COMPARISON OF FAKE NEWS DETECTION USING MACHINE LEARNING AND DEEP LEARNING TECHNIQUES

Saeed Amer Alameri, Masnizah Mohd

Center for Cyber Security (CYBER),
Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor Malaysia.

Email: salameri@seiyunu.edu.ye, masnizah.mohd@ukm.edu.my

ABSTRACT

Fake news has spread widely on the Web in recent years due to the massive amount of information exchanged on the digital media. Intelligent, exciting, and different techniques applied to detect fake news as recently, Machine Learning (ML) techniques. Determining the best model for fake news detection using several datasets is better than using a single dataset. This study aims to determine the best-performing model among two ML models: Naïve Bayes (NB), Support Vector Machine (SVM), and three deep learning models: Long Short-Term Memory (LSTM), Neural Network with Keras (NN-Keras), and Neural Network with TensorFlow (NN-TF) for detecting fake news to curb the prevalence of fake news on the Web and social media platforms. The five ML models were examined using two different English language news datasets. The experiment was run five times after shuffling the news articles in the datasets, and the results recorded and then averaged. Each time the dataset instances are being shuffled. The performance of models was evaluated using four metrics, accuracy, precision, recall and F1-Score. The obtained results showed that deep learning models have achieved better accuracy than traditional ML models. The LSTM model has outperformed all other ML models examined. It achieved an average accuracy of 94.21%. The NN-Keras has also produced a good performance with an average accuracy of 92.99%. The order of the words carries important information and plays a significant role in the fake news classification, where our LSTM makes a prediction based on this.

Keyword: Fake news detection, Machine learning, Deep learning techniques, LSTM

I. INTRODUCTION

Nowadays, websites and social media have become the main source of information. Fake news is a type of information that are transferred and exchanged throughout these sites' visitors and users. The fake news simply defined as “a news article that is intentionally and verifiably false" (Bondielli & Marcelloni 2019). It contains false information; but, can be verified as false by several sources. Because fake news has a bad impact on different life aspects, fake news detection is a necessary process to curb the spread of fake news on the Web. Automatic fake news detection is now become a hot research area due to the large online information on social media websites, and due to the increasing number of users on these platforms.

Fake news is defined as intentionally and verifiably false news published by a news outlet (Zhou & Zafarani 2018). Nowadays, fake news has become widely spread and has bad effects on many aspects of life, such as political, economic, and education. It is typically generated for commercial interests to attract viewers and collect advertising revenue. According to the statistics, two million accounts across the world are closed in WhatsApp platform every month to limit the spread of fake news or misinformation (Kong et al. 2020).

During the coronavirus disease 2019 (COVID-19) pandemic, there was fake news on COVID-19 on social media has been countered. For example, the website https://sebenarnya.my/ has published around more than 160 fake news in Malaysia only in March 2020. Producing fake news about the COVID-19 is still continuous. So, identifying fake news in social networks is very important because they have an impact that can be tremendous due to the massive user numbers globally, which is further boosted by the extensive information sharing and propagation among these users.

In an attempt to reveal the fake news due to the increase of false posts or misinformation, several fact-checking websites have been deployed to expose or confirm stories: The most popular websites that expose fake news include snopes.com, factcheck.org, fakenewswatch.com, and politifact.com, etc. Locally, in Malaysia, there is sebenarnya.my by the Malaysian Communications and Multimedia Commission (MCMC). These websites play a crucial role in combating fake news, but they require expert analysis, which inhibits a timely response. As a response, numerous articles and blogs have been written to raise public awareness and provide tips on differentiating truth from falsehood. Nowadays, with the advent of digital technology and increasing the amount of information that people access and share every day, this information might be unverified or might be assumed as true. The term fake news appears in recent years to spread in some websites and social media platforms. Using machine learning techniques is become more popular in detecting and classifying fake news.

Curci et al. (2018) have compared five machine learning models to identify which model is the best to classify the fake news. They found that the best performing model was Long Short-Term Memory (LSTM) networks implementation. He used a dataset of news articles that are written in a formal English language.

Based on the literature review, Curci et al. (2018) and Abdullah-All-Tanvir et al. (2019), some studies have been conducted to determine which machine learning algorithms are the best to detect fake news. This study (Curci et al. 2018) has achieved good results (high accuracy) when applied some deep learning models such as LSTM by implementation on a dataset. However, still, these studies disapproved that these models can be generalized to different news from different sources. So, implementing these models on a different dataset (different topics of news or various resources of the news) is needed to be applied to approve its effectiveness in different datasets. Moreover, if these models (Naïve Bayes (NB), Support Vector Machine (SVM) and LSTM) achieved high accuracy, there is also a need to examine these models with fake news of varied languages. That means there is a need to apply the best model in a different language dataset. Different datasets mean different news datasets collected from different news sources. This includes thousands of news articles crawled from different websites where these news articles written by different authors, and each author or website has own writing style. Finally, collecting datasets for fake news detection is very import due to these datasets can be used again in future work. So, it is necessary to see how does the dataset affect on classification performance. In case if there were a bad impact, how to overcome this negative impact on the classification of fake news

To sum up, our problem is that the previous studies conducted to determine the best approaches for fake news detection were not implemented on different datasets. These studies claimed that some deep learning approaches such as (LSTM and Neural Network (NN)) outperformed some basic machine learning such as (Naïve Bayes and SVM) on detecting fake news but with the implementation of a single dataset, not several datasets. Re-implementation of these studies on several datasets is needed before making any further improvement or changes in those models that have outperformed. This paper aims to compare traditional machine learning models ((Naïve Bayes (NB) and Support Vector Machine (SVM)) with deep learning models (Long Short-Term Memory (LSTM), Neural Network with Keras (NN-Keras), and Neural Network with TensorFlow (NN-TF)) using two different datasets of news articles.

This paper consists of five (5) sections. Section I discuss the background of this study including the definition of fake news. Section II discuss the related works. Section III introduces the method used in the study. Section IV presents the results of the work and discussion. Lastly, section V concludes the paper with a summary of the findings and recommended future work.

II. RELATED WORK

Granik & Mesyura (2017) have implemented the NB approach on Facebook news posts. The classification accuracy achieved was around 74%. The dataset used was BuzzFeed News dataset which is collected from nine Facebook pages. The size of the dataset was 2282 posts. The dataset can be considered a small-sized dataset. NB used bag of words features in this study as a method of feature extraction. Most of the studies use NB as a baseline for their work because it is easy to implement, fast, and can outperform the more powerful alternatives, especially for small sample sizes.

A hybrid deep learning model of LSTM and CNN models was proposed by Ajao et al. (2018) to detect and classify fake news from Twitter posts. The dataset used in that study was consisted of around 5800 tweets. Using deep learning models enables automatic feature extraction. Embedding layer was used where the tweets cleaned and prepared such that each word is one-hot encoded. The proposed work using this deep learning approach achieves 80% accuracy. Using deep learning models gives better results than using traditional machine learning models due to the advantages the deep learning approaches have such as memory and the role of hidden layers play.

Five machine learning models: NB, SVM, LSTM, NN-TensorFlow and NN-Keras have been implemented using a corpus of labelled real and fake news articles to build a classifier that can make decisions about information based on the content of news articles from the corpus. The Doc2Vec embeddings has been used in all models applied by this study (Curci et al. 2018) to generate feature vector of each article; except LTSM model for which the text encoding and word embedding is used. The dataset used in this study (Curci et al. 2018) was crawled from Kaggle (Kaggle 2018). According to a quick look at the dataset, the majority of articles that this dataset contains were about political news. It has total news around 20K records (articles). 10K articles of it were real news, and 10K of it were fake news as well. A full training dataset has the following attributes: id, title, author, text, and label (1 for unreliable; and 0 for reliable).

The text classification approaches were used in that study (Curci et al. 2018). LSTM gave the best results and achieved high accuracy and F1-Score in comparison to all other models. NN-Keras has also achieved a high accuracy after LSTM. The reason that makes LSTM performs well is that the text is inherently a serialized object. And most models use the Doc2Vec to get their feature vectors and hence, they rely on the Doc2Vec to extract the order information and perform classification on it. Moreover, the LSTM model preserves the order of words using a different preprocessing method and makes a prediction based on both the words and their order.

Rathod et al. (2020) have examined four basic machine learning models: Random Forest, XGBoost, Naive Bayes, and Logistic Regression to detect or classify fake news articles. The dataset used in that study (Rathod et al. 2020) was the same Dataset 1 used in this project where this dataset was published on Kaggle. The real news articles were collected from various publications such as New York Times, Washington Post. The fake news articles were collected from 244 unreliable sources and then were identified and reported as fake by BS Detector. Each model was tested against the header and content of the articles. The standard Term Frequency-Inverted Document Frequency (TF-IDF) method and google word2vec were used as the vectorization technique in this study to represent data. The vectors obtained from Google’s Word2Vec word embeddings were inputs to train the models that have been examined in this study. The results of this study showed that using word2vec as vectorization to represent data was significantly improved accuracy for most of the models that were used in the study. One of the explanations of the increase in the accuracy is that word2vec adds meaning to the words. So, models seem to perform better.

Ahmed et al. (2017) have presented a model for fake news detection. Two different feature extraction techniques and six machine learning models have been examined, Stochastic Gradient Descent (SGD), Support Vector Machines (SVM), Linear Support Vector Machine (LSVM), K-Nearest Neighbour (KNN), Decision Trees (DT) and Logistic regression (LR). The dataset used was a new dataset collected by their team by compiling publicly available news article on Kaggole. They also tested their model on dataset of (Horne & Adali 2017), which is accessible to the public. The dataset used consists of 12,600 fake news, and 12,600 truthful articles. Each of the article's length is bigger than 200 characters. The results showed that the Linear SVM model achieved an accuracy of 92% and TF-IDF outperformed Term Frequency (TF) using two different datasets. Moreover, this study showed that when using unigram features, the highest accuracy has been achieved. In other words, with the increase of n-gram (Tri-gram, Four-gram), the accuracy of the algorithm decreases.

Kong et al. (2020) have been applied neural network models fed with two types of word vectors, N-gram vectorization and sequence vectorization. These models are trained on title news and content news to detect fake news. US English news dataset is collected and combined from two different Kaggle sources such that both of them are having similar attributes of news title, news content and news labels (“0” for real news and “1” for fake news). The dataset used was containing a total of 9,805 articles; fake and real. The study showed that training model with N-gram vectorization was performing better than training with sequence vectorization. Models trained with N-gram vectors on news content achieved the best accuracy and recall with 90.3% and 97.50% respectively. The study above (Kong et al. 2020) showed the NN-Keras produced better but does not address other deep learning models such as LSTM model which achieved better accuracy than NN-Keras model as in some studies reviewed above (Curci et al. 2018).

Deepak & Chitturi (2020) have proposed a method for fake news detection by combining two deep learning models, LSTM and Gated Recurrent Unit (GRU), with several word vector representations. The models also combined with online data mining, which works as additional features to the dataset. George Mclntires Fake News Dataset was chosen for the testing. It has 10,558 rows(entries) where it includes fake news and real news in 1:1 ratio. LSTM model achieved better performance with an accuracy of 94% when additional online data are combined with the model. As seen in this study, the suitable word vectorization for the LSTM model was word2vec, which used in similar studies (Curci et al. 2018; Rathod et al. 2020). Besides, the LSTM model showed better-performing in terms of precision, recall, and F1-Score for detecting fake news in this study.

Acharya et al. (2019) have compared three machine learning models, Support Vector Machine (SVM), Naïve Bayes (NB), and Logistic Regression, to classify fake news. Two datasets were used in their study, the ISOT dataset with a size of 31K news articles and the Liar Liar dataset with a size of 12.8K news articles. For the extraction of features, Term Frequency (TF) and Term Frequency-Inverted Document Frequency (TF-IDF) were used, where the features play a main role in the detection of fake news. In that study, the comparison results between the models showed that the Naïve Bayes had achieved better accuracy for the Term Frequency (TF) feature. Logistic Regression has also produced better accuracy with (TF-IDF). Additionally, due to the big size of the ISOT dataset comparing with the Liar Liar dataset, the ISOT dataset gave better results than the Liar Liar dataset.

Amine et al. (2019) have suggested a method to predict fake news by combining deep learning models with different metadata of news articles such as text, title, and author. It used word embedding technique with convolutional neural networks (CNNs). The study tested three models: CNN (Text only), CNNs (Text+Title), and CNNs (Text+Author). The last model, CNNs (Text+Author), achieved the best accuracy among the other models examined in this study with an accuracy of 96.00%. This study (Amine et al. 2019) implemented on a single dataset proposed by kaggle.com, which containing 20K news articles; 10K are fake articles, and 10K are real articles.

Based on the literature reviewed above, there are several datasets used to examined different machine learning and deep learning models. The common datasets used in these studies reviewed above are datasets that are crawled from Kaggle. These datasets are most contains full news articles. Six studies (Ahmed et al. 2017; Amine et al. 2019; Curci et al. 2018; Kong et al. 2020; Rathod et al. 2020) out the literature are used Kaggle datasets. Using full news articles is better than using short news articles such as tweets to train models. Using news content produced better accuracy than using news titles (Kong et al. 2020). The length of news content is longer than the news titles; thus, using full news articles can train models better.

The common methods used in the studies reviewed above are both machine learning and deep learning methods. NB and SVM is an example of machine learning models that widely used in studies. Recently, these methods are used as a baseline for the work. As reviewed, NB used in (Acharya et al. 2019; Curci et al. 2018; Granik & Mesyura 2017; Gupta et al. 2013; Rathod et al. 2020). In some studies, NB did not achieve better results, but it used at least as a baseline. Thus; it was recommended to used NB in our work. On the other hand, The common method used which achieved better results is the LSTM model and NN-Keras. In (Curci et al. 2018), LSTM model achieved the best accuracy, and NN-Keras models achieved competitive results. LSTMs can keep the ”Short Term Memories” for ”Longer” since there are different gates - input, output and forget to control the information flow. That makes LSTM is suitable for our work. In (Ajao et al. 2018) the LSTM model outperformed CNN model, but the dataset used was short texts with 5,800 tweets.

Additionally, most of the studies reviewed above have used different methods to extract features of the text news. The popular methods used to extract features from the text news articles which are used in studies reviewed above are, bag of words, one-hot encoding, TF, TF-IDF, word2vec and Doc2Vec for word embeddings. Among these features, it can be noted that using word2vec as vectorization technique has outperformed using standard TF-IDF method (Rathod et al. 2020) and TF-IDF has outperformed TF (Ahmed et al. 2017). So, in this work, only two features were used, word2vec and Doc2Vec, as previous studies showed that they are produced better results when used with machine learning and deep learning models. Due to word2vec adds meaning to words, it is highly recommended to use word2vec as vectorization technique to improve the accuracy of models.

III. METHOD

A. Experimental Design

The research design that has been followed in this project is provided in this section. By applying these steps, the study objectives have been achieved. The main phases of this research approach that have been taken are shown below:

- Phase 1: Preparing datasets
    - Collecting datasets.
    - Preprocessing datasets.
- Phase 2: Implementation.
- Phase 3: Evaluation and comparison results.

B. Dataset Description

i) Dataset 1

The first dataset (Dataset 1) was drawn from many different news sources. The sources are around 250 websites and blogs. These sources are companies' names that published the article news. For example, the real news was published by Washington Post, Reuters, Guardian and Cable News Network (CNN) news network, where the fake news was published in companies' websites such as 100percentfedup, 21st centurywire and activist post. It is available on the Kaggle website (Chan 2019). As the previous dataset, this dataset contains real and fake article news. The size of the dataset is around 27K articles: 12K of it are fake article news, and the other 15K are real article news. Each article has the following columns: Title: the title of the article, Content: the text of the article, Publication: which company published this article, and Label: real or fake.

ii) Dataset 2

The dataset 2 used in this project was collected by the Information Security and Object Technology (ISOT) research group, which is known as ISOT dataset (ISOT 2020). The dataset contains two types of articles, fake and real news. This dataset was collected from real-world sources; the truthful articles were obtained by crawling articles from reuters.com (News website). As for the fake news articles, they were collected from different sources. The fake news articles were collected from unreliable websites that were flagged by politifact.com (a fact-checking organization in the USA) and Wikipedia. The dataset contains different types of articles on various topics; however, the majority of articles focus on political and world news topics. In this project, only politics news type (for real) and politics type (for fake) were selected to be our dataset. The overall number of article news that has been selected was 18113 articles.

iii) Preprocessing Datasets

Before presenting our dataset to the source code, a little bit of preprocessing on the data have been done to train our models. The preprocesses include: Data cleaning, Inconsistent data representation, Missing data handling (removing missing values in the data), Removing the repeated instances (duplicated records) and removing the articles that are not crawled or represented in the file correctly. After preprocessing the datasets, we have obtained cleaned datasets which are now ready for processing. Table 3.2, and Table 3.3 give a summary of instances number in our datasets; dataset 1, and dataset 2; before and after; the preprocessing respectively.

Table **Error! No text of specified style in document.**.1 Dataset 1 before and after preprocessing

Dataset	Before preprocessing	After preprocessing
Real articles	15,712	15,607
Fake articles	12,999	12,738
Total	28,711	28,345

Table **Error! No text of specified style in document.**.2 Dataset 2 before and after preprocessing

Dataset	Before preprocessing	After preprocessing
Real articles	11,272	11,271
Fake articles	6,841	6,430
Total	18,113	17,701

C. Classification Models

Five models that have been selected. As this study is conducted to compare the results obtained using several datasets with the results obtained by the study of Curci et al. (2018), which used five classification models so, NB, SVM, NN-Keras, NN-TensorFlow, and LSTM have been used in this study. We implemented the classification using these models with the same parameters used in the study of Curci et al. (2018) to ensure there is no bias when we compare our results with the results obtained by Curci et al. (2018). The experiment of this study implemented five times for each model due to some models use random parameters.

i) Naïve Bayes (NB)

Naïve Bayes (NB) used in this project as it is a popular baseline model used in researches related to classification problems. The embeddings used for NB here are generated using the Doc2Vec model. The goal is to produce a vector representation of each article. After preprocessing the text news articles, a list of words has produced, which can be input into the Doc2Vec algorithm to produce a 300-length embedding vector for each article. The scikit-learn implementation of Gaussian Naive Bayes (GaussianNB) has been used in our project. GaussianNB is a built-in class inside scikit-learn machine learning library.

ii) Support Vector Machine (SVM)

SVM is one of the most used techniques in classification problems such as text classification (fake news detection). Using kernel trick in SVM makes it to give good results in the classification problems. In this paper, Doc2Vec is implemented to which generates vector representations for words for our SVM model. The Radial Basis Function kernel has been used. Two Doc2Vec feature vectors will be close to each other if their corresponding documents are similar, so the distance computed by this common kernel function should still represent the original distance.

iii) Neural Networks - Keras

A feed-forward neural network model has been implemented in this study using Keras. Nowadays, neural networks are used widely in Natural Language Processing (NLP) applications. To present the news article to our model, Doc2Vec has been used to generate a 300-length embedding vector for each news article. Three hidden layers have been used in this implementation. The first and second layer used 256 neutrons for each. The third layer used 80 neutrons. These layers are interspersed with dropout layers to avoid overfitting. The activation function used here is the Rectified Linear Unit (ReLU), which most deep networks use it. ReLU has been found to train faster and have less expensive operations when compared to sigmoid/ tanh. The value of the learning rate used in this model is (0.01), and training steps are (10000). This model has a fixed-size input $x \in R^{1 \times 300}$

iv) Neural Networks – TensorFlow

The neural network model has been implemented in this study using TensorFlow. The same implementation of the previous model (using Keras) has also been done here with simple differences. Doc2Vec has been used to generate a 300-length embedding vector for each news article to be an input for our model. Three hidden layers have been used in this implementation. All layers had 300 neutrons for each. The activation function used here is the Rectified Linear Unit (ReLU). The value of the learning rate used in this model is (0.001), and training steps are (20000). This model has a fixed-size input $x \in R^{1 \times 300}$

v) Long Short-Term Memory (LSTM)

Since the order of words is important in our model (LSTM), the Doc2Vec cannot be used or not suitable here because it will convert the entire document to one vector, and thus; the information of words order will be lost. So, word embedding will be used instead. After doing the preprocesses like removing special characters from text and preparing our text for processing, the frequency of each word in our dataset has counted. Then, the 5000 most common words have been found and given each one a unique integer ID. For example, the most common word had ID 0, and the second most common one had 1, and so on. The 5000 common words that have been chosen cover most of the text, as shown in Figure 3.6 and Figure 3.7. After that, each common word has been replaced with its assigned ID and deleted all uncommon words.

Since the LSTM model requires a fixed input vector length, the list longer than 500 numbers was truncated. For those lists shorter than 500 words, 0's have been put at the beginning of the list. The data with only a few words have been deleted since they do not carry enough information for training. By doing this, the original text string had been transferred to a fixed-length integer vector while preserving the words order information. Finally, word embedding has been used to transfer each word ID to a 32-dimensional vector. The word embedding will train each word vector based on word similarity. If two words frequently appear together in the text, they are thought to be more similar, and the distance of their corresponding vectors is small.

D. Evaluation Metrics

Four metrics have been selected, which are accuracy, precision, recall and F1-score to provide a comparison of performance between the models used in this study.

i) Accuracy

Accuracy measure is mostly used in machine learning to evaluate the performance for classification models. It tells how well the classifier correctly identifies an instance of the dataset, or as a percentage of the total number of predictions that are true. For binary classification problems, it can be calculated in terms of confusion as following:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

ii) Precision

Sometimes classification accuracy is not enough to measure the performance of the models. So, more performance measure is needed to evaluate our models. Precision is one of these metrics that can be used to measure the performance of the models. It is defined as the number of correctly classified positive examples divided by the number of examples labelled by the model (Sokolova & Lapalme 2009). The mathematical formula of precision using terms of confusion matrix is as:

$$Precision = \frac{TP}{TP + FP}$$

iii) Recall

Recall or sometimes known as sensitivity defines as the number of correctly classified positive examples divided by the number of positive examples in the data (Sokolova & Lapalme 2009). The mathematical formula of this measure using terms of confusion matrix is as:

$$Recall = \frac{TP}{TP + FN}$$

iv) F1-Score

F1-Score is one of the common evaluation metrics used in text classification problems. It is the harmonic mean between precision and recall. Thus; it balances between precision and recall. It tells you how precise your classifier is (how many instances it classifies correctly), as well as how robust it is (it does not miss a significant number of instances). Mathematically, it can be calculated in terms of confusion matrix by using the following equation.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

IV. RESULTS & DISCUSSION

The experiment was performed on the (Dataset 1) and (Dataset 2). The results of the implementation of five machine learning algorithms using these dataset were stated in fig……… These results are the average of five times of implementation as it is explained in the previous section.

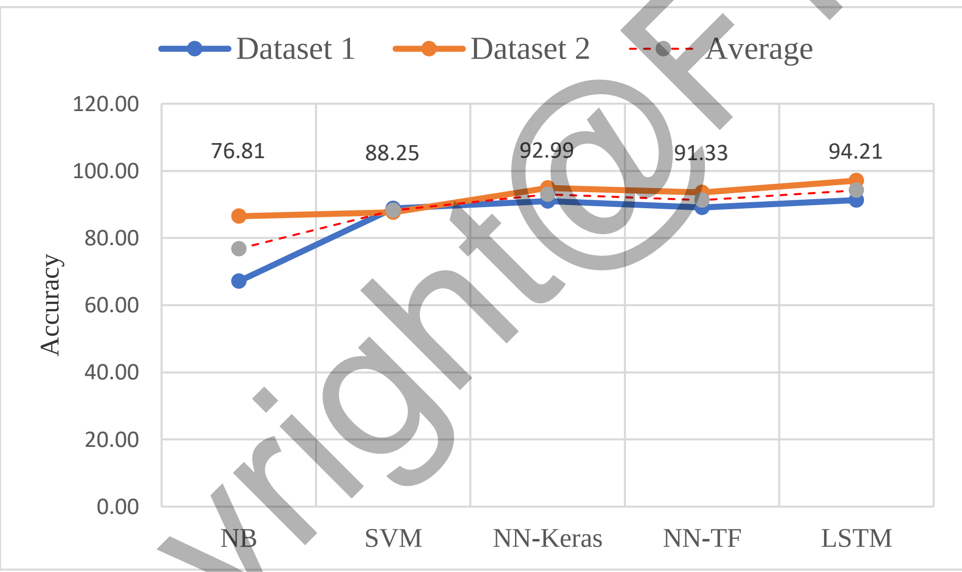

Figure 1 Average of accuracy for our models

Figure 1 shows the performance of the five models used on Dataset 1 and Dataset 2. It is worth noting that LSTM outperformed all other models in terms of accuracy, precision, recall and F1-Score. LSTM achieved an average accuracy of 94.21%, and NN-Keras has been showing good classification performance where it achieved an average accuracy of 91.33%. LSTM has achieved better results than other models due to it classifies the text depending on the words and its order. So, more features are extracted that LSTM classifies the text based on them. Considering the word order makes LSTM achieves higher accuracy than others.

In contrast, NB model produced the lowest accuracy among the models examined in this study using Dataset 1 and Dataset 2 because NB Cannot incorporate feature interactions and the predictors/features are independent. In general, the deep machine learning models, LSTM, NN-Keras and NN-TensorFlow have produced better classification performance than traditional machine learning, NB and SVM.

By comparing our results with the results obtained by Curci et al. (2018), it is noted that the LSTM model is the best in comparison to all other models, NB, SVM, NN-Keras, NN-TF for detecting fake news. As shown in Table 4.6, all studies; including this study; agreed that the LSTM model had been outperformed all other models as shown in Table 4.6. So, it is highly recommended for future studies to conduct more improvement on the LSTM model to detect fake news where it was examined by several studies and approved itself as the best classifier for fake news. This clearly states the most important finding of this study which is that the LSTM model is the best performing model out of the models, Naïve Bayes (NB), Support Vector Machine (SVM), Long Short-Term Memory (LSTM), Neural Network with Keras (NN-Keras), and Neural Network with TensorFlow (NN-TF) in fake news detection.

Table 3 Comparison of our results with previous researches

Study	Performance Measurements	Machine Learning Models				
		NB	SVM	NN-Keras	NN-TF	LSTM
Current study using Dataset 1	Accuracy	67.12	88.83	91.02	89.06	91.31
Current study using Dataset 2		86.50	87.68	94.96	93.60	97.11
Curci et al. (2018)		71.68	88.23	92.66	90.83	94.58
Current study using Dataset 1	Precision	68.68	88.67	90.94	89.04	91.43
Current study using Dataset 2		85.48	87.77	94.87	93.22	97.36
Curci et al. (2018)		73.31	88.70	92.66	90.82	94.58
Current study using Dataset 1	Recall	68.27	89.00	90.93	88.83	91.16
Current study using Dataset 2		85.17	85.47	94.25	92.92	96.44
Curci et al. (2018)		71.48	88.15	92.66	90.83	94.58
Current study using Dataset 1	F1-Score	67.07	88.76	90.93	88.91	91.19
Current study using Dataset 2		85.32	86.07	94.53	93.07	96.83
Curci et al. (2018)		71.05	88.17	92.65	90.82	94.58

All classification models except SVM model had achieved a higher accuracy when Dataset 2 was used, and they achieved a little bit lower accuracy when Dataset 1 was used comparing to Dataset 1 and also comparing to the dataset that has been used by Curci et al. (2018). One of the possible explanations of outperforming models when Dataset 2 was applied is that the nature of Dataset 2 has trained our models well. This means that the models are trained well when Dataset 2 has used. Dataset

2 is containing news articles from variety sources. Each source has its own writing style. So, the training process of models might not get enough data on training for each writing style. For example, Dataset 1 was drawn from different news sources which are more than 250 websites. In Dataset 2, the true news had been crawled from one source, which is reuters.com (website news). The number of articles for each source of the 250 websites in Dataset 1 definitely will not equal the number of articles that are obtained from once source (reuters.com) in Dataset 2. This explains that the news articles of reuters.com (real news article) which are in Dataset 2 have trained well the classification models because they are enough number of articles whereas the number of news articles for each source of the 250 which are in Dataset 1 is limited in training.

Besides, when comparing Curci et al. (2018) with results in Dataset 1, it is noted that the performance of models when using (Curci et al. 2018) dataset was higher than using Dataset 1 for all metrics used, Accuracy, precision, recall and F1-Score. This is because Dataset 1 has news articles from several sources which reach to more than 250 news sources. So, each news source in Dataset 1 may bring a few news articles and hence; the writing styles in Dataset 1 are varied. Therefore; the training models was not well due to the few data or a smaller number of news articles for each style. Although the instances of Dataset 1 are 27K, which are more compared to (Curci et al. 2018) dataset, but the Dataset 1 is a little bit skewed where it contains 12K fake articles and 15K real articles in Dataset 1. In (Curci et al. 2018) dataset, 10K fake article and 10K real articles are included. This might be contributed to make (Curci et al. 2018) dataset get better results than Dataset 1.

V. CONCLUSION

The fake news is become nowadays more spread on the Web. Detection fake news automatically is needed especially with the advent of technology and increasing of the amount of news data on social media platforms and spreading this fake news quickly. With using machine learning techniques which achieved good results in classification problems, fake news detection is one of these problems where a lot of researches applied machine learning techniques to solve this type of problems. This study aims to provide a comparison between five machine learning models to determine the best model for fake news detection. The classification models are NB, SVM, NN-Keras, NN-TensorFlow and LSTM. A preprocessing on our dataset has been conducted to prepare them for the processing. The implementation phase applied to classify news articles into fake or real using the datasets prepared previously. The evaluation of the models has been implemented at the end, followed by comparing the obtained results with the results of previous studies. The results obtained showed that deep learning models like LSTM, NN-Keras and NN-TensorFlow outperformed traditional machine learning models like NB and SVM. LSTM classifier provides the best results among the other deep learning. This finding agreed with the previous studies that showed the outperforming of LSTM model in the classification of fake news. In the future, it is recommended to re-apply this study with different datasets with different languages of news such as Malay news or Arabic news. In this case, the suitable tools for cleaning datasets such as removing stop words should be considered.

REFERENCE

Abdullah-All-Tanvir, Mahir, E.M., Akhter, S. & Huq, M.R. 2019. Detecting Fake News using Machine Learning and Deep Learning Algorithms. *2019 7th International Conference on Smart Computing and Communications, ICSCC 2019* 1–5.

Acharya, A., Jathan, A., Anchan, D. & Kottary, K.D. 2019. Fake News Detection Using Machine Learning

Ahmed, H., Traore, I. & Saad, S. 2017. Detection of online fake news using n-gram analysis and machine learning techniques. *International conference on intelligent, secure, and dependable systems in distributed and cloud environments* hlm. 127–138. Springer.

Ajao, O., Bhowmik, D. & Zargari, S. 2018. Fake News Identification on Twitter with Hybrid CNN and RNN Models. *International Conference on Social Media & Society* hlm. 226–230. Copenhagen, Denmark.

Amine, B.M., Drif, A. & Giordano, S. 2019. Merging deep learning model for fake news detection. *2019 International Conference on Advanced Electrical Engineering (ICAEE)* hlm. 1–4. IEEE.

Bondielli, A. & Marcelloni, F. 2019. A survey on fake news and rumour detection techniques. *Information Sciences 497*: 38–55.

Chan, A. 2019. Fake News Classifier - Final Project https://www.kaggle.com/anthonyc1/fake-news-classifier-final-project/notebook?scriptVersionId=6878570 [7 March 2020].

Curci, J., Khara, K., Pillai, A. & Qin, R. 2018. Fake News Detection. Boston University.

Deepak, S. & Chitturi, B. 2020. Deep neural approach to Fake-News identification. *Procedia Computer Science 167*: 2236–2243.

Granik, M. & Mesyura, V. 2017. Fake News Detection Using Naive Bayes Classifier. *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)* hlm. 900–903

Kaggle. 2018. Fake News https://www.kaggle.com/c/fake-news/data [7 March 2020].

Kong, S.H., Tan, L.M., Gan, K.H. & Samsudin, N.H. 2020. Fake News Detection using Deep Learning. *2020 IEEE 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE)* hlm. 102–107. IEEE.

Rathod, D., Salvi, H., Mehta, M. & Vegesna, A. 2020. Comparison of Techniques to Detect Fake News. *International Journal for Research in Applied Science & Engineering Technology (IJRASET) 8*(IV)

Sokolova, M. & Lapalme, G. 2009. A systematic analysis of performance measures for classification tasks. *Information Processing and Management 45*(4): 427–437. http://dx.doi.org/10.1016/j.ipm.2009.03.002.

Zhou, X. & Zafarani, R. 2018. Fake News: A Survey of Research, Detection Methods, and Opportunities