Journal of Theoretical and Applied Information Technology
15th April 2021. Vol.99. No 7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FAKE NEWS DETECTION MODEL BASED ON CREDIBILITY MEASUREMENT FOR INDONESIAN ONLINE NEWS

[1,2]RAKHMAT ARIANTO, [3]SPITS WARNARS HARCO LESLIE, [4]YAYA HERYADI, [5]EDI ABDURACHMAN

[1]Department of Information Technology, Malang State Polytechnic, Indonesia

[2,3,4,5]Doctoral of Computer Science, Bina Nusantara University, Indonesia

E-mail: [1]arianto@polinema.ac.id

ABSTRACT

The spread of fake news is a problem faced by all active internet users, especially in Indonesian society. Fake news can have an impact on readers' misperceptions, causing harm to certain individuals or groups so that the Indonesian government issued Law number 19 of 2016 which serves to protect internet users from misinformation from fake news. The research that has been done in detecting fake news in Indonesian is still very much dependent on the results of detection from third parties and the Indonesian government in determining whether a news title is included in fake news or not, but if no similarity is found it will be considered as factual news so that this research proposes a fake news detection model based on the level of credibility of online news headlines. This model has 5 stages, namely: Scrapping Web, Document Similarity, Online News Search, Online News Scoring, and Classification. This research has also tested the use of the K-Means, Support Vector Machine with various kernel, and Multilayer Perceptron methods to obtain optimal classification. The results showed that at the Document Similarity stage an optimal threshold value is needed at 0.6, while the Classification stage determines that the most effective method on the data used is the Multilayer Perceptron with the provisions of Hidden Layer 30,20,10 so that you get a mean accuracy of 0.6 and accuracy maximum of 0.8.

Keywords: *Fake News Detection, Online News, Credibility, Text Mining, Classification*

1. INTRODUCTION

News is info concerning current events. this could be provided through many various media: word of mouth, printing, communicating systems, broadcasting, transmission, or through the testimony of observers and witnesses to events. Common topics for news reports embody war, government, politics, education, health, the surroundings, economy, business, fashion, and amusement, moreover as athletic events, far-out or uncommon events. Government proclamations, regarding royal ceremonies, laws, taxes, public health, and criminals, are dubbed news since precedent days. Technological and social developments, typically driven by government communication and undercover work networks, have inflated the speed with that news will unfold, moreover as influenced its content [1]. However, due to the ease with which people can access and share news online, there are some irresponsible people who share fake news.

The spread of fake news is one of the problems faced by the Indonesian people. With fake news, a certain person or group can harm other people or the target group by using fake news. This is because based on previous research [2] fake news can change a person's behavior or people's thoughts on the subject in fake news so that they carry out harmful behavior in accordance with the fake news they receive. Fake news is easier to spread because internet users in the world are always experiencing an increase due to social media and massive internet usage [3]. In Indonesia, according to a survey conducted by the Indonesian Internet Service Providers Association (APJII) in 2019, it was stated that internet users in Indonesia reached 171.17 million people or 64.8% of the total population of Indonesia with an increase of 10.12% from the previous year's survey. The survey on the main reasons for using the internet reached 24.7% for accessing communication media and 18.9% while the second reason for using the internet was 19.1% for accessing social media and 7% for accessing online news portals[4]. This is also reinforced by research conducted[5], [6] which states that the lack of literacy culture in Indonesian society is one of the strongest supporting factors for the rapid spread of fake news.

The Indonesian government already has a legal basis to take action against perpetrators of the spread of fake news, one of which is Law Number 19 of 2016 [7] which regulates the use and dissemination of information to the public. However, the spread of fake news remains unsettling for the people of Indonesia where on the website pages of the Ministry of Communication and Information Technology of the Republic of Indonesia (Kemkominfo RI) detects fake news that is updated every day, and the number of fake news will increase in events that take people's attention, especially political events, political policies, health, and economics [8].

To avoid fake news, many of research has been done to detect fake news in Indonesian online news, but these studies still revolve around the detection of fake news based on the measurement of the similarity of news headlines entered by users with the results of detection of fake news from third parties such as Masyarakat Anti Fitnah Indonesia (MAFINDO) and the Kemkominfo RI Therefore, has to wait for the detection results first, then it can be determined that a news headline is fake news or factual news. In addition, some of the research that has been studied (listed in the Related Works section) does not reveal the solution if there is no similarity between the news headlines entered by the user and the results of detection from third parties and will even be immediately considered factual news without further analysis.

While the International Federation of Library Associations and Institutions (IFLA)[9]–[11] publishes ways to check the truth of news received, including consider the source of the news or place of the news publication, understand the headline well in order to get an overview of the content of the news and check the date of publication to find out the update of the news. With these problems, a computational framework is proposed to assist the public in checking the truth of the news it receives through 4 stages, namely: Matching news headlines against a collection of fake news from the Ministry of Communication and Information and MAFINDO, measuring the level of credibility of online news based on the credibility value of news headlines, the popularity of news publication place, and time of news publication, measuring the level of credibility of news on social media based on the time of news publication, social media accounts that publish news, and news messages published.

2. RELATED WORKS

2.1. Detection of Fake News in Indonesian Language News

Several studies have detected fake news using online news data in Indonesian, such as in research conducted by Pratiwi et al. [12] where the detection of fake news was carried out by collecting news that had been labeled as fake news from the turnbackhoax.id site and then the similarity level was calculated. Headlines entered by the user and yielded an accuracy value of 78.6%. Next, in a study conducted by Al-Ash and Wibowo [13] where the detection of fake news was carried out by measuring the level of similarity of the resulting phrases and entities in the title sentence entered by the user with a collection of news from turnbackhoax.id and successfully detecting fake news with an accuracy value of 96.74%. Whereas in the research of Prasetijo et al. [14], performed a comparative analysis of the performance of the classification algorithm between Support Vector Machine and Stochastic Gradient Descent to classify fake news from news titles entered by users based on news titles labeled fake news taken from the turnbackhoax.id site where the SGD algorithm can detect fake news better than SVM. In a study conducted by Al-Ash et al. [15], which corrected the weaknesses of previous studies using ensemble learning techniques using the Naïve Bayes algorithm, Support Vector Machine, and Random Forest which resulted in a better accuracy rate of 98% in detecting fake news. Next, in a study conducted by Fauzi et al [16] where fake news detection was carried out using data from social media Twitter and processed using the TF-IDF algorithm and the Support Vector Machine classification algorithm, resulting in the detection of fake news with an accuracy of 78.33%. Whereas the research of Prasetyo et al. [17], analyzed the performance of the LSVM algorithm, Multinomial Naïve Bayes, k-NN, and Logistic Classification in detecting fake news from user input news headlines based on labeled news obtained from the Ministry of Communication and Information of Central Java Province and TurnBackHoax and producing an algorithm. The most optimal is Logistic Classification with an accuracy value of 84.67% which was compared to research from [18]. Furthermore, in Rahmat et al's research [19], fake news was detected from the URL entered by the user then the contents of the URL would be classified as fake news or factual news based on labeled data taken from TurnBackHoax using the Support Vector Machine classification algorithm with a Linear kernel resulting in detection Fake news with 85% accuracy which was compared to research from [14], [20]–[22]. Research [23] has also classified fake news using the Supervised Binary Text Classification algorithm based on previous research [24] and [25] so that an accuracy value of 0.83 is obtained. Meanwhile, research [26] conducted a classification of fake news and sentiment analysis of Indonesian language news using the Naïve Bayes algorithm which was optimized with the Particle Swarm Optimization algorithm. This research was conducted on the basis of the results of previous studies [27]–[30] so that the results showed that the higher the level of negative sentiment in news, the higher the hoax rate. the news. Furthermore, in research [31] classification of fake news has been carried out based on news that has been labeled fake news with news headlines entered by the user using the Smith-Waterman algorithm with an accuracy value of 99.29%. Research [32] has also classified fake news by utilizing feedback from site page users using the Naïve Bayes Classifier algorithm which is based on the results of previous research [33] dan [21] , with the results obtained as 0.91 for precision and 1 for recall. Then in research [34] has also classified fake news using the Term Frequency / Term Document Matrix algorithm and combined it with the k-Nearest Neighbor algorithm so that an accuracy value of 83.6% is obtained.

Based on the literature study that has been explained, it shows that the research carried out on the detection of fake news in Indonesian, the majority only relies on the results of detection from parties who have carried out the labeling of news headlines such as MAFINDO or the Ministry of Communication and Informatics of the Republic of Indonesia and then measuring the level of similarity of news headlines already has a label with the headline entered by the user and involves a classification or clustering algorithm but still excludes the features of how to detect fake news such as those published by IFLA or some fake news detection research conducted on news data other than in Indonesian.

2.2. Fake News Detection

In contrast to research on fake news detection in Indonesian language news, research on fake news detection with data in English or international languages is more varied by paying attention to important features in news published online, such as in research [35] which conducted fake news detection research based on several factors, namely: text characteristics of articles, response characteristics of published news, characteristics of news sources where the analysis includes the structure of the URL, the credibility of the published news sources, and the profile of the journalists who write the news so it is proposed. a model called CSI (Capture, Score, Integrate) which also uses the LSTM algorithm. Using the model on 2 real-world datasets shows the high accuracy of CSI in classifying fake news articles. Apart from accurate predictions, the CSI model also produces a latent representation of users and articles that can be used for separate analyzes.

Next, in the research conducted by [36] regarding rumor detection carried out by dividing which rumors are included in true and false rumors, the results of the separation will be re-analyzed by 3 MIT and Wellesley College undergraduate students so as to produce accurate validation and detection of fake news containing BOT. The results of his research concluded that the spread of fake news was still faster, deeper, and wider due to human intervention to spread the fake news.

Furthermore, in research [37] raises the problem of the spread of fake news that occurred during the US Presidential election in 2016 where Donald Trump spread fake news attacking his competitor, Hillary Clinton, through the social media Twitter. Based on these problems, a model that detects fake news on social media is proposed by classifying the news propagation paths where the model contains (1) The modeling of the propagation path of each news as a multivariate time series where each tuple shows the characteristics of the users involved in spreading the news. (2) Time series classification with recurring and convolutional networks to predict whether the news gave is fake. The results of the experiments conducted show that the proposed model is more effective and efficient than the existing model because the proposed model only depends on the characteristics of users in general compared to more complex features such as linguistic features and language structures that are widely used in the model. existed before.

In research [38] analyzing the detection of fake news can be categorized into 3 categories, namely:

Knowledge-Based or what is commonly called Fact-Checking can be further divided into 2 types, namely: (i) Information Retrieval which is exemplified in research [39] which proposes the detection of fake news by identifying inconsistencies between extracted claims on news sites and the documents in question, [40] detect fake news by calculating the frequency of news that supports a claim, and [41] who questioned the two previous approaches regarding the level of credibility of news used as claims about expertise, trustworthiness, quality, and reliability. (ii) Semantic Web or it can be referred to as Linked Open Data (LOD) is exemplified by research from [42] which "disturbs" the questionable claim to inquire about the knowledge base, using variations in results as an indicator of the support offered by the knowledge base. for these claims, [43] use the shortest path distance between concepts in the knowledge graph, [44] using predictive algorithms on URLs, and all three approaches are not suitable for new claims without an appropriate entry in knowledge-based, while the knowledge base can be manipulated [45].

Context-Based Fake News Detection is a category of fake news detection that analyzes information data and patterns of fake news spread. Several studies that fall into this category are [46] which state that the author's information from the news is an important factor in detecting fake news, [47] trying to determine the truth of claims based on conversations that are appearing on Twitter as one of *RumourEval's* tasks, [48] stated that social media Facebook has a lot of news that has no basis to spread widely and is supported by people who side with conspiracy theories who will be the first to spread the news, whereas in research [49]–[52] have similarities in the proposed model, namely detecting fake news by analyzing the dissemination of false information on social media.

Style-Based Fake News Detection is the detection of fake news based on linguistic forensics and combined with the *Undeutsch* hypothesis which is one of the forensic psychology states about real life, events that are experienced themselves differ in content and quality of the events imagined. This basis is used as a fraud detector on a large scale to detect uncertainty in social media posts as exemplified in the study [53] and [54].

In research [55] conducted a survey of strategies for detecting fake news. In his research, the detection of false news divided into four categories, namely: Knowledge-Based, Style-Based, Propagation-Based, and Source-Based wherein each of these categories has a connectedness that incorporation of each category in the detection of false news very recommended.

Next in research [56] with the problem of spreading fake news carried out by most people without checking the facts of the news on social media, a solution is proposed by utilizing news dissemination on social media to get an efficient and informative latent representation. from news articles. By modeling the echo chamber as a closely connected community in a social network, we present news articles as a 3-mode structure tensor - <News, Users, Communities> and propose a tensor factorization-based method to encode news articles in a latent embedding space that preserves the community structure as well as modeling community and news article content information through the combined tensor-matrix factorization framework. The results of his research found that content modeling information and echo space information (community) together help improve detection performance. Further two additional tasks are proposed to verify the generalizability of the proposed method and then demonstrate its effectiveness over the basic method for the same.

In research [57] proposes a model that contains several indicators to detect credibility-based fake news which is a combination of several research indicators that have been carried out such as the Reputation System, Fact-Checking, Media Literacy Campaigns, Revenue Model, and Public Feedback. This research presents a set of 16 indicators of article credibility, focusing on article content as well as external sources and article metadata, refined over several months by a diverse coalition of media experts. In addition, it presents the process of collecting these credibility indicator annotations, including platform design and annotator recruitment, as well as a preliminary data set of 40 articles annotated by 6 trained annotators and rated by domain experts.

Furthermore, in research [58] proposed a fake news detection model based on the effect of the spread of news or information published on online news or social media by assessing the effects of the news spread. In the proposed model, there are 3 factors in calculating the effect of news dissemination, namely: (1) Scope: using the help of the Text Razer application to find out the coverage of news or new information is given. (2) Publishing Site's Reputation: Used Google Search API to search several websites that publish the same news as the news entered and then compared with the 100 most popular news sites in the US. If you have one of the sites that have been registered, you will get a high score. (3) Proliferator's Popularity: The spread of posts is a crucial characteristic that can influence the impact of fake news. Popular and trusted social media users not only have huge followers, but their posts also receive tons of likes and shares.

The study [59] raised the problem of several shortcomings of the detection of fake news that had been published, including the weakness in the detection of fake news with a content-based approach as exemplified in the study [60] is that this approach can is debunked by fairly sophisticated fake news that doesn't immediately come off as fake. Furthermore, most linguistic features are language-dependent, limiting the generality of this approach. Whereas in a demographic-based approach, users are very dependent on the availability of data on age, gender, education, and political affiliation [61], then the social network structure requires the availability of data connections between users on social media [62] and [63], as well as on user reactions which may contain items from news or "Likes" from other users that are very easy to manipulate [64]

Next in the research [65] tries to offer solutions to several problems, namely (1) What features can detect fake news with high accuracy, (2) Will the combination of word embeddings and linguistic features improve performance in fake news detection tasks? and (3) Which ML algorithm is the most accurate in detecting fake news in several data sets. Based on these questions, research was carried out by (1) Perform an extensive feature set evaluation study, which is to conduct an extensive feature set evaluation study with the aim of selecting an effective feature set to detect fake news articles. (2) Perform an extensive Machine Learning (ML) classification algorithm benchmarking study, which is testing several algorithms included in Machine Learning including Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision Tree (DT), Ensemble Algorithm between AdaBoost and Bagging. (3) Set certain rules and a solid methodology for creating an unbiased dataset for Fake News Detection, namely compiling a dataset consisting of a collection of fake news that has been proven to be fake news from trusted sources and factual news taken from trusted news sources or no fake news detected by the sources used. (4) Quality results in fake news detection, namely calculating the accuracy of all algorithms used to determine fake news from the dataset used.

Next, the research [66] raises the problem of fake news that was rampant in the US Presidential election in 2016 where according to cybersecurity firm Trend Micro, the use of propaganda in elections through social media is the cheapest solution for politicians to manipulate. voting results and distort public opinion. More broadly, the negative effect of fake news is also felt by product owners, customers, and online shops due to opinions containing fake news given to reviews about the products they sell on social media or official selling platforms. So that a fake news detection model is proposed by utilizing the N-Gram algorithm and testing several Machine Learning classification algorithms to determine which algorithm is the most optimal in detecting fake news based on the level of accuracy of each of the machine learning classification algorithms tested, namely: K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Logistic Regression (LR), Linear Support Vector Machine (LSVM), Decision tree (DT), and Stochastic Gradient Descent (SGD).

The next study, namely research [67] raised the problem of the spread of fake news specifically published through a web page. Pages published by suspicious sites are considered unreliable, and pages published by legitimate media outlets as reliable. It uses all kinds of active political misinformation that go beyond ancient partisan biases, and take into account not only unreliable sites that publish fake stories, but also sites that have misleading headline patterns, thinly sourced claims, and which promote conspiracy theories. Based on these problems, a feature called Topic-Agnostic Features is proposed which contains: (1) Morphological Features, this feature is arranged according to the frequency (number of words) of the morphological pattern in the text. This pattern is obtained by tagging a portion of the speech, which assigns each word in the document to a category based on its definition and context. (2) Psychological Features, a psychological feature that captures the total percentage of semantic words in the text. Obtained semantics of words using a dictionary that has a list of words that express psychological processes (personal attention, affection, perception). (3) Readability Features, this feature captures the ease or difficulty of understanding the text. This feature is obtained through the readability score and the calculation of the use of characters, words, and sentences. (4) Web-Markup Features, This feature captures web page layout patterns. The web-markup features include the frequency (number of appearances) of the ad, the presence of the author's name (binary value), and the number of times different categories of the tag group. (5) Feature Selection, this feature uses a combination of four different methods, namely Shannon Entropy, Tree-Based Rule, L1 Regularization, and Mutual Information. The output of this feature will be normalized and the geometric mean to get the value, and (6) Classification, this feature will compare three Machine Learning classification methods, namely K-Nearest Neighbor, Support Vector Machine, and Random Forest. With the SVM settings used Linear Kernel and Cost 0.1. The results of this study when compared with FNDetector [68] get better accuracy.

The research [69], explains the user's perception of false information, the dynamics of the propagation of false information on the online social network, detection, and handling of false information, and false information in the political field. False information can be divided into several types, including (1) Fabricated, information that is completely fictional and has no relationship with existing factual information. (2) Propaganda, false information that aims to harm the interests of certain parties and usually has a political context. (3) Conspiracy Theory, Refers to information that tries to explain a situation or event using a conspiracy without evidence. (4) Hoaxes, news that contains false or inaccurate facts and is presented as valid facts. (5) Biased or One-sided, Refers to information that is very one-sided or biased. In a political context, this type is known as Hyperpartisan news [38] and is news that is very biased towards a person/party/situation/event. (6) Rumors, Refers to information whose truth is ambiguous or has never been confirmed. (7) Clickbait, Refers to the deliberate use of misleading headlines and thumbnails of content on the Web. (8) Satire News, information that contains a lot of irony and humor. Meanwhile, the types of perpetrators from spreading false information consisted of several types, namely; (1) Bots, in the context of false information, bots are programs that are part of a bot network (Botnet) and are responsible for controlling the online activities of several fake accounts with the aim of spreading false information, (2) Criminal / Terrorist Organizations, criminal gangs and organizations terrorists use OSN as a means to spread false information to achieve their goals (3) Activist / Political Organization, Various organizations share false information to promote their organization, bring down other rival organizations, or push certain narratives to the public (4) Governments, Historically, government engaging in the spread of false information for a variety of reasons. Recently, with the rise of the Internet, governments have made use of social media to manipulate public opinion on certain topics. In addition, there are reports that foreign governments share false information about other countries in order to manipulate public opinion on certain topics concerning that particular country. (5) Hidden Paid Posters, they are a special group of users who are paid to spread false information about certain content or target certain demographics, (6) Journalists, individuals who are the main entities responsible for spreading information both to the online world and to the offline world. However, in many cases, journalists are met in the midst of controversy for posting false information for various reasons (7) Useful Idiots, users who share false information mainly because they are manipulated by the leader of an organization or because they are naive. Usually, useful idiots are normal users who are not fully aware of the goals of the organization, therefore it is very difficult to identify them (8) “True Believers” and Conspiracy Theorists, Refers to individuals who share false information because they truly believe they are sharing truth and that others need to know (9) Individuals who have the advantage of False Information, Referring to various individuals who would benefit personally by spreading false information (10) Trolls, The term troll is used mostly by the Web community and refers to users who aim to do things that annoy or annoy other users, usually for their personal entertainment.

This research [70] raises the problem of choosing the most optimal prediction algorithm used in detecting fake news with various features, namely: Features extracted from news content, features extracted from news sources, and features extracted from social network structures. While the algorithms tested are K-Nearest Neighbor (KNN), Naïve Bayes (NB), Random Forest (RF), Support Vector Machine using the RBF kernel, and XGBoost (XGB), each of which has a calculated level of effectiveness using ROC Curve and Macro F1 Score. Classifications that get the best performance value are Random Forest with AUC values 0.85 ± 0.007 and F1 0.81 ± 0.008 and XGB with AUC values 0.86 ± 0.006 and F1 0.81 $\pm$ 0.011 while the error rate obtained from this study is 40%.

Next, in the research [24] the detection of fake news on social media was carried out using the Text Mining algorithm and the classification of text mining results by comparing 23 Supervised Artificial Intelligence algorithms, namely: BayesNet, JRip, OneR, Decision Stump, ZeroR, Stochastic Gradient Descent (SGD), CV Parameter Selection (CVPS), Randomizable Filtered Classifier (RFC), Logistic Model Tree (LMT), Locally Weighted Learning (LWL), Classification Via Clustering (CvC), Weighted Instances Handler Wrapper (WIHW), Ridor , Multi-Layer Perceptron (MLP), Ordinal Learning Model (OLM), Simple Cart, Attribute Selected Classifier (ASC), J48, Sequential Minimal Optimization (SMO), Bagging, Decision Tree, IBk, and Kernel Logistic Regression (KLR) with using Dataset from BuzzFeed Political News, Random Political News, and ISOT Fake News. The results of this study indicate that the Decision Tree gets the Mean Accuracy, Mean Precision, and Mean F-measure values with 0.745, 0.741, and 0.759 while the ZeroR, CVPS, and WIHW algorithms get the highest value on Mean Recall with a value of 1.

Furthermore, research [71] reveals that the main problem of detecting fake news can be divided into 2 aspects, namely: First, news falsehood can come from various perspectives, which are outside the boundaries of traditional textual analysis. Second, the fakeness detection results require further explanation, which is important and necessary for the end user's decision. Based on these problems, a model called XFake is proposed which contains 3 different frameworks, namely MIMIC, ATTN, and PERT.

The next research is research conducted by [72] where research was carried out to answer the problem of how to detect fake news by using Multi-Criteria Decision Making (MCDM) to analyze the credibility of news. This research is inspired by research [73] which uses the factors of the number of user friendships in one application, the number of reviews or comments written by users, the length of reviews written by users, the users' ratings of restaurants, the distance between the results. evaluations from users with global assessments, and standard images, as well as research from [74] which measures the credibility of User-Generated Content on social media using the MCDM paradigm containing Ordered Weighted Averaging (OWA) Operators and Fuzzy Integrals. The features used in his research are Structure Feature, User Related Feature, Content Related Feature, and Temporal Feature. Where all features are used and classified using OWA, Support Vector Machine, Decision Tree, KNN, Naïve Bayes, and Random Forest then evaluated using accuracy, Precision, Recall, F1-Score, and Area Under the ROC Curve. The best results are obtained in the OWA algorithm with the use of 50% and 75% features based on the Accuracy, Precision, and F1-Score values.

Next, in research [75] conducted research on measuring the credibility of news by using emotional signals on social media. This is based on research [36] investigating true and false rumors on Twitter. However, they do not explore the effectiveness of emotions in automatic false information detection, [60] use linguistic information from claims to address the problem of credibility detection, and [76] suggest that a claim can be strengthened by taking supporting evidence taken from the Web. So a system called EmoCred was proposed which incorporated emotional signals into the LSTM to distinguish between credible claims or not. The research explores three different approaches to generating emotional signals from claims: (i) a lexicon-based approach based on the number of emotional words that appear in the claim. (ii) an approach that calculates the emotional intensity expressed in claims using a lexicon of emotional intensity. (iii) neural networks that predict the level of emotional intensity that can be triggered to the user.

In the study [77], this study seeks to understand how individuals process new information and further explore the user's decision-making process behind why and with whom users choose to share this content. The study was conducted by involving 209 participants with measurements made including News Articles, Sharing Knowledge, Credibility, Political Interest, Religiosity, Distraction, and Devices used.

The next research is [78] which proposes a new model for detecting fake news on the social media Facebook. The features used in detecting fake news are Facebook Reaction and the polarity, Vector Space Model, Sentiment analysis, Correlation Coefficient. The results show the success of the proposed model, but include only text comments, and need to include other types of comments including images. In addition, this paper only includes English, therefore, incorporating a multilingual component to the proposed approach is one of the key factors in the future.

The next research is research [79] which proposes a fake news detection model using a rule-based concept where the research is based on previous research [80] which detects fake news using a combination of the Convolutional Neural algorithm. Network (CNN), Bi-directional Long Short Term Memory (Bi-LSTM), and Multilayer Perceptron (MLP) with an accuracy of 44.87% and research [64] that classifies fake news using Logistic Regression and Boolean Crowdsourcing using a dataset of 15,500 Facebook Post and 909,236 users get 99% accuracy. The research conducted was divided into 3 stages, namely (1) Data Gathering, Hoax Detection Generator Data Process, which contained 3 processes, namely Preprocessing, Labeling, and Categorization which resulted in 4 categories of fake news, namely: Hoax, Fact, Information, and Unknown. The next stage is Analysis and Detection Process which contains 2 processes, namely Multi-Detection Language which will contain pattern selection, pattern retrieval, and score assignment and Validity Result which contains Similarity Text to produce fake news word data in percentage form. The results of this study resulted in 12000 reliable datasets.

Furthermore, in research [81], improving the performance of the detection of fake news by adding a synonym count feature as a novelty in their research using the basic model of the research [82]–[84] based on Stance Distance (SD) and Deep

Learning. The steps taken in completing the research are (1) Train Data, which uses 49,972 news with 30% used as validation data, 25,413 data is used as test data. (2) Augmentation, which aims to increase the amount of relevant data based on the desired data. From this process, the resulting amount of data becomes 68,472 news. (3) Preprocessing, which consists of lower-case processes, removing symbols, stopword removal, and tokenizing. (4) Vectorization, which breaks the word into several letter combinations (sub-words) using n-gram with the rule of n = 3 based on research [85]. (5) Modeling, the model used is a two-way LSTM that can go forward and backward, which makes this method better able to read and remember the memory of the previous Bi-LSTM unit. The results of this study obtained F1 0.24 with the detectable classes being Agree, Disagree, and Discuss.

2.3. News Recommendation System

In this study, several studies on the News Recommendation System were also studied as the basis for the formation of the proposed model, especially in the Online News Scoring section of the calculation of Time Credibility. Some of these studies include:

A study [86] conducted a research survey on online news recommendations that focused on the sensitivity of a news session which could be determined by the user himself. This problem is supported by previous research, namely research from [87] which states that online news considers factors such as very short news duration, novelty, popularity, trends, and high magnitude of news that comes every second and [88] who make online news recommendations based on the recommendation paradigm, user modeling, data dissemination, recency, measurement beyond accuracy, and scalability. The conclusion drawn from the survey conducted was that there were many unique challenges associated with the News Recommendation Systems, most of which were inherited from the news domain. Of these challenges, issues related to timeliness, readers' evolving preferences for dynamically generated news content, the quality of news content, and the effect of news recommendations on user behavior are most prominent. A general recommendation algorithm is not sufficient in the News Recommendation System as it needs to be modified, varied, or expanded to a large extent. Recently, Deep Learning based solutions have overcome many of the limitations of conventional recommenders.

Next in the research [89] which conducted research with the problems of the News Recommender System at that time still experiencing problems in the section on the use of news filtering, lack of transparency, diversity, and user control. So that qualitative research is carried out which will answer several problems at once, namely outlining the concept of user control because it reduces all those worries at once: empowering users to apply the News Recommender System more to the needs and interests of users, increasing trust and satisfaction, requiring the News Recommender System to be more transparent and explainable, and reduces the influence of blind spot algorithms. The study was conducted by creating four focus groups, or moderated think-aloud sessions, with newsreaders to systematically study how people evaluate different control mechanisms (in the input, algorithm, and output phases) in a News Recommender Prototype.

Furthermore, in research [90] which investigates, designs, implements, and evaluates a deep learning meta-architecture for news recommendations, in order to improve the accuracy of recommendations provided by news portals, meeting the dynamic information needs of readers in challenging recommendation scenarios. In his research, he focuses on several factors, namely: title, text, topic, and the entity mentioned (for example, people, places). A publisher's reputation can also add credence or discredit an article. News articles also have a dynamic nature, which changes over time, such as popularity and novelty. Global factors that may influence article popularity are usually related to breaking events (for example, natural disasters, or the birth of an actual family member). There are also popular topics, which may continue to be of interest to users (for example, sports) or may follow several seasons (for example, football during the World Cup, politics during presidential elections, and so on). The user's current context, such as location, device, and time of day is also important for determining his short-term interests, as users may vary during and outside of business hours. The proposed solution is to create a framework that includes: (i) A new clustering algorithm called Ordered Clustering (OC) which is able to group news items and users based on the nature of the news and the user's reading behavior. (ii) User profile model created from user explicit profiles, long-term profiles, and short-term profiles. Short-term and long-term profiles were collected from users' reading behavior. (iii) News metadata model which combines two new properties in user modeling, namely: ReadingRate and HotnessRate. Meanwhile, to enrich the news metadata, a new property is defined called Hotness. (iv) News selection model based on submodularity model to achieve diversity in news recommendations. The results show that HYPNER achieved an 81.56% increase in F1 scores and 5.33% in terms of diversity compared to an existing recommendation system called SCENE.

Next is a study [91] that focuses on automatically developing News Recommender Systems whose selection is not influenced by the habit of reading news from certain users. When recommending news in a non-personalized way, there are three basic metrics of interest: recency, importance (analogous to relevance in personalized recommendations), and diversity of recommended news. In resolving these problems, a reciprocal analysis of the current importance achieved by the current news recommendation strategy is carried out and shows them as sub-optimal, proposes a simple but overlooked strategy that selects stories based on their future impact, and demonstrates that it has the potential to achieve novelty importance. Better than the current strategy, proposed practical implementation of future impact-based recommendation strategies, leveraging popularity signals and editorial judgments in predicting future impacts, and developing approaches to eliminate the possibility of having a temporal coverage bias in the recommended stories. The results showed that the Future-impact + diversity + sectional composition factor got an accuracy value of 0.841, Precision 0.627, and Recall 0.742 on The Guardian data and an accuracy value of 0.923, Precision 0.847, and Recall 0.806 on NYTimes data.

Furthermore, in research [92] which focuses on indicators of online news which include update and popularity where the official website of the Chinese University is used, which has recorded 39,990,200 visits to the site between March 1, 2017, to 30. April 2017 with records including user IP, date of user visit, Time of visit, Method (GET), Links visited, and HTTP Status. Stored records are erased for corrupted data (errors generated when the server logs incorrectly and is easily recognizable because they do not match normal data patterns in the same field) and which experience redundancies resulting from including unsuccessful requests, delivery requests data, and requests for images, styles, scripts, and other resources. The next step is to define a session. Each session consists of all records originating from a single visit to the site, and a user may have more than one session to visit the site multiple times during two months. So different users are identified by their IP address (User-IP field); for the same user, if the time interval between two recordings exceeds 30 minutes, they will be split into different sessions resulting in 839,685 sessions. The Mann-Whitney U test was conducted to examine the relationship between news recency/popularity and user clicking behavior. Significant results were obtained for recency ($Z = -15{,}366$, $p < .05$) and popularity ($Z = -17{,}889$, $p < .05$).

Research [93] focuses on the development of a time-based approach to news publication that skips the session of consideration, which summarizes articles that users have interacted with within a short period of time. Research is conducted by formulating news recommendation problems and presenting the main dynamics that govern them, such as news updates, the life span of news articles or topic categories, and the use of sliding time windows to forget old news articles. The proposed method builds a Content-Based user profile, by identifying the main categories of news articles that are of interest to the user (eg, politics, sports, etc.). The sliding time window is used to reduce the impact of previously clicked news articles. In addition, to reveal short-term user profiles, it analyzes the latest news articles that users have read (i.e. in their last session).

Next, research [94] focuses on news published online which is an important source of information that can be used for detection and tracking of events and to analyze the relationship between temporal publishing between different news streams, where this research is a development of previous research. [95] and [96]. So that detection, tracking, and prediction of events from various news streams are carried out and analyzed the temporary publishing patterns of news cables on various platforms and their timeliness in reporting events. The research was conducted with an approach based on discrete dynamic topic modeling and the Hidden Markov Model for event detection and tracking. Then, predict the events that will persist in the next part of the time, which can be important for predicting the facts that will be popular in the future. The use of detected events to group news documents according to the events it describes. Two assessment functions are proposed to rank news cables based on their timeliness by testing methodologies using various collections of news articles and tweets. The experimental results show that, compared to the traditional dynamic topic model, the proposed approach can detect emerging topics (events) in a timely manner.

Furthermore, research [97] states that timeliness is important for Session-Based Recommendation Systems because user preferences, popularity or item characteristics, and temporal semantic information are always changing, which requires a timely recommendation algorithm to capture these changes in a timely manner. The research was undertaken

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

with a focus on proposing a new framework for session-based recommendations, which uses attention mechanisms to capture user behavior and item characteristics while applying a cross-time mechanism to study temporal information. The proposed model derives item dynamic features from its historical users and temporal semantic features from all interaction data, which are integrated into the attention network, resulting in improved timekeeping. Experiments were carried out extensively on three baseline data sets. Several detailed comparative experiments were conducted to demonstrate the benefits and advantages of TEAN. The results of this study have been compared with previous studies (1) POP, Season-Based Recommendation System based on news popularity, (2) Item-KNN, (3) UNLFA [98], (4) FPMC, (5) GRU4Rec, (6) NARM [99], (7) SR-GNN [100], (8) ATRank, (9) DCN-SR [101]. The results of the study indicate that TEAN achieves the best performance in terms of P @ 20 and MRR @ 20 across the three datasets. Compared to ATRank, TEAN increased 1.23%, 1.75%, 1.63% at P @ 20, and 3.41%, 5.21%, 2.43% at MRR @ 20 respectively across the three datasets.

3. MODELLING

In this study, a model was built to detect fake news based on the credibility assessment of online news sources as illustrated in Figure 1. The proposed model has several stages, namely: (1) Web Scrapping, which aims to retrieve a collection of fake news that has been detected by a third party, namely TurnBackHoax and fake news reporting on the website of the Ministry of Communication and Information of the Republic of Indonesia (Kemkominfo RI). MAFINDO website is used as a source of labeled datasets because the website initiates the trend of detecting fake news based on public reporting. The Ministry of Communication and Information also detects fake news on news circulating in the community as a form of attention from the Government of the Republic of Indonesia in reducing the spread of fake news. (2) Document Similarity Measurement, which aims to calculate the level of similarity between news headlines entered by the user and news headlines on the labeled dataset. This is done as a first step in detecting news headlines so that it can speed up the process of detecting fake news and implementing the results of previous research. If at this initial stage, news headlines that are similar to news titles in the labeled dataset can be determined, it can be directly determined that the news headlines entered by the user are fake news. (3) Online News Search, aims to find news that is similar to the news headline entered by the user. This process is done as a form of solution if in the process Document Similarity does not produce a similarity level value above the threshold so it is necessary to look for a collection of online news headlines that are similar to the headlines entered by users. (4) Online News Scoring, this stage aims to calculate the credibility level of online news sources that have been collected at the Online News Search stage where the calculation is based on 3 determining factors, namely: (a) Time Credibility where the credibility of an online news source is determined by the time of publication of the news with the longer terms the news is published, it will get a smaller credibility value and vice versa. (b) Message Credibility, which measures the similarity of news titles obtained at the Online News Search stage because online search results do not always get news titles that are exactly similar. The higher the similarity value that online news titles have with user input, the higher the credibility value will be. (c) Website Credibility, where the measurement of the level of credibility is based on 3 things, namely: global website rankings, website rankings according to visitor countries and online news sources, and the number of links contained in these online news sources. On the ranking of website pages globally and by country, it will get a high credibility score if it gets a small value, while on the number of links, the more links in an online news source, the higher the credibility score will be. (5) The results of the credibility assessment will be classified to determine the news titles entered by the User, including fake news or fact where in this research tested using the K-Means, Support Vector Machine, and Multilayer Perceptron methods to find out which method can classify optimally.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

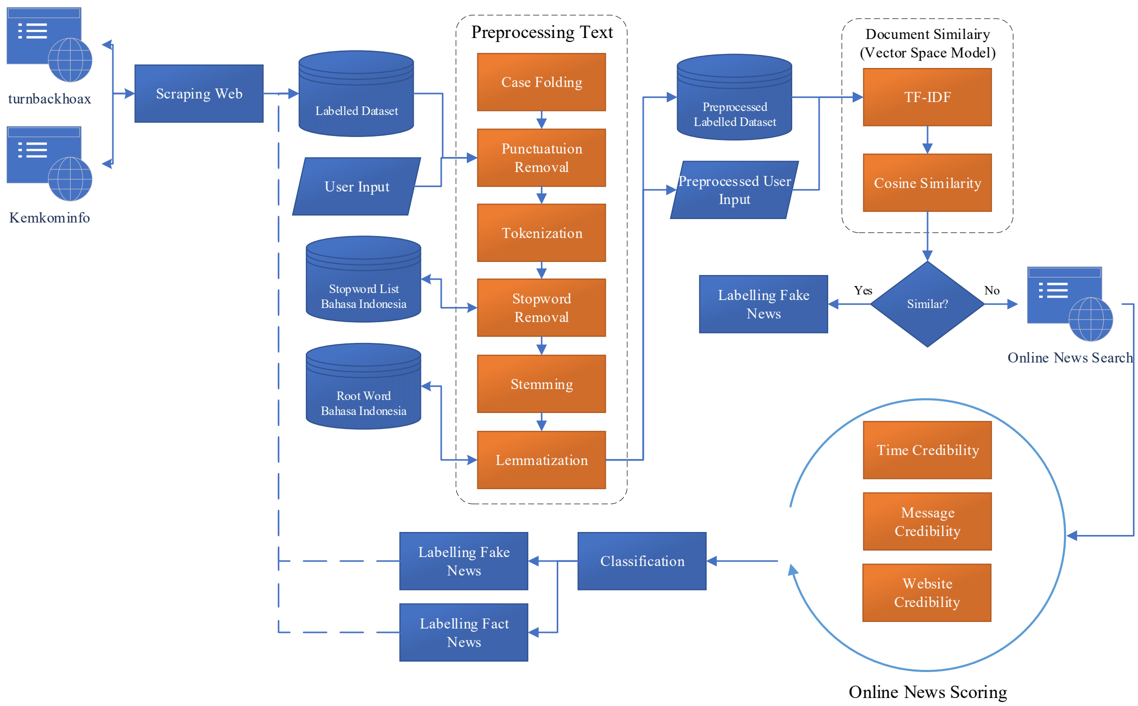

Figure 1. Model of Fake News Credibility Scoring

4. RESULT AND ANALYSIS

5.1. Scraping Web

In the web scrapping stage, two main sources of labeled datasets were used, namely: sites managed by the Ministry of Communication and Information of the Republic of Indonesia [8] and sites managed by MAFINDO [102]. The site managed by the Ministry of Communication and Information only has two labels, namely DISINFORMATION and HOAKS where the two labels are classified as fake news so that 5405 news titles are labeled as fake news within the publication period of February 2, 2018 to June 10, 2020. While on the site MAFINDO, there are several labels, among others: to get 3633 news titles that are included in the fake news category and 1069 news titles that are included in the fact news category within the publication period of 31 July 2015 to 11 June 2020.

5.2. Preprocessing Text

At this stage, a preprocessing of news titles from user input is carried out as well as a collection of news titles in the dataset of labeled news titles as a result of the Web scraping stage. The steps taken are (1) Case Folding, which aims to change all letters in a sentence to lowercase. (2) Punctuation Removal, which aims to remove all punctuation marks in sentences (3) Tokenization, which aims to cut each sentence into a collection of words in the form of an Array. (4) Stopword Removal, which aims to remove words that do not have an important meaning in sentences or words that appear too often. The Stopwordd list used is the Indonesian Language Stopword [103] (5) Stemming word list, which aims to remove the additions, insertions, and endings of each word in a sentence. (6) Lemmatization, which aims to return the form of words resulting from the Stemming process into standard words in accordance with the writing rules of the Big Indonesian Dictionary.

5.3. Document Similarity

Measurement of the level of similarity of news titles entered by the user with a collection of news headlines from the labeled dataset uses the Vector Space Model method where each document will be weighted using TF-IDF and the weighting results will be calculated for its proximity using Cosine Similarity. The weighting of each document is done using a formula:

$$W_{i_j} = tf_{i_j}\, x\, idf_j \qquad (1)$$

$$W_{i_j} = tf_{i_j}\, x \log\left(\frac{D}{df_j}\right) \qquad (2)$$

Where W_{i_j} is the weight of the term (t_j) to the document (d_i). Meanwhile tf_{i_j} is the number of occurrences of the term t_j in the document d_i. D is the number of all documents in the database and df_j is the number of documents containing the term (t_j) (at least one word is a term (t_j)). Regardless of the value of $\left(tf_{i_j}\right)$, if $D = df_j$, then the result will be 0 (zero), because the result is $\log 1$, for the IDF calculation. For this reason, a value of 1 can be added on the IDF side, so that the weight calculation is as follows:

$$W_{i_j} = tf_{i_j}\, x \log\left(\frac{D}{df_j}\right) + 1 \quad (3)$$

Meanwhile, to calculate the closeness of each document weight, the Cosine Similarity formula is used as follows:

$$cos\theta = \frac{A.B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i}} \quad (4)$$

Where A_i and B_i are components of vector A and B respectively. The coming about similitude ranges from -1 meaning precisely inverse, to 1 meaning precisely the same, with showing orthogonality or decorrelation, whereas in-between values show middle of the road closeness or divergence.

If the results of the closeness of each document have been obtained, the next step is to determine the threshold value of all document similarity values by validating each value by using 16-Fold Cross-Validation to produce the most optimal threshold value at 0.6 as in Figure 2.

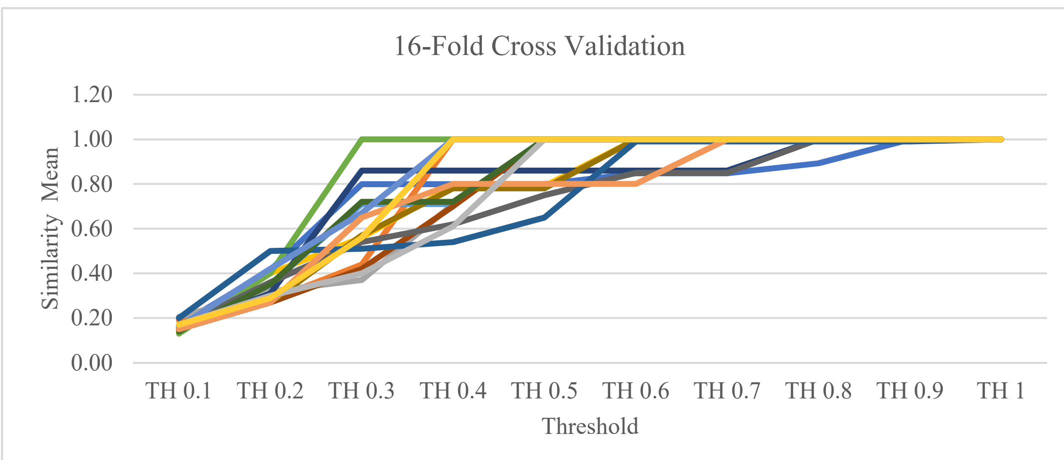

Figure 2. The Result of Determining the Threshold Value

5.4. Online News Search

If the input of news titles from the User is not detected at the Document Similarity stage, the news titles will be searched online using the Google search engine platform so that a collection of news related to the news titles entered by the User will be obtained. This research setting used the Googlesearch-python 2020.0.2 [104] program library in the Python programming language. The news title entered by the User will be input into the search engine without going through the Preprocessing Text process by adding the word "news" in front of the sentence and "-youtube" to eliminate search results from youtube.com so that more search results will be on the site page in the form of news.

5.5. Online News Scoring

5.5.1. Time Credibility

The credibility measurement at this stage uses the publication date of the online news sources that have been collected. Based on previous research studies that have been carried out, online news publications carried out by journalists are always published at the same time and pay attention to the time of occurrence of each news that will be published so it is assumed that if the collection of online news collected has a high closeness to the publication date then will score a high level of credibility.

This research was carried out in several stages, namely: (1) The publication date of each online news will be calculated the difference from the date the User entered the news title. (2) The results of the difference from that date, data normalization is carried out using the Min-Max to get the results in a value between 0 to 1. (3) Data Normalization Results, the Mean Arithmetic value will be searched to represent the value of the set of date differences with units day. (4) Tested against news headlines labeled as fake news and also on news titles labeled as factual news.

Tests that have been carried out at the Time Credibility stage use online news titles labeled fake news and online news titles labeled factual news, so the mean value of the difference in date from fake news titles is 624.95 after data normalization has been carried out, has a mean value of 0.3051 while the difference is the date using factual reports shows the mean value of 24.78 after data normalization is carried out, has a mean value of 0.1003. The two results show that fake news has a lag or publication time that is not close to each other so that it gets a higher Mean value compared to using fact news because in fact news, the search results for similar news headlines using Google, are published nearby.

5.5.2. Message Credibility

At this stage, the credibility level of the titles of each URL that has been collected is calculated at the Online News Search stage with the following steps: (1) The title of the news entered by the User will be Preprocessing Text. (2) The collection of news headlines that have been collected is also carried out by Preprocessing Text. (3) User's news headline is positioned as a query and the collection result collection will be positioned as Document. (4) Weighting of each Document is performed on the query. (5) Document Similarity method is performed to determine the similarity level of the Document to the query. (6) Arithmetic mean is performed to represent the similarity level result set from the Document.

In testing using fake news, the mean value of the similarity level value is 0.04185073935818, while in testing using fact news, the mean value of the similarity level value is 0.3406004840781. At this stage, data normalization is not carried out because the results of the level of similarity of online news titles are in the range of 0 to 1. The results obtained in testing using fake news have a smaller Mean value because online news titles are obtained from searching online news titles using Google tends to have a low degree of similarity. Whereas in the test using fact news, it has a higher Mean value due to the tendency of various online news sources to publish news titles that have a high level of similarity.

5.5.3. Website Credibility

At this stage, URL data will be used from each of the news titles that have been obtained in the previous stage and information will be taken about page rankings in Indonesia, global rankings, and the number of links on web pages using the Alexa Rank service [105].

The value obtained from the Website Credibility stage on news headlines labeled as fake news gets a mean value of 1848299.6 on the Alexa Traffic Rank, 502721.05 on the Country Rank, and 5121.55 with the results of Data Normalization 0.19, 0.05, and 0.11. Whereas in the trial using news labeled as factual news, the results showed 238257.44 on the Alexa Traffic Rank, 5401.44 on the Country Rank, and 8810 on Sites Linking In with the results of Data Normalization 0.08, 0.07, and 0.19.

Based on these results, fake news tends to have a smaller Alexa Traffic Rank and Country Rank compared to the results of using fact news. This is influenced by the collection of fake news source URLs that have been collected is less popular or do not have a good track record of internet users, while the Sites Linking In value, fake news has a smaller inclination value compared to the results of using fact stories. This is influenced by sources that publish fake news having the availability of links on their web pages are few.

The problem encountered when carrying out the Website Credibility stage is that some URLs from sources, both in the use of fake news and factual news, are not detected by Alexa Rank so that the number 9999999 is used as a consequence value so that the data is outside the collection of data that has been detected. When the ranking data have a difference that is too far between the smallest value and the largest value, then the results of data normalization only produce numbers 0 and 1 so that there is no visible difference between the data.

5.6. Classification

At this stage, 2000 data were used, consisting of 1000 data labeled fake news which were taken from the turnbackhoax.id site and fake news reports on the Ministry of Communication and Information website which were not included in the Scraping Web stage and the next 1000 data were news titles taken from news sources. online which is often used by the Indonesian people such as detik.com, kompas.com, and tribunnews.com where the headline is assumed to be factual news due to its high popularity and the provisions of journalists who publish news based on applicable laws. All data will be tested using three types of methods, namely: K-

Journal of Theoretical and Applied Information Technology
15th April 2021. Vol.99. No 7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Means Clustering, Support Vector Machine with various kinds of kernels used, and Multilayer Perceptron with different rules for the number of hidden layers. In the SVM and Multilayer Perceptron methods, the K-Fold Cross Validation method will be used to determine the accuracy, precision, and recall values of the tests carried out.

Testing of test data begins with the K-Means method as a representative of the data clustering method. The steps taken in this test follow the K-Means algorithm [106] are: (1) Determine the number of clusters, namely 2 clusters because it will look for groups of fake news and fact news. (2) Determine the centroid point of each random data group. (3) Calculate the distance of each data to the centroid point that has been determined by the formula:

$$d(x_i, x_j) = \left(|x_{i1} - x_{j1}|^g + \cdots + |x_{ip} - x_{jp}|^g\right)^{\frac{1}{g}} \quad (5)$$

Where:

- g = 1, to calculate the Manhattan distance
- g = 2, to calculate the Euclidean distance
- g = ∞, to calculate the Chebychev distance
- x_i, x_j are two pieces of data that will be calculated the distance
- p = dimensions of a data

Based on the test results with the terms of searching for 2 clusters and limiting iterations up to 1000 times, it is found that Cluster 0 can be equated with label 0 or fake news has a number of members of 786 documents and Cluster 1 which can be equated with label 1 or fact news of 1214 documents as shown. shown in Figure 3 so that the accuracy obtained from testing using the K-Means Clustering method is 80.4% and has an error rate of 19.5%.

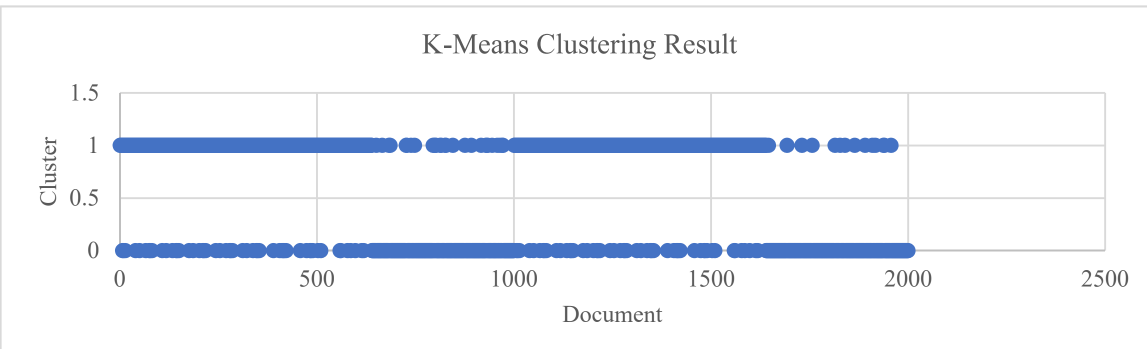

Figure 3. K-Means Test Result

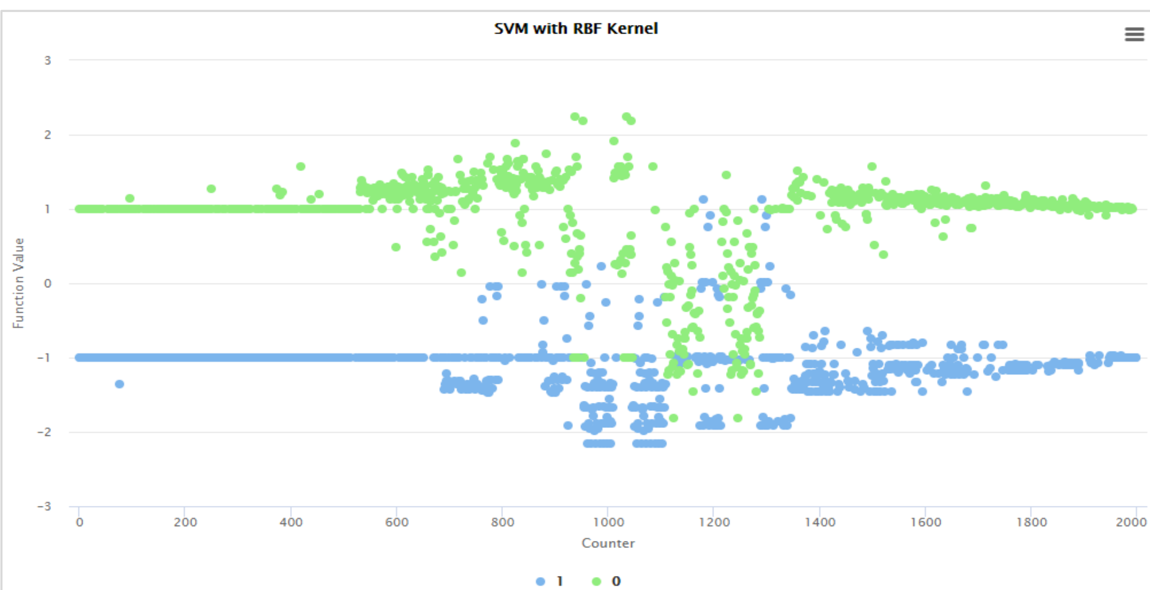

Figure 4. Support Vector Machine with Dot Product Kernel Testing Result

The next test is that 2000 test data results from Online News Scoring will be classified using the Support Vector Machine (SVM) method. In testing using SVM, it will be tested with a dot product kernel, radial., And polynomial. Each data will be calculated using the SVM formula [107], namely:

$$\left[\frac{1}{n}\sum_{i=1}^{n}\max\left(0,1-y_i(w^Tx_i-b)\right)\right]+\lambda||w||^2 \qquad (6)$$

Where, n is the amount of data to be processed, y_i is the value 1 or -1 which indicates the class of x_i, w^T is the transpose value of the normal vector, b is the bias value, and λ is the value that determines the amount of the margin

Based on testing the test data using the Support Vector Machine method by testing the Dot Product, Radial, and Polynomial kernels as shown in Table 1, it can be concluded that with the test data used, the most optimal kernel is the Radial Basis Function (RBF) with a Precision value of 96.87%, Recall 88.20%, and an Accuracy value of 92.65% so that this value will be compared with other test methods.

Table 1. Comparison of Kernel Usage In SVM

Kernel	Precision	Recall	Accuracy
Dot Product	84.29%	73.20%	79.80%
Radial Basis Function	96.87%	88.20%	92.65%
Polynomial	77.50%	58.10%	70.50%

The next test is to use the Multilayer Perceptron method wherein the testing process changes the number of hidden layers where what is being tested is a combination of (1) 10,10,10, (2) 10,20,30, (3) 30,20,10, (4) 20,20,20, and (5) 30,30,30 and each test was carried out with an epoch number of 100. For each test with a different number of hidden layers, the best accuracy value will be determined.

Based on Table 4:22 shows the results of the Multilayer Perceptron test with a different number of hidden layers, it can be concluded that the highest accuracy value with a value of 87.20%, 90.62% precision, and 83.30% recall is the number of hidden layers 20,20,20 so that the accuracy value is will be compared with accuracy values from other methods.

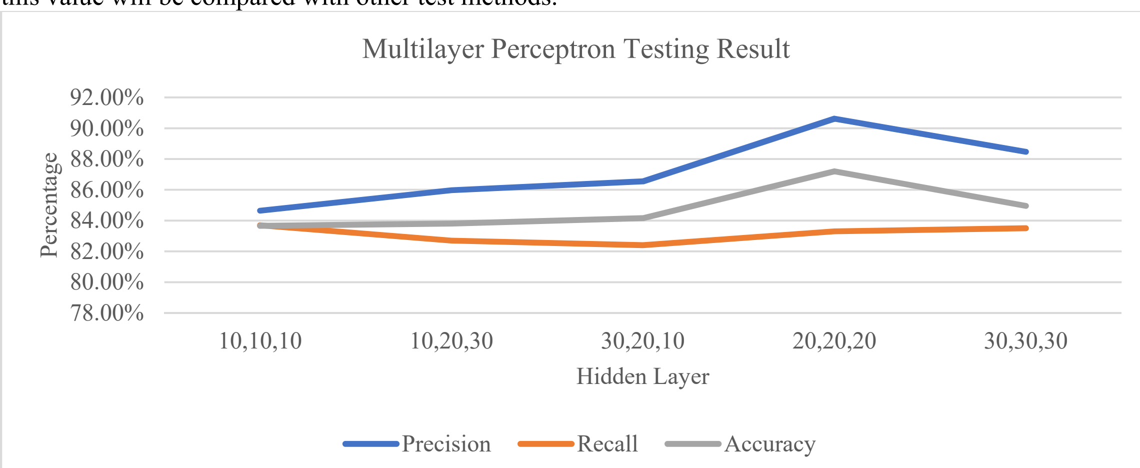

Figure 5. Comparison of Hidden Layer Combinations on The Multilayer Perceptron

Testing data test consisting of 2000 headlines with the number of news labeled fake news as many as 1000 titles and 1000 headlines labeled fact news by using the K-Fold Cross Validation technique in each test used K = 10 so that the comparison value of Accuracy was obtained. and the error rate as shown in Table 2. Based on this comparison, it shows that the most optimal method used in determining fake news or factual news with the test data used is the Support Vector Machine method using the Radial Based Function kernel which has an Accuracy value of 92.65% and a Classification Error of 7.35%.

Table 2. The Results of The Accuracy and Error Rate of Each Method

Method	Accuracy	Error
K-Means	80.41%	19.59%
SVM (RBF Kernel)	92.65%	7.35%
MLP (20,20,20)	90.62%	9.38%

At the time of testing, there were several detection errors that were influenced by several things, such as: (1) In the data normalization process using the Min Max method, if the data has a small or large difference it will still be converted to 0 and 1. For example, when the data is [5,6], the results of normalization are [0,1] as well as data [10, 100] will result in normalization [0,1] so that it will affect the mean value and the classification results. (2) At the Time Credibility stage, the detection of the publication date of online news, not all places for online news publication can be detected by using libraries that are owned by the Python Programming Language so that in this study, it will be replaced with the largest value in the Mean value of the difference in publication dates, namely 1. This This happens because the templates used by each online news source are different, so even though the news is included in factual news, which should have a small value, it will be replaced with the largest value. (3) At the Message Credibility stage, the process of assessing the level of similarity of online news titles always gets the best value if the sentence title contains the same sentence with the actual meaning being the opposite as in the online news title containing the word "Fact:". "Check the Facts", and so on. So that in the Message Credibility stage you will get the best score even though you should get a bad score. (4) At the Website Credibility stage, the online news source ranking measurement based on Alexa Rank will get a good ranking because the news source includes fake news headlines but the news source is a well-known online news source, for example, online news sources include the title "*foto satelit pangkalan militer china di natuna, cek faktanya*" where the title is an online news title that was detected on Turnbackhoax as fake news but the title is used by online news sources detikcom so that the Alexa Rank ranking will get a good ranking.

5. CONCLUSION

Features used in detecting fake news are the online news publication date from the source obtained as a solution to find out that news is not past news, online news headlines that describe the content of the news well, and online news sources are news sources that are trustworthy is not a news source specifically designed to spread fake news.

The model proposed for detecting fake news based on the level of credibility consists of: Scraping Web to get a collection of online news that has been determined to be fake news, Document Similiarity which functions to measure the level of similarity of news titles with news titles labeled as fake news by the Ministry of Communication and Information and Turnbackhoax, Time Credibility which functions as a measure of credibility level based on the time of publication of news, Message Credibility which functions as a measure of the credibility level of news headlines obtained from online news searches, Website Credibility which functions to calculate the level of credibility of online news sources, and Classification using the Support Vector Method Machine with Radial Based Function kernel.

In the data normalization process using the Min Max method, if the data has a small or large difference it will still be changed to 0 and 1. This is because the pattern of the Min Max method is looking for the smallest and greatest values without paying attention to the Standard Deviation of the normalized data set.

In the Time Credibility stage, it shows that online news detected as factual news will get a smaller Mean value when compared to online news detected as fake news. This is influenced by the fact that several online news sources will publish online news simultaneously or in a small time difference.

Errors that occur at the Time Credibility stage are influenced by the ability of the libraries used in the Python Programming Language to detect the date of news publication against each of the templates used by online news sources.

In the Message Credibility stage, it shows that online news that is detected as factual news will get a greater value than online news that is detected as fake news because some online news sources publish news headlines that have a high degree of similarity.

Errors that occur at the Message Credibility stage are influenced by the process of assessing the level of similarity of online news headlines always getting the best value if the sentence title contains the same sentence with the actual meaning is the opposite as in online news headlines containing the word "Fact:". "Check the Facts", and so on.

In the Website Credibility process, online news detected as factual news will get a smaller Alexa Traffic Rank, a smaller Country Rank, and a larger Site Linking In when compared to news headlines that are detected as fake news due to factual news. will be published more by reputable news sources than fake news.

Errors in the Website Credibility process affected by the ranking of online news sources based on Alexa Rank will get a good ranking because the news source includes fake news headlines but the news source is a well-known online news source.

Testing methods carried out at the Classification stage using 1000 news headline test data detected as fake news and 1000 news headlines that are assumed to be factual news and the method used is K-Means with an accuracy value of 80.41%, SVM with the RBF kernel get an accuracy value of 92.65%, and Multilayer Perceptron with an accuracy value of 90.62% so that the most optimal method for the test data used is the Support Vector Machine method with the Radial Based Function kernel.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

For further research, in the data normalization process, it is necessary to use a method other than Min Max or to be replaced with Standardized Data which has a Standard Deviation factor in transforming data so that the range between data can be detected properly.

At the Time Credibility stage, it is necessary to provide a solution to determine the value obtained if the publication date of online news sources is not detected by the librarian used in the Python programming language so that online news titles labeled factual news will not have an identical value with the value of online news that is labeled. fake news.

At the Message Credibility stage, it is necessary to detect types of sentences that have opposite meanings, detect types of interrogative sentences, and types of sentences that contain the word negation because some online news sources use sentence patterns with sentences meaning the opposite, interrogative sentences and contain the word negation as news headlines.

At the Website Credibility stage, it is necessary to filter the news headlines obtained from fake news detection news sources such as Turnbackhoax or the Ministry of Communication and Information because the results of the ranking by Alexa Traffic both sites have good ratings and improvements at the Message Credibility stage will also affect the site address that is processed in the Website Credibility stage.

The use of the structure analysis feature of online news site pages that pays attention to the number of advertisements on the page because in previous studies it has been revealed that every advertisement included in a web page has paid attention to the number of visitors from the site page so that it can increase the credibility of the news page.

News writers on online news are one of the important factors in measuring the level of credibility of the online news so that a list of journalists who have high credibility is needed or can be measured from the published history of the news and need to measure the sentiment analysis of each news published by each -Each journalist.

The use of entities in previous studies can be used as a measure of the credibility of published news because according to journalism research, figures in published news will increase the level of readership of the news.

REFRENCES:

[1] D. Harper, "'News' on Online Etymology Dictionary," *Online Etymology Dictionary*. https://www.etymonline.com/word/news (accessed Mar. 16, 2021).

[2] J. De keersmaecker and A. Roets, "'Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions," *Intelligence*, vol. 65, pp. 107–110, Nov. 2017, doi: 10.1016/j.intell.2017.10.005.

[3] O. K. Lekik, S. Palinggi, and I. C. Ranteallo, "The Descriptive Analysis of Hoax Spread through Social Media in Indonesia Media Perspective:," in *Proceedings of the 1st International Conference on Anti-Corruption and Integrity*, Jakarta, Indonesia, 2019, pp. 276–286, doi: 10.5220/0009441402760286.

[4] Asosiasi Penyelenggara Jasa Internet Indonesia, "Penetrasi & Profil Perilaku Pengguna Internet Indonesia Tahun 2018," 2019. [Online]. Available: www.apjii.or.id.

[5] T. Paterson, "Indonesian cyberspace expansion: a double-edged sword," *J. Cyber Policy*, vol. 4, no. 2, pp. 216–234, May 2019, doi: 10.1080/23738871.2019.1627476.

[6] S. E. Rahmaniah, Rupita, and Hayat, "STOP HOAX Indonesia: Digital Literacy and Education to Prevent Hoax and Hate Speech In the Regional Head Election of West Kalimantan 2020.," *Talent Dev. Excell.*, vol. 12, no. 2, 2020.

[7] R. Indonesia, *Undang-Undang Republik Indonesia Nomor 19 Tahun 2016 Tentang Perubahan Atas Undang-Undang Nomor 11 Tahun 2008 Tentang Informasi Dan Transaksi Elektronik*. 2016.

[8] Kementerian Komunikasi dan Informatika, *Laporan Isu Hoaks*. 2020.

[9] Internasional Federation of Library Associations, *How to Spot Fake News*. 2017.

[10] S. Creagh and W. Mountain, *How we do Fact Checks at The Conversation*. 2017.

[11] IFCN, *International Fact-Checking Network fact-checkers' code of principles*. 2016.

[12] I. Y. R. Pratiwi, R. A. Asmara, and F. Rahutomo, "Study Of Hoax News Detection Using Naïve Bayes Classifier In Indonesian Language," in *2017 11th International Conference on Information & Communication Technology and System (ICTS)*, 2017, pp. 73–78, doi: 10.1109/ICTS.2017.8265649.

[13] H. S. Al-Ash and W. C. Wibowo, "Fake News Identification Characteristics Using Named

Entity Recognition and Phrase Detection," in *2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE)*, Kuta, Jul. 2018, pp. 12–17, doi: 10.1109/ICITEED.2018.8534898.

[14] A. B. Prasetijo, R. R. Isnanto, D. Eridani, Y. A. A. Soetrisno, M. Arfan, and A. Sofwan, "Hoax detection system on Indonesian news sites based on text classification using SVM and SGD," in *2017 4th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, Semarang, Oct. 2017, pp. 45–49, doi: 10.1109/ICITACEE.2017.8257673.

[15] H. S. Al-Ash, M. F. Putri, P. Mursanto, and A. Bustamam, "Ensemble Learning Approach on Indonesian Fake News Classification," in *2019 3rd International Conference on Informatics and Computational Sciences (ICICoS)*, Semarang, Indonesia, Oct. 2019, pp. 1–6, doi: 10.1109/ICICoS48119.2019.8982409.

[16] A. Fauzi, E. B. Setiawan, and Z. K. A. Baizal, "Hoax News Detection on Twitter using Term Frequency Inverse Document Frequency and Support Vector Machine Method," *J. Phys. Conf. Ser.*, vol. 1192, p. 012025, Mar. 2019, doi: 10.1088/1742-6596/1192/1/012025.

[17] A. Prasetyo, B. D. Septianto, G. F. Shidik, and A. Z. Fanani, "Evaluation of Feature Extraction TF-IDF in Indonesian Hoax News Classification," in *2019 International Seminar on Application for Technology of Information and Communication (iSemantic)*, Semarang, Indonesia, Sep. 2019, pp. 1–6, doi: 10.1109/ISEMANTIC.2019.8884291.

[18] M. Aldwairi and A. Alwahedi, "Detecting Fake News in Social Media Networks," *Procedia Comput. Sci.*, vol. 141, pp. 215–222, 2018, doi: 10.1016/j.procs.2018.10.171.

[19] M. A. Rahmat, Indrabayu, and I. S. Areni, "Hoax Web Detection For News in Bahasa Using Support Vector Machine," in *2019 International Conference on Information and Communications Technology (ICOIACT)*, Yogyakarta, Indonesia, Jul. 2019, pp. 332–336, doi: 10.1109/ICOIACT46704.2019.8938425.

[20] S. Gilda, "Notice of Violation of IEEE Publication Principles: Evaluating machine learning algorithms for fake news detection," in *2017 IEEE 15th Student Conference on Research and Development (SCOReD)*, Putrajaya, Dec. 2017, pp. 110–115, doi: 10.1109/SCORED.2017.8305411.

[21] M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," in *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, Kiev, May 2017, pp. 900–903, doi: 10.1109/UKRCON.2017.8100379.

[22] I. Santoso, I. Yohansen, Nealson, H. L. H. S. Warnars, and K. Hashimoto, "Early investigation of proposed hoax detection for decreasing hoax in social media," in *2017 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)*, Phuket, Nov. 2017, pp. 175–179, doi: 10.1109/CYBERNETICSCOM.2017.8311705.

[23] A. Rusli, J. C. Young, and N. M. S. Iswari, "Identifying Fake News in Indonesian via Supervised Binary Text Classification," in *IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, Bali, Indonesia, 2020, pp. 86–90, doi: 10.1109/IAICT50021.2020.9172020.

[24] F. A. Ozbay and B. Alatas, "Fake news detection within online social media using supervised artificial intelligence algorithms," *Phys. Stat. Mech. Its Appl.*, vol. 540, p. 123174, Feb. 2020, doi: 10.1016/j.physa.2019.123174.

[25] C. Conforti, N. Collier, and M. Pilehvar, "Towards Automatic Fake News Detection: Cross-Level Stance Detection in News Articles," Mar. 2019, doi: 10.17863/CAM.37758.

[26] H. A. Santoso, E. H. Rachmawanto, A. Nugraha, A. A. Nugroho, D. Rosal Ignatius Moses Setiadi, and R. S. Basuki, "Hoax classification and sentiment analysis of Indonesian news using Naive Bayes optimization," *TELKOMNIKA Telecommun. Comput. Electron. Control*, vol. 18, no. 2, p. 799, Apr. 2020, doi: 10.12928/telkomnika.v18i2.14744.

[27] P. Assiroj, Meyliana, A. N. Hidayanto, H. Prabowo, and H. L. H. S. Warnars, "Hoax News Detection on Social Media: A Survey," in *2018 Indonesian Association for Pattern Recognition International Conference (INAPR)*, Jakarta, Indonesia,

Sep. 2018, pp. 186–191, doi: 10.1109/INAPR.2018.8627053.
[28] A. Bondielli and F. Marcelloni, "A survey on fake news and rumour detection techniques," *Inf. Sci.*, vol. 497, pp. 38–55, Sep. 2019, doi: 10.1016/j.ins.2019.05.035.
[29] N. A. Miftahul Huda and I. Sembiring, "The Use of Soft Systems Methodology to Resolve Hoax News Problems in Indonesia," in *2018 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE)*, Yogyakarta, Indonesia, Nov. 2018, pp. 65–68, doi: 10.1109/ICITISEE.2018.8720966.
[30] A. Pandhu Wijaya and H. Agus Santoso, "Improving The Accuracy of Naïve Bayes Algorithm for Hoax Classification Using Particle Swarm Optimization," in *2018 International Seminar on Application for Technology of Information and Communication*, Semarang, Sep. 2018, pp. 482–487, doi: 10.1109/ISEMANTIC.2018.8549700.
[31] S. Y. Yuliani, S. Sahib, M. F. Abdollah, Y. S. Wijaya, and N. H. M. Yusoff, "Hoax news validation using similarity algorithms," *J. Phys. Conf. Ser.*, vol. 1524, p. 012035, Apr. 2020, doi: 10.1088/1742-6596/1524/1/012035.
[32] B. Zaman, A. Justitia, K. N. Sani, and E. Purwanti, "An Indonesian Hoax News Detection System Using Reader Feedback and Naïve Bayes Algorithm," *Cybern. Inf. Technol.*, vol. 20, no. 1, pp. 82–94, Mar. 2020, doi: 10.2478/cait-2020-0006.
[33] S. M. Sirajudeen, N. F. A. Azmi, and A. I. Abubakar, "ONLINE FAKE NEWS DETECTION ALGORITHM," *J. Theor. Appl. Inf. Technol.*, vol. 95, no. 17, pp. 4114–4122, 2017.
[34] E. Zuliarso, M. T. Anwar, K. Hadiono, and I. Chasanah, "Detecting Hoaxes in Indonesian News Using TF/TDM and K Nearest Neighbor," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 835, p. 012036, May 2020, doi: 10.1088/1757-899X/835/1/012036.
[35] N. Ruchansky, S. Seo, and Y. Liu, "CSI: A Hybrid Deep Model for Fake News Detection," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, Singapore Singapore, Nov. 2017, pp. 797–806, doi: 10.1145/3132847.3132877.
[36] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, Mar. 2018, doi: 10.1126/science.aap9559.
[37] Y. Liu and Y. B. Wu, "Early Detection of Fake News on Social Media Through Propagation Path Classification with Recurrent and Convolutional Networks," in *AAAI*, 2018, pp. 354–361, [Online]. Available: https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16826.
[38] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein, "A Stylometric Inquiry into Hyperpartisan and Fake News," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Melbourne, Australia, Jul. 2018, pp. 231–240, doi: 10.18653/v1/P18-1022.
[39] O. Etzioni, M. Banko, S. Soderland, and D. S. Weld, "Open information extraction from the web," *Commun. ACM*, vol. 51, no. 12, p. 68, 2008, doi: 10.1145/1409360.1409378.
[40] A. Magdy and W. Nayer, "Web-based Statistical Fact Checking of Textual Documents," in *SMUC '10 Proceedings of the 2nd international workshop on Search and mining user-generated contents*, Toronto, ON, Canada, 2010, pp. 103–109, doi: 10.1145/1871985.1872002.
[41] A. L. Ginsca, A. Popescu, and M. Lupu, "Credibility in Information Retrieval," *Found. Trends® Inf. Retr.*, vol. 9, no. 5, pp. 355–475, 2015, doi: 10.1561/1500000046.
[42] Y. Wu, P. K. Agarwal, C. Li, J. Yang, and C. Yu, "Toward computational fact-checking," in *Proceedings of the VLDB Endowment*, 2014, pp. 589–600, doi: 10.14778/2732286.2732295.
[43] G. L. Ciampaglia, P. Shiralkar, L. M. Rocha, J. Bollen, F. Menczer, and A. Flammini, "Computational fact checking from knowledge networks," *PLoS ONE*, vol. 10, no. 10, 2015, doi: 10.1371/journal.pone.0141938.
[44] B. Shi and T. Weninger, "Fact Checking in Heterogeneous Information Networks," in *Proceedings of the 25th International Conference Companion on World Wide Web*, Republic and Canton of Geneva, Switzerland, 2016, pp. 101–102, doi: 10.1145/2872518.2889354.
[45] S. Heindorf, M. Potthast, B. Stein, and G. Engels, "Vandalism Detection in Wikidata,"

in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, Indianapolis Indiana USA, Oct. 2016, pp. 327–336, doi: 10.1145/2983323.2983740.

[46] Y. Long, Q. Lu, R. Xiang, M. Li, and C.-R. Huang, "Fake News Detection Through Multi-Perspective Speaker Profiles," in *Proceedings of the Eighth International Joint Conference on Natural Language Processing*, 2017, vol. Volume 2:, no. 8, pp. 252–256, [Online]. Available: http://www.aclweb.org/anthology/I17-2043.

[47] L. Derczynski, K. Bontcheva, M. Liakata, R. Procter, G. Wong Sak Hoi, and A. Zubiaga, "SemEval-2017 Task 8: RumourEval: Determining rumour veracity and support for rumours," in *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, Stroudsburg, PA, USA, 2017, pp. 69–76, doi: 10.18653/v1/S17-2006.

[48] D. Mocanu, L. Rossi, Q. Zhang, M. Karsai, and W. Quattrociocchi, "Collective attention in the age of (mis)information," *Comput. Hum. Behav.*, vol. 51, pp. 1198–1204, Oct. 2015, doi: 10.1016/j.chb.2015.01.024.

[49] D. Acemoglu, A. Ozdaglar, and A. ParandehGheibi, "Spread of (mis)information in social networks," *Games Econ. Behav.*, vol. 70, no. 2, pp. 194–227, Nov. 2010, doi: 10.1016/j.geb.2010.01.005.

[50] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang, "Prominent Features of Rumor Propagation in Online Social Media," in *2013 IEEE 13th International Conference on Data Mining*, Dallas, TX, USA, Dec. 2013, pp. 1103–1108, doi: 10.1109/ICDM.2013.61.

[51] J. Ma, W. Gao, and K.-F. Wong, "Detect Rumors in Microblog Posts Using Propagation Structure via Kernel Learning," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Stroudsburg, PA, USA, 2017, vol. 1, pp. 708–717, doi: 10.18653/v1/P17-1066.

[52] S. Volkova, K. Shaffer, J. Y. Jang, and N. Hodas, "Separating Facts from Fiction: Linguistic Models to Classify Suspicious and Trusted News Posts on Twitter," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Vancouver, Canada, 2017, pp. 647–653, doi: 10.18653/v1/P17-2102.

[53] P. Bourgonje, J. Moreno Schneider, and G. Rehm, "From Clickbait to Fake News Detection: An Approach based on Detecting the Stance of Headlines to Articles," in *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism*, Copenhagen, Denmark, 2017, pp. 84–89, doi: 10.18653/v1/W17-4215.

[54] W. Y. Wang, "'Liar, Liar Pants on Fire': A New Benchmark Dataset for Fake News Detection," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Stroudsburg, PA, USA, 2017, vol. 2, pp. 422–426, doi: 10.18653/v1/P17-2067.

[55] X. Zhou and R. Zafarani, "Fake news: A survey of research, detection methods, and opportunities," *ArXiv Prepr. ArXiv181200315v2*, 2020.

[56] S. Gupta, R. Thirukovalluru, M. Sinha, and S. Mannarswamy, "CIMTDetect: A Community Infused Matrix-Tensor Coupled Factorization Based Method for Fake News Detection," in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Barcelona, Aug. 2018, pp. 278–281, doi: 10.1109/ASONAM.2018.8508408.

[57] A. X. Zhang *et al.*, "A Structured Response to Misinformation: Defining and Annotating Credibility Indicators in News Articles," in *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18*, Lyon, France, 2018, pp. 603–612, doi: 10.1145/3184558.3188731.

[58] S. B. Parikh, V. Patil, R. Makawana, and P. K. Atrey, "Towards Impact Scoring of Fake News," in *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, San Jose, CA, USA, Mar. 2019, pp. 529–533, doi: 10.1109/MIPR.2019.00107.

[59] F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, "Fake News Detection on Social Media using Geometric Deep Learning," *ArXiv190206673 Cs Stat*, Feb. 2019, Accessed: Oct. 23, 2020. [Online]. Available: http://arxiv.org/abs/1902.06673.

[60] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, "Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking," in *Proceedings of*

the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 2017, pp. 2931–2937, doi: 10.18653/v1/D17-1317.

[61] K. Shu, S. Wang, and H. Liu, "Understanding User Profiles on Social Media for Fake News Detection," in *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, Miami, FL, Apr. 2018, pp. 430–435, doi: 10.1109/MIPR.2018.00092.

[62] K. Shu, S. Wang, and H. Liu, "Beyond News Contents: The Role of Social Context for Fake News Detection," in *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, Melbourne VIC Australia, Jan. 2019, pp. 312–320, doi: 10.1145/3289600.3290994.

[63] K. Shu, H. R. Bernard, and H. Liu, "Studying Fake News via Network Analysis: Detection and Mitigation," in *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*, N. Agarwal, N. Dokoohaki, and S. Tokdemir, Eds. Cham: Springer International Publishing, 2019, pp. 43–65.

[64] E. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, and L. de Alfaro, "Some Like it Hoax: Automated Fake News Detection in Social Networks," *ArXiv170407506 Cs*, Apr. 2017, Accessed: Nov. 04, 2020. [Online]. Available: http://arxiv.org/abs/1704.07506.

[65] G. Gravanis, A. Vakali, K. Diamantaras, and P. Karadais, "Behind the cues: A benchmarking study for fake news detection," *Expert Syst. Appl.*, vol. 128, pp. 201–213, Aug. 2019, doi: 10.1016/j.eswa.2019.03.036.

[66] H. Ahmed, I. Traore, and S. Saad, "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques," in *Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*, Cham, 2017, pp. 127–138, doi: 10.1007/978-3-319-69155-8_9.

[67] S. Castelo *et al.*, "A Topic-Agnostic Approach for Identifying Fake News Pages," in *Companion Proceedings of The 2019 World Wide Web Conference*, San Francisco USA, May 2019, pp. 975–980, doi: 10.1145/3308560.3316739.

[68] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic Detection of Fake News," in *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico, USA, Aug. 2018, pp. 3391–3401, [Online]. Available: https://www.aclweb.org/anthology/C18-1287.

[69] S. Zannettou, M. Sirivianos, J. Blackburn, and N. Kourtellis, "The Web of False Information: Rumors, Fake News, Hoaxes, Clickbait, and Various Other Shenanigans," *J. Data Inf. Qual.*, vol. 11, no. 3, pp. 1–37, Jul. 2019, doi: 10.1145/3309699.

[70] J. C. S. Reis, A. Correia, F. Murai, A. Veloso, F. Benevenuto, and E. Cambria, "Supervised Learning for Fake News Detection," *IEEE Intell. Syst.*, vol. 34, no. 2, pp. 76–81, Mar. 2019, doi: 10.1109/MIS.2019.2899143.

[71] S. Yang, K. Shu, S. Wang, R. Gu, F. Wu, and H. Liu, "Unsupervised Fake News Detection on Social Media: A Generative Approach," *Proc. AAAI Conf. Artif. Intell.*, vol. 33, pp. 5644–5651, Jul. 2019, doi: 10.1609/aaai.v33i01.33015644.

[72] G. Pasi, M. De Grandis, and M. Viviani, "Decision Making over Multiple Criteria to Assess News Credibility in Microblogging Sites," in *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Glasgow, United Kingdom, Jul. 2020, pp. 1–8, doi: 10.1109/FUZZ48607.2020.9177751.

[73] M. Viviani and G. Pasi, "Quantifier Guided Aggregation for the Veracity Assessment of Online Reviews: VERACITY ASSESSMENT OF ONLINE REVIEWS," *Int. J. Intell. Syst.*, vol. 32, no. 5, pp. 481–501, May 2017, doi: 10.1002/int.21844.

[74] G. Pasi and M. Viviani, "Application of Aggregation Operators to Assess the Credibility of User-Generated Content in Social Media," in *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Foundations*, Cham, 2018, pp. 342–353, doi: https://doi.org/10.1007/978-3-319-91473-2_30.

[75] A. Giachanou, P. Rosso, and F. Crestani, "Leveraging Emotional Signals for Credibility Detection," in *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Paris France, Jul. 2019, pp. 877–880, doi: 10.1145/3331184.3331285.

[76] K. Popat, S. Mukherjee, A. Yates, and G. Weikum, "DeClarE: Debunking Fake News and False Claims using Evidence-Aware

Deep Learning," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, 2018, pp. 22–32, doi: 10.18653/v1/D18-1003.

[77] M. A. Stefanone, M. Vollmer, and J. M. Covert, "In News We Trust?: Examining Credibility and Sharing Behaviors of Fake News," in *Proceedings of the 10th International Conference on Social Media and Society*, Toronto ON Canada, Jul. 2019, pp. 136–147, doi: 10.1145/3328529.3328554.

[78] A. M. Idrees, F. Kamal, and A. I., "A Proposed Model for Detecting Facebook News' Credibility," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 7, 2019, doi: 10.14569/IJACSA.2019.0100743.

[79] S. Yuliani, M. Faizal, S. Sahib, and Y. Supriadi, "A Framework for Hoax News Detection and Analyzer used Rule-based Methods," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 10, 2019, doi: 10.14569/IJACSA.2019.0101055.

[80] A. Roy, K. Basak, A. Ekbal, and P. Bhattacharyya, "A Deep Ensemble Framework for Fake News Detection and Classification," *ArXiv181104670 Cs*, Nov. 2018, Accessed: Dec. 07, 2020. [Online]. Available: http://arxiv.org/abs/1811.04670.

[81] Ghinadya and S. Suyanto, "Synonyms-Based Augmentation to Improve Fake News Detection using Bidirectional LSTM," in *2020 8th International Conference on Information and Communication Technology (ICoICT)*, Yogyakarta, Indonesia, Jun. 2020, pp. 1–5, doi: 10.1109/ICoICT49345.2020.9166230.

[82] S. Kumar, R. Asthana, S. Upadhyay, N. Upreti, and M. Akbar, "Fake news detection using deep learning models: A novel approach," *Trans. Emerg. Telecommun. Technol.*, vol. 31, no. 2, Feb. 2020, doi: 10.1002/ett.3767.

[83] T. Saikh, A. Anand, A. Ekbal, and P. Bhattacharyya, "A Novel Approach Towards Fake News Detection: Deep Learning Augmented with Textual Entailment Features," in *Natural Language Processing and Information Systems*, Cham, 2019, pp. 345–358.

[84] A. Thota, P. Tilak, S. Ahluwalia, and N. Lohia, "Fake news detection: a deep learning approach," *SMU Data Sci. Rev.*, vol. 1, no. 3, p. 10, 2018.

[85] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," *Trans. Assoc. Comput. Linguist.*, vol. 5, pp. 135–146, Dec. 2017, doi: 10.1162/tacl_a_00051.

[86] P. Symeonidis, L. Kirjackaja, and M. Zanker, "Session-aware news recommendations using random walks on time-evolving heterogeneous information networks," *User Model. User-Adapt. Interact.*, vol. 30, no. 4, pp. 727–755, Sep. 2020, doi: 10.1007/s11257-020-09261-9.

[87] A. Lommatzsch, B. Kille, and S. Albayrak, "Incorporating context and trends in news recommender systems," in *Proceedings of the International Conference on Web Intelligence*, Leipzig Germany, Aug. 2017, pp. 1062–1068, doi: 10.1145/3106426.3109433.

[88] M. Karimi, D. Jannach, and M. Jugovac, "News recommender systems – Survey and roads ahead," *Inf. Process. Manag.*, vol. 54, no. 6, pp. 1203–1227, Nov. 2018, doi: 10.1016/j.ipm.2018.04.008.

[89] J. Harambam, D. Bountouridis, M. Makhortykh, and J. van Hoboken, "Designing for the better by taking users into account: a qualitative evaluation of user control mechanisms in (news) recommender systems," in *Proceedings of the 13th ACM Conference on Recommender Systems*, Copenhagen Denmark, Sep. 2019, pp. 69–77, doi: 10.1145/3298689.3347014.

[90] G. de Souza Pereira Moreira, "CHAMELEON: a deep learning meta-architecture for news recommender systems," in *Proceedings of the 12th ACM Conference on Recommender Systems*, Vancouver British Columbia Canada, Sep. 2018, pp. 578–583, doi: 10.1145/3240323.3240331.

[91] A. Chakraborty, S. Ghosh, N. Ganguly, and K. P. Gummadi, "Optimizing the recency-relevance-diversity trade-offs in non-personalized news recommendations," *Inf. Retr. J.*, vol. 22, no. 5, pp. 447–475, Oct. 2019, doi: 10.1007/s10791-019-09351-2.

[92] T. Jiang, Q. Guo, Y. Xu, Y. Zhao, and S. Fu, "What Prompts Users to Click on News Headlines? A Clickstream Data Analysis of the Effects of News Recency and Popularity," in *Information in Contemporary Society*, vol. 11420, N. G. Taylor, C. Christian-Lamb, M. H. Martin,

and B. Nardi, Eds. Cham: Springer International Publishing, 2019, pp. 539–546.

[93] G. Sottocornola, P. Symeonidis, and M. Zanker, "Session-based News Recommendations," in *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18*, Lyon, France, 2018, pp. 1395–1399, doi: 10.1145/3184558.3191582.

[94] I. Mele, S. A. Bahrainian, and F. Crestani, "Event mining and timeliness analysis from heterogeneous news streams," *Inf. Process. Manag.*, vol. 56, no. 3, pp. 969–993, May 2019, doi: 10.1016/j.ipm.2019.02.003.

[95] S. A. Bahrainian, I. Mele, and F. Crestani, "Predicting Topics in Scholarly Papers," in *Advances in Information Retrieval*, vol. 10772, G. Pasi, B. Piwowarski, L. Azzopardi, and A. Hanbury, Eds. Cham: Springer International Publishing, 2018, pp. 16–28.

[96] I. Mele, S. A. Bahrainian, and F. Crestani, "Linking News across Multiple Streams for Timeliness Analysis," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, Singapore Singapore, Nov. 2017, pp. 767–776, doi: 10.1145/3132847.3132988.

[97] D. Chen, X. Zhang, H. Wang, and W. Zhang, "TEAN: Timeliness enhanced attention network for session-based recommendation," *Neurocomputing*, vol. 411, pp. 229–238, Oct. 2020, doi: 10.1016/j.neucom.2020.06.063.

[98] X. Luo, M. Zhou, S. Li, D. Wu, Z. Liu, and M. Shang, "Algorithms of Unconstrained Non-negative Latent Factor Analysis for Recommender Systems," *IEEE Trans. Big Data*, pp. 1–1, 2019, doi: 10.1109/TBDATA.2019.2916868.

[99] J. Li, P. Ren, Z. Chen, Z. Ren, T. Lian, and J. Ma, "Neural Attentive Session-based Recommendation," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, Singapore Singapore, Nov. 2017, pp. 1419–1428, doi: 10.1145/3132847.3132926.

[100] S. Wu, Y. Tang, Y. Zhu, L. Wang, X. Xie, and T. Tan, "Session-Based Recommendation with Graph Neural Networks," *Proc. AAAI Conf. Artif. Intell.*, vol. 33, pp. 346–353, Jul. 2019, doi: 10.1609/aaai.v33i01.3301346.

[101] W. Chen, F. Cai, H. Chen, and M. de Rijke, "A Dynamic Co-attention Network for Session-based Recommendation," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, Beijing China, Nov. 2019, pp. 1461–1470, doi: 10.1145/3357384.3357964.

[102] MAFINDO, *TurnBackHoax – Masyarakat Anti Fitnah Indonesia*. 2017.

[103] masdevid, "ID-Stopwords," *Github*, Jan. 02, 2016. https://github.com/masdevid/ID-Stopwords/blob/master/id.stopwords.02.01.2016.txt (accessed Aug. 16, 2018).

[104] N. Vikramaditya, "googlesearch-python 2020.0.2," *PYPI*, Jul. 06, 2020. https://pypi.org/project/googlesearch-python/ (accessed Aug. 18, 2020).

[105] Alexa Internet, Inc, "How are Alexa's traffic rankings determined?," *Alexa Rank*, 2019. https://support.alexa.com/hc/en-us/articles/200449744-How-are-Alexa-s-traffic-rankings-determined- (accessed Aug. 07, 2020).

[106] P.-N. Tan, M. Steinbach, A. Karpatne, and V. Kumar, *Introduction to data mining*, Second edition. NY NY: Pearson, 2019.

[107] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.