the Neural Network (NN) method and obtained an average accuracy of 56%.

There are several studies on clickbait detection, but a review study on its detection is not yet available. A literature review is needed to determine various previous methods and approaches. Therefore, this literature review 1) provides a summary of the earlier methods used, and 2) analyzes several methods and previous research. Furthermore, the structures discussed in this article are research methods, results, discussion, and conclusion.

2. Methodology

This study referred to the guidelines by Kitchenham & Charters [11], which explained the steps for carrying out a systematic literature review, such as:

1. Defining research questions
   It aims to determine the methods used in clickbait detection. The research questions are:
   Q1: What dataset is currently available?
   Q2: What methods are used for its detection?
   Q3: What differences have been found from the method used in its detection?
2. Literature
   The keywords to search for this research literature are clickbait detection methods, and detection review. The search was carried out on several sites such as IEEE, ACM, Elsevier, and Google Scholar.
3. Study criteria selection
   These keywords refer to various literature, with the researcher establishing several specific criteria to identify the most relevant literature for the study. The requirements are:
   - It describes the background, objectives, techniques, or methods used for clickbait detection.
   - It is presented in English or Indonesian.
4. Findings
   The identification process resulted in 21 works of literature, which are presented in Table 1.

Table 1. The findings of the literature study

Year of publication	References	Total
2016	Agrawal [2], Biyani et al. [12], Chakraborty et al. [6], and Potthast et al. [4]	4
2017	Anand et al. [1], Cao et al. [13], Dimpas et al. [14], Fu et al. [15], Kumar et al. [16], Manjesh et al. [17], Rony et al. [18]	8
2018	Geçkil et al. [19], Klairith & Tanachutiwat [20], Maulidi et al. [9], Pandey & Kaur [21], Shu et al. [22], Wongsap et al. [23], Yavi [10], Zheng et al. [24]	8
2019	Dong et al. [8]	1
Total		21

3. Result and Discussion

The researcher identified several detection steps from the literature. These include (1) Collection of clickbait and non-clickbait news articles from several social media and sites, with labeling initially carried out manually by several volunteers [17]. (2) Data preprocessing in the form of cleaning. (3) Analysis of feature selection. (4) Its classification, which explains the techniques used in clickbait detection.

3.1. Data collection

Data were collected from several online or social media sites such as Facebook, Twitter, and Reddit. While the samples representing the clickbait title were obtained from several websites such as /r/SavedYouAClick, @HuffPoSpoilers (Twitter), and StopClickbait (Facebook), which provided information on clickbait-articles and educated the public on ways to avoid such news. Meanwhile, the non-clickbait news was obtained from Wikinews [6], a trusted site that established a set of specific writing procedures that must be obeyed by each of its contributors. Furthermore, it also verifies articles to ensure their reliability. Online literature studies collected are in English (EN), Thai (TH), and Chinese (ZH). Table 2 explains the online version of the dataset.

http://doi.org/10.26594/register.v6i1.1561

ISSN 2502-3357 (Online)
N. A. Zuhroh & N. A. Rakhmawati./Register 6 (1) 2020 pp. 1-10 ISSN 2503-0477 (Print)

Table 2. Online dataset available

No	Dataset	Language	Size	URL	Types
1	Agrawal [2]	EN	Clickbait: 814 Non-Clickbait: 1574	https://github.com/pfrcks/clickbait-detection	News title
2	Chakraborty et al. [6]	EN	Clickbait: 7500 Non-Clickbait: 7500	https://github.com/bhargaviparanjape/clickbait/tree/master/dataset	News title
3	Potthast et al. [25]	EN	Total: 38517	https://www.clickbait-challenge.org/	Tweet
4	Klairith & Tanachutiwat [20]	TH	Clickbait: 15000 Non-Clickbait: 15000	https://github.com/xbklairith/thai-web-crawler	News title
5	Wongsap et al. [23]	TH	Clickbait: 2000 Non-Clickbait: 2000	https://github.com/9meo/Thai-Clickbait	News title
6	Zheng et al. [24]	ZH	Clickbait: 7461 Non-Clickbait: 7461	https://github.com/chenjinyuan87/cbcnn	News title

Table 3. Features used

No	Author	Feature
1	Potthast, Köpsel, Stein, & Hagen [4]	*Teaser message*; link; *meta information*.
2	Chakraborty, Paranjape, Kakarla, & Ganguly [6]	Sentence structure, word patterns; the language often used in clickbait; *n-gram*.
3	Anand, Chakraborty, & Park [1]	*Distributed word embedding*; *character level word embedding*.
4	Biyani, Tsioutsiouliklis, & Blackmer [12]	Content: writing format, unigram, bigram; The similarity of titles with paragraphs 1-n; Use of informal language and pronouns; URL.
5	Manjesh, Kanakagiri, Vaishak, Chettiar, & Shobha [17]	Sentiment analysis; Legibility; Average sentence length and number of words; syntactic structure.
6	Kumar, Khattar, Gairola, Lal, & Varma [16]	*Distributed word embedding*; *character level word embedding*; *document embedding*; *pretrained cnn features*.
7	Cao, Le, Zhang, & Lee [13]	*Post text*/title; content; the relation between title and content
8	Pandey & Kaur [21]	Sentence composition; word structure; language analysis; *n-gram*.
9	Maulidi, Ayilillahi, Isyiriyah, & Palandi [9]	Frequent words; the number of punctuations.
10	Yavi [10]	Frequent words
11	Klairith & Tanachutiwat [20]	*Character level word embedding*; *word level embedding*.
12	Wongsap, Prapphan, Lou, Kongyoung, Jumun, & Kaothanthong [23]	Frequent words; *n-gram*.
13	Shu, Wang, Le, Lee, & Liu [22]	Readability; *word dictionary matching*; *n-gram*; *part-of-speech tags*.

3.2. Data preprocessing

Preprocessing is a data process used to improve classification performance. Data consists of noise (errors or unexpected data) and is inconsistent. Therefore, this technique is carried out to obtain the best dataset amount, structure, and format suitable for each algorithm [26]. It is carried out by erasing all special characters (such as punctuation), leaving only alphanumeric characters [14]. In addition, news title characters need to be changed into lowercase letters to reduce errors during data interpretation, while eliminating stopwords [9, 14].

The results of data analysis by Chakraborty et al. showed that clickbait articles use more stopwords than non-clickbait. Another study carried out in the Thai-language dataset [20] showed that the researchers segmented words using the maximal matching algorithm due to its differing writing format (not using spaces) with Latin. In other studies, based on observations made, clickbait news titles usually contain characters such as "!" and "?" therefore, two datasets are used to determine their impact.

N. A. Zuhroh & N. A. Rakhmawati./Register 6 (1) 2020 pp. 1-10
ISSN 2502-3357 (Online)
ISSN 2503-0477 (Print)

The first dataset consists of several titles containing punctuation, while the second does not contain any [23].

3.3. Feature analysis

Feature analysis needs to be carried out before classification to determine the syntactic and semantic structure patterns in the clickbait title. It was also conducted due to the differences in writing structures. Table 3 describes the features used in each clickbait detection studies.

Table 4. Word embedding model

No	Model	Reference	Total
1	Word2Vec	[2, 21, 24]	3
2	Word2Vec with SkipGram model	[15]	1
3	Word2Vec Extended SkipGram model	[7, 16, 18]	3
4	Word2Vec with Continuous Bag of Words (CBOW) model	[1, 14]	2
5	GloVe	[17, 21]	2

In addition, another technique used for feature extraction is word embedding, which is a representation of distributed words. This technique is often used in Natural Language Processing (NLP) [27] to represent words into a vector in order to produce context or information on semantic and syntactic similarities, as well as their relationship with other words. Table 4 shows some of the word embedding models and algorithms used in clickbait detection.

Anand et al. [1] used Word2Vec and Continuous Bag of Words (CBOW) to obtain lexical and semantic features. The combination of BiLSTM, distributed word embedding, and character-level word embedding showed a better performance than BiRNN and BiGRU. Word2Vec consists of CBOW and SkipGram Mikolov et al. [28], with CBOW used to predict words based on the context. Besides, Mikolov et al. also developed Extended SkipGram as a new distributed word embedding model developed from SkipGram Mikolov et al. [29]. This technique maps sentences to vectors and uses softmax as classifiers Rony et al. [7]. Rony et al. stated that this technique measures the distance between the title and the first paragraph (intro). The test results showed that the method works better than ordinary SkipGram.

Another algorithm used in word embedding techniques is GloVe, which is an unsupervised learning algorithm used to obtain the vector representation of a word Pennington et al. [30]. In a study conducted by Pandey et al., GloVe and Word2Vec algorithms were used as a comparative analysis. The results obtained showed that BiLSTM and GloVe have the best level of accuracy [21].

3.4. Clickbait classification

Table 5. Classification algorithms in machine learning

No	Machine Learning Algorithms	Reference	Total
1	Logistic Regression (LR)	[13, 15, 19, 21, 4, 22]	6
2	Naive Bayes (NB)	[19, 4, 23, 10]	4
3	Random Forest (RF)	[13] [6] [15] [19] [21] [4] [22]	7
4	Support Vector Machine (SVM)	[6] [15] [21] [22] [23]	5
5	Decision Tree (DT)	[6, 15, 22, 23]	4
6	Gradient Boosted Decision Tree (GBD)	[12]	1
7	Gaussian Naive Bayes (GNB)	[17]	1
8	Bernoulli Naive Bayes (BNB)	[17]	1
9	Multinomial Naive Bayes (MNB)	[17]	1
10	XGBoost	[22]	1
11	AdaBoost	[22]	1
12	GradBoost	[22]	1

Currently, machine learning is widely used in various case studies, such as clickbait detection, due to its ability to allow a program to learn datasets [31]. One method used by machine learning in detecting clickbait is the classification algorithm. According to Table 5, several studies use machine learning classification techniques to detect clickbait. Initial research on clickbait detection through a machine learning approach was conducted by Potthast et al. In this study, 215 features were grouped into three categories, namely 1) the teaser message, 2) the linked web page, and 3) the meta-information [4]. The study compared three machine learning classification algorithms, namely Logistic Regression, NB, and

http://doi.org/10.26594/register.v6i1.1561

ISSN 2502-3357 (Online)
ISSN 2503-0477 (Print)

Random Forest. Logistic regression (LR) is a classification algorithm used to obtain nonlinear curves that match the data using different target variables [32]. NB is one of the simplest methods for supervised learning and data mining [33]. It is an algorithm obtained from the Bayes theorem concept and works well on large volumes of data [34]. While Random forest (RF) is a popular algorithm due to its accuracy even with limited samples with quite a lot of features [35]. The results of comparisons conducted by Potthast et al. showed that RF has the best performance than others [4].

Chakraborty et al. analyzed the dataset linguistic using Stanford CoreNLP [5] to obtain the features used in the classification process using SVM, DT, and RF classification algorithms. The study utilized sentence structure, word patterns, language characteristics, and n-grams as the features. SVM is one of the supervised learning algorithms used for classification or regression [36] and to determine the maximum marginal hyperplane [37]. While DT uses tree concept to deduce classification rules based on practical examples [38]. Each node in the tree represents a parameter, while its branch represents a possible value of the connected top node [38]. Chakraborty et al. found different sentence structures in clickbait and non-clickbait titles. Clickbait titles tend to be wordy, contain stopwords and slangs, and hyperbolic with 40 frequent words used as the classification features [6]. Furthermore, Biyani et al. defined several types of clickbait titles using GBD [39]. The clickbait title is usually written in a hyperbole, ambiguous, or unclear statement that affects the emotions of the reader (increase the user's curiosity) [12].

The study of clickbait detection in Indonesian-language articles used NB as a classification method [10]. It measured its effect based on the number of "shares" and "likes" in a news article on Facebook, while other similar research uses Neural Network (NN) with Backpropagation algorithm [9]. Neural Network is a technology that is able to study the input of previous and new data obtained. The process of learning data and NN prediction is carried out through a network of neurons that are connected and arranged into a layered structure [40]. Conversely, Backpropagation is a hierarchical design consisting of layers or rows of processing units that are fully interconnected [41].

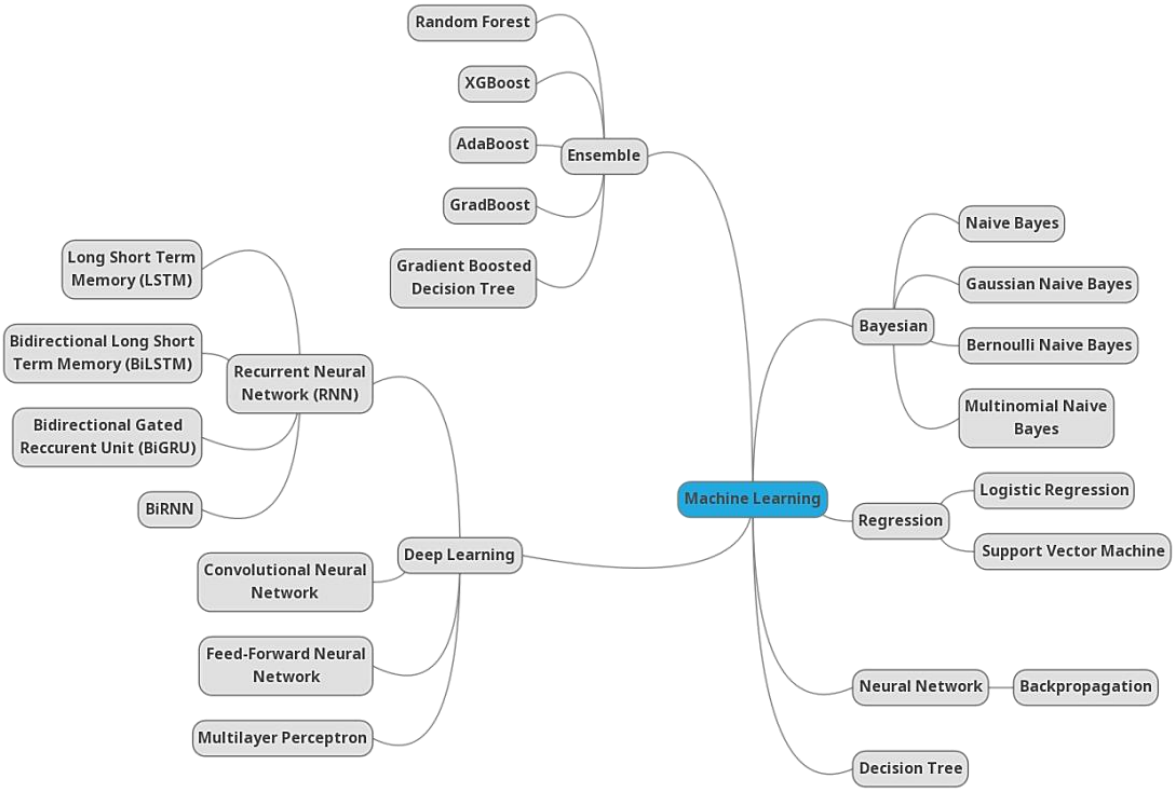

Figure 1. Machine Learning is grouped based on similarity

Table 6 shows some methods of deep learning (DL) approach, which is made and functions based on Artificial Neural Networks (ANN). The science of human cognition is used to understand DL [42], while [15] used the Convolutional Neural Network (CNN) as a classification method. CNN functions automatically to provide a vector that is used as a representation of news headlines. It calculates the phrase vector for every single phrase in a sentence [43]. In addition, its purpose in clickbait detection is

http://doi.org/10.26594/register.v6i1.1561

N. A. Zuhroh & N. A. Rakhmawati./Register 6 (1) 2020 pp. 1-10
ISSN 2502-3357 (Online)
ISSN 2503-0477 (Print)

to represent news headlines as vectors which have fixed size, continuous, and real values as classifiers [15]. Another method used to detect clickbait is Feed-Forward Neural Network (FNN), an algorithm consisting of several layers of neurons, each of which is determined by a set of synaptic weights that are appropriate [44].

Table 6. The method uses a Deep Learning and Neural Network approaches

No	Deep Learning & NN methods	Reference	Total
1	Convolutional Neural Network (CNN)	[2, 15, 24]	3
2	Recurrent Neural Network (RNN)	[14]	1
3	Long Short Term Memory (LSTM)	[17]	1
4	Bidirectional Long Short Term Memory (BiLSTM)	[1, 14, 20]	3
5	Bidirectional Recurrent Neural Network (BiRNN)	[1]	1
6	Bidirectional Gated Recurrent Unit (BiGRU)	[1, 8, 20]	3
7	Feed-Forward Neural Network (FNN)	[20]	1
8	Multilayer Perceptron (MLP)	[17, 21]	2
9	Backpropagation	[9]	1

Recurrent Neural Network (RNN) is a class of ANN and a system that functions in decision making by considering the things and information that already exists. It processes data sequentially, and not suitable for long processing as it is hampered by the learning process. Therefore, the LSTM model is developed to overcome this problem [45].

Manjesh et al. conducted a study using LSTM by analyzing sentiment to determine the classification of a sentence by indicating whether it is positive or negative. The analysis found that non-clickbait titles naturally tend to use more neutral language [17]. This study also analyzed the average number of words used in the clickbait title, which showed that non-clickbait articles tend to contain shorter phrases due to its ability to represent the content of news directly [17].

Manjesh et al. [17] compared several machine learning algorithms using deep learning methods (LSTM and MLP). This study showed that the deep learning method works better than the machine algorithm [17]. Figure 1 show the mind mapping of machine learning of several methods found based on similarity.

3.5. Differences in each approach

The explanation above shows that there are two approaches in the classification method (machine and deep learning algorithm). Some of the steps of the approaches and current research are as follow:

a. Machine Learning

Here are some steps for the GNB, BNB, and MNB algorithms [17]:

1) Data collection/corpus

This stage consists of data collection and classification methods obtained from several news sites and social media. The initial classification was conducted manually by several volunteers through voting.

2) Analysis

Feature analysis is performed to determine semantic and syntactic differences in sentence structure.

3) Classification and testing

At this stage, the research datasets were divided into training, validation, and testing with a ratio of 70:20:10 [17]. The phrases that often appear in clickbait and non-clickbait articles are calculated with N-grams for each title created. Data training is carried out until it gets maximum and constant accuracy values, then dataset testing is obtained.

b. Deep Learning and Neural Network

Anand et al. [1] and Klairith & Tanachutiwat [20] used a deep learning approach to process information hierarchically and automatically while capturing new levels of data abstraction effectively. Although it is capable of handling large data dimensions, a model also requires more examples to carry out maximum exploration [46] through the following steps:

1) Data collection/corpus

Data is collected from several sites or social media. The initial classification is conducted manually using a machine learning method.

http://doi.org/10.26594/register.v6i1.1561

ISSN 2502-3357 (Online)
ISSN 2503-0477 (Print)

2) Embedding layer
It functions as input for the hidden layer and transforms each word into embedded features.
3) Hidden layer and Model
It is a layer that is between the input and output layers that functions to process and study the input data from the embedded layer.
4) Output Layer
At this layer, the model is able to classify clickbait and non-clickbait titles.
5) Testing
Tests were carried out to determine the performance of the model.

c. State of The Art
Dong et al. [8] analyzed the similarities between the title and content of a story. This study used a deep learning approach that is carried out through the following stages:
1) Analysis of latent representation
At this stage, researchers conducted data preprocessing and transformed it into vectors (word embedding) with BiGRU used to determine hidden representations of the hidden layer.
2) Analysis of similarities
Researchers calculated the suitability of the title and content using the cosine similarity algorithm.
3) Prediction
The results obtained in stages 1 and 2 were combined using a fully connected layer to map hidden representations as input to the Multilayer Perceptron.
4) Testing
At this stage, testing was performed on two different databases and obtained accuracy results above 85%.

4. Conclusions

Various researches have recently been conducted on clickbait detection. However, there is limited research on the detection of clickbait in Indonesian-language articles due to the lack of availability of online datasets because majorities are in English, Thai, and Chinese. The 21 literature studies showed that clickbait is detected using a machine learning classification algorithm with deep learning and neural network approaches. Therefore, machine learning-based research often uses the Random Forest algorithm, while deep and neural network makes use of CNN, BiLSTM, BiGRU, and Multilayer Perceptron.

5. References

[1] A. Anand, T. Chakraborty and N. Park, "We Used Neural Networks to Detect Clickbaits: You Won't Believe What Happened Next!," in *39th European Conference on IR Research*, Cham, 2017.

[2] A. Agrawal, "Clickbait detection using deep learning," in *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, Dehradun, India, 2016.

[3] G. Loewenstein, "The psychology of curiosity: A review and reinterpretation," *Psychological Bulletin,* vol. 116, no. 1, pp. 75-98, 1994.

[4] M. Potthast, S. Köpsel, B. Stein and M. Hagen, "Clickbait Detection," in *European Conference on Information Retrieval ECIR*, Cham, 2016.

[5] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard and D. McClosky, "The Stanford CoreNLP Natural Language Processing Toolkit," in *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, Baltimore, Maryland, USA, 2014.

[6] A. Chakraborty, B. Paranjape, S. Kakarla and N. Ganguly, "Stop Clickbait: Detecting and preventing clickbaits in online news media," in *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, San Francisco, CA, USA, 2016.

ISSN 2502-3357 (Online)
ISSN 2503-0477 (Print)

[7] M. M. U. Rony, N. Hassan and M. Yousuf, "Diving Deep into Clickbaits: Who Use Them to What Extents in Which Topics with What Effects?," in *The 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '17)*, Sydney, Australia, 2017.

[8] M. Dong, L. Yao, X. Wang, B. Benatallah and C. Huang, "Similarity-Aware Deep Attentive Model for Clickbait Detection," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining PAKDD 2019: Advances in Knowledge Discovery and Data Mining*, Macau, China.

[9] R. Maulidi, M. F. Ayilillahi, L. Isyiriyah and J. F. Palandi, "Penerapan Neural Network Backprogpagation Untuk Klasifikasi Artikel Clickbait," in *Seminar Nasional FST 2018*, Malang, 2018.

[10] A. F. Yavi, "Klasifikasi Artikel Berbahasa Indonesia untuk Mendeteksi Clickbait menggunakan Metode Naïve Bayes," *J-INTECH Journal of Information and Technology,* vol. 6, no. 1, pp. 141-147, 2018.

[11] B. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering," 2007.

[12] P. Biyani, K. Tsioutsiouliklis and J. Blackmer, ""8 amazing secrets for getting more clicks": detecting clickbaits in news streams using article informality," in *AAAI'16 Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, Phoenix, Arizona, 2016.

[13] X. Cao, T. Le, J. (. Zhang and D. Lee, "Machine Learning Based Detection of Clickbait Posts in Social Media," 2017.

[14] P. K. Dimpas, R. V. Po and M. J. Sabellano, "Filipino and english clickbait detection using a long short term memory recurrent neural network," in *International Conference on Asian Language Processing (IALP)*, Singapore, Singapore, 2017.

[15] J. Fu, L. Liang, X. Zhou and J. Zheng, "A Convolutional Neural Network for Clickbait Detection," in *2017 4th International Conference on Information Science and Control Engineering (ICISCE)*, Changsha, China, 2017.

[16] V. Kumar, D. Khattar, S. Gairola, Y. K. Lal and V. Varma, "Identifying Clickbait: A Multi-Strategy Approach Using Neural Networks," in *SIGIR '18 The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, Ann Arbor, MI, USA, 2018.

[17] S. Manjesh, T. Kanakagiri, P. Vaishak, V. Chettiar and G. Shobha, "Clickbait Pattern Detection and Classification of News Headlines Using Natural Language Processing," in *2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)*, Bangalore, India, 2017.

[18] M. M. U. Rony, N. Hassan and M. Yousuf, "BaitBuster: A Clickbait Identification Framework," in *The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, New Orleans, Louisiana, USA, 2018.

[19] A. Geçkil, A. A. Müngen, E. Gündogan and M. Kaya, "A Clickbait Detection Method on News Sites," in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Barcelona, Spain, 2018.

[20] P. Klairith and S. Tanachutiwat, "Thai Clickbait Detection Algorithms Using Natural Language Processing with Machine Learning Techniques," in *International Conference on Engineering, Applied Sciences, and Technology (ICEAST)*, Phuket, Thailand, 2018.

[21] S. Pandey and G. Kaur, "Curious to Click It?-Identifying Clickbait using Deep Learning and Evolutionary Algorithm," in *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Bangalore, India, 2018.

[22] K. Shu, S. Wang, T. Le, D. Lee and H. Liu, "Deep Headline Generation for Clickbait Detection," in *IEEE International Conference on Data Mining (ICDM)*, Singapore, Singapore, 2018.

[23] N. Wongsap, T. Prapphan, L. Lou, S. Kongyoung, S. Jumun and N. Kaothanthong, "Thai Clickbait Headline News Classification and its Characteristic," in *International Conference on Embedded*

 http://doi.org/10.26594/register.v6i1.1561

ISSN 2502-3357 (Online)
ISSN 2503-0477 (Print)

Systems and Intelligent Technology & International Conference on Information and Communication Technology for Embedded Systems (ICESIT-ICICTES), Khon Kaen, Thailand, 2018.

[24] H.-T. Zheng, J.-Y. Chen, X. Yao, A. K. Sangaiah, Y. Jiang and C.-Z. Zhao, "Clickbait Convolutional Neural Network," *symmetry*, vol. 10, no. 5, 2018.

[25] M. Potthast, T. Gollub, M. Hagen and B. Stein, "The Clickbait Challenge 2017: Towards a Regression Model for Clickbait Strength," 2017.

[26] S. García, J. Luengo and F. Herrera, Data Preprocessing in Data Mining, New York: Springer, 2015.

[27] Y. Li, L. Xu, F. Tian, L. Jiang, X. Zhong and E. Chen, "Word Embedding Revisited: A New Representation Learning and Explicit Matrix Factorization Perspective," in *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)*, Buenos Aires, Argentina, 2015.

[28] T. Mikolov, K. Chen, G. Corrado and J. Dean, "Efficient Estimation of Word Representations in Vector Space," 2013.

[29] T. Mikolov, I. Sutskever, I. Sutskever, G. Corrado and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," in *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, Lake Tahoe, Nevada, United States, 2013.

[30] J. Pennington, R. Socher and C. D. Manning, "GloVe: Global Vectors for Word Representation," in *In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, Doha, Qatar, 2014.

[31] A. Géron, Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, 1st ed., Gravenstein Highway North, Sebastopol: O'Reilly, 2017.

[32] V. Kotu and B. Deshpande, Data Science: Concepts and Practice, 2nd ed., Cambridge, United States: Morgan Kaufmann, 2019.

[33] M. Langarizadeh and F. Moghbeli, "Applying Naive Bayesian Networks to Disease Prediction: a Systematic Review," *Acta Inform Med*, vol. 24, no. 5, p. 364–369, 2016.

[34] M. F. Hornick, E. Marcadé and S. Venkayala, Java Data Mining Strategy, Standard, and Practice: A Practical Guide for Architecture, Design, and Implementation, San Francisco: Morgan Kaufmann, 2007.

[35] R. Pal, "Predictive modeling based on multivariate random forests," in *Predictive Modeling of Drug Sensitivity*, Elsevier, 2017, pp. 189-218.

[36] M. Binkhonain and L. Zhao, "A review of machine learning algorithms for identification and classification of non-functional requirements," *Expert Systems with Applications: X*, vol. 1, no. April, p. 100001, 2019.

[37] J. Han, M. Kamber and J. Pei, Data Mining Concepts and Techniques, Third Edition ed., Wyman Street, Waltham: Morgan Kaufmann, 2012.

[38] G. Shi, "Decision Trees," in *Data Mining and Knowledge Discovery for Geoscientists*, Elsevier, 2014, pp. 111-138.

[39] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367-378, 2002.

[40] X.-S. Yang, "Neural networks and deep learning," in *Introduction to Algorithms for Data Mining and Machine Learning*, Elsevier, 2019, pp. 139-161.

[41] R. Hecht-Nielsen, "Theory of the Backpropagation Neural Network," in *Neural Networks for Perception Computation, Learning, and Architectures*, Elsevier, 1992, pp. 65-93.

[42] R. Nisbet, G. Miner and K. Yale, "Deep Learning," in *Handbook of Statistical Analysis and Data Mining Applications*, 2nd ed., Elsevier, 2018, pp. 741-751.

[43] E. Fathi and B. M. Shoja, "Deep Neural Networks for Natural Language Processing," in *Handbook of Statistics*, vol. 38, Elsevier, 2019, pp. 229-316.

 http://doi.org/10.26594/register.v6i1.1561

ISSN 2502-3357 (Online)
N. A. Zuhroh & N. A. Rakhmawati./Register 6 (1) 2020 pp. 1-10 ISSN 2503-0477 (Print)

[44] S. Theodoridis, "Neural Networks and Deep Learning," in *Machine Learning A Bayesian and Optimization Perspective*, Elsevier, 2015, pp. 875-936.

[45] J.-T. Chien, "Deep Neural Network," in *Source Separation and Machine Learning*, Elsevier, 2019, pp. 259-320.

[46] F. C. Pereira and S. S. Borysov, "Machine Learning Fundamentals," in *Mobility Patterns, Big Data and Transport Analytics Tools and Applications for Modeling*, Elsevier, 2019, pp. 9-29.