Fake News Detection using Deep Learning

Sheng How Kong
School of Computer Sciences
Universiti Sains Malaysia
Pulau Pinang, Malaysia
sheng-how.kong@student.usm.my

Li Mei Tan
School of Computer Sciences
Universiti Sains Malaysia
Pulau Pinang, Malaysia
limeitan@student.usm.my

Keng Hoon Gan
School of Computer Sciences
Universiti Sains Malaysia
Pulau Pinang, Malaysia
khgan@usm.my

Nur Hana Samsudin
School of Computer Sciences
Universiti Sains Malaysia
Pulau Pinang, Malaysia
nurhana.samsudin@usm.my

***Abstract*—FAKE news has proliferated to a big crowd than before in this digital era, the main factor derives from the rise of social media and direct messaging platform. Techniques of fake news stories detection ingenious, varied, and exciting. This study aims to apply natural language processing (NLP) techniques for text analytics and train deep learning models for detecting fake news based on news title or news content. Solution proposed in this study aims to be applied in real-world social media and eliminate the bad experience for user to receive misleading stories that come from non-reputable source. For NLP techniques, text pre-processing such as regular expression, tokenization, lemmatization and stop words removal are used before vectorizing them into N-gram vectors or sequence vectors using terms frequency inverse document frequency (TF-IDF) or one-hot encoding respectively. Then, TensorFlow is chosen as the framework to be used with built in Keras deep learning libraries that is having a large community and number of commits on Tensorflow GitHub repository that can be enough to build deep learning neural network models. Results from the models are showing that models trained with news content can achieve better performance with computation time being sacrificed while models trained with news title require less computation time to achieve good performance. Also, overall performance of models fed with N-gram vectors are slightly better than models fed with sequence vectors.**

Keywords—Fake news, natural language processing, deep learning, neural network, TensorFlow, Keras

I. INTRODUCTION

Fake news is one of the biggest scourges in digital connected world. It is defined as a subject including news, data, report and information that wholly or partly false. The impact of fake news from personnel until society is huge and no longer limit to conflict. It is a wildfire and will influence many people in every day. Fakes news created a threaten to country's security, economy, prosperity and individual. People might do not aware how fake news could impact to matter surrounding them on how to handle when it happens. Billions of articles created every day on the web, people might be the helping hand by spreading this news without knowing this news is real or fake. A simple action has become a serious issue if there is no control gate to prevent fake news stories being spread aggressively.

In Malaysia, there is a party Anti-Fake News Act (AFNA) introduce by government indicate that fake news stories can be interpret in the pattern of features, audio, visual or any from social media applications capable of suggesting ideas or words [1]. There's a lot of social media application available in the market, "Facebook", "WhatsApp", YouTube, etc. Let's take WhatsApp as an example, referring to the statistics computed, there was a total of 2 million of accounts being closed in every month by prevent the spread of fakes news [2]. In addition, it not just in touch with friends and family but also part of politics. In 2018, Brazilian has poison by the fake news via WhatsApp. The reason was because total of 44% of Brazilian voters use WhatsApp to know about their country political and electronic information [3]. WhatsApp has spent many efforts to develop an automated process to remove millions of fake news. From these millions of deleted accounts, it is difficult to know how many misclassified as fake news.

II. PROBLEM STATEMENT

Social media provides easier and accessible to connect and communicate with people, how we interact the quality of interpersonal relationships. However, the quality of interpersonal is at risk. Internet is interwoven into daily life, people has made social media applications indispensable to the life. Variety usage of social media continue increases with advance technology that made the world a domestics world where people connect to anyone freely from diverse place. Interaction with strangers across worldwide has make it possible and opportunity for hacker to access individual's vital information and cybercrime happen. As a result, social media has a huge impact that could affect interpersonal relationship when people depend strongly and allow social media control the communication.

Spreading of fake news and misleading information can eventually cause confusions and rumors circulating around and the victims could be badly impacted, which one of the worst impacts is committing suicide. Existing system on the topics of deep learning for deception fake news detection has been focusing online review and publicly of the posting by social media. It is difficult to detect fake news because it can be existing in variety of pattern and there's a huge leap in NLP frameworks.

III. PROPOSED SOLUTION

To avoid misinformation and fake news spreading around, fake news detection is much needed, and this project proposes 4 similar neural network models are to be trained and each model is fed with different text vectors of news title and news content so to be compared on the model performance. The 4 similar neural network models to be trained are as follows:

1. Model 1: Fed with N-gram vectors of news title
2. Model 2: Fed with N-gram vectors of news content
3. Model 3: Fed with sequence vectors of news title
4. Model 4: Fed with sequence vectors of news content

To develop the proposed solution, machine learning workflow suggested by Google developers is used because it is shown and explained with working example in spam mail classification that also involves text analytics and classification [4]. The workflow begins with dataset collection and then followed by data exploration, data cleaning and preparation, deep learning model training, model evaluation and finally model deployment as shown in Fig 1 [4].

Fig. 1. Machine learning workflow [4].

A. *Data Collection and Cleaning*

News dataset can be collected from many different sources such as Kaggle, UCI Machine Learning and more. However, the focus of this project is to train the deep learning model classifying real or fake news based on the news title or news content without knowing where the sources of the news coming from. As such, the model trained may not be over-dependent on news sources and this can help in generalization of the model [5].

In this project, English language news is chosen mainly because there are more open-source news datasets that can be obtained from the internet and there are also more text analytics and classification works have been done on English language. Also, English news is popular in Malaysia and therefore, driving the motivation of this work. Due to time constraints and the amount of efforts required to collect Malaysia English News, US English news dataset is collected and combined from 2 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) while other fields such as the URL sources and News ID are removed in the dataset cleaning stage [6] [7].

At the data cleaning stage, besides removing the ID and URL columns, all the empty, repetitive and problematic rows are also removed from the dataset. This is because there are only less than 3% of the total rows in the combined dataset are having empty or repetitive values in any of the columns. The problematic rows are also removed because some of the news title or content are having newlines ("\n") or tabs ("\t") that makes the texts in the (.csv) file go into new row or column, causing mess to the dataset.

B. *Data Exploration*

After data collection, the data exploration is serving the purpose of understanding the characteristics and finding the patterns of the data [8]. Also, exploring data can also help to minimize the risk of having extremely imbalanced data which can greatly affect model trained in the later stage [9]. With the dataset being collected and cleaned, data exploration is performed to show the distribution or ratio of fake new and real news, word counts or even plotting a word cloud to show the most frequently occurring words.

C. *Data Preparation*

Data preparation is a stage where the data is being prepared and transformed into a context which the machine can understand and therefore, feeding into the machine learning model to be trained [10]. Regular expression is a fundamental tool by specifying text search string and excellent at pattern matching in natural language processing [11]. It performs a basic activity including word segmentation and normalization, sentence segmentation, and stemming. Word tokenization is cascade of a simple regular expression substitution [11]. Stop words is the word that commonly apply into the sentence and will be remove when building the index. For example, "the", "a", "on", "are", "around", and there is a long list that could easily search from website.

In this project, data preparation involves the preprocessing of the news titles and contents by using text processing methods such as regular expression to remove punctuations and special characters, tokenization to split the text into words, lemmatization to transform the words back to its root word and stop words removal to remove the common and meaningless words [12]. After obtaining the keywords, vectorization is to be performed to convert the keywords into numerical vectors which is understandable by machines [13]. In this project, 4 similar deep learning models will be trained with N-gram vectorization and sequence vectorization based on news title and news content as both these methods could be applied in different applications of social media.

N-gram model is contiguous sequence of **n** items from a sample of text and widely used tool in language processing. It estimates from preceding words, compute the probability by counting in corpus and normalizing to a sentence/ any other sequence of words [11]. TF-IDF is a short for term frequency in information retrieval [4]. It is a statistic that reflect the important of a word in the document. It helps to increase the proportional where a word appears in number of times and adjust for the fact that some word that exists consistently in general [4]. As machine can only recognize numbers, one-hot encoding is also another way to convert words data to vectors of number (e.g. binary vectors, "0" and "1") [14].

In this project, unigram (N=1), bi-gram (N=2) and tri-gram (N=3) vectorization is used with term frequency inverse document frequency (TF-IDF) encoding and output with numerical vectors that provide the statistics of each unigram, bi-gram or trigram [13]. To perform N-gram vectorization, python TFIDFVectorizer methods from NLTK library is used to transform the news titles and contents into N-gram and compute the TF-IDF and output a sparse matrix of n x m where n refers to the number of rows and m refers to the size of vocabulary.

Unlike N-gram vectorization, sequence vectorization will turn the whole sequence of words into numerical vectors based on the token indexes [13]. Then, one-hot encoding is used to encode the sequence vectors into matrices of 0 and 1 where the matrices will have n-dimensional spaces which the "n" refers to the size of vocabulary [13]. To perform sequence vectorization with one-hot encoding, Keras text tokenizers, text to sequence and sequence to matrix methods are used and the final output will be the words from the news title or content in in the form of numerical vectors format.

D. Model Training

Based on some relevant works and researches, it is found that many of the commonly known machine learning algorithms such as Decision Tree and BayesNet are not performing so well as expected and not being able to achieve at least 90% of accuracy and recall rate [15] [16]. Also, referring to the work by Poddar, Amali and Umadevi, the performance of their neural network model can only achieve 49.9% accuracy which is relatively low when compared to other machine learning algorithms [16]. Therefore, this project work targets not only to accurately detect fake news but also make improvement on neural network model.

With all the pre-processed new titles and content in vectors form, Keras neural network models with some dense layers are built and trained using Tensorflow framework to perform classification task of detecting the fake news. In the Keras neural network model, the layers 1, 3 and 5 are using rectified linear unit (RELU) as the activation function with 64 nodes, 16 nodes and 2 nodes respectively. In between these 3 layers, 2 dropout layers of 20% dropout rate are also added to the network. The purpose of adding dropout layers is used to avoid from overfitting by dropping some node and therefore, generalize better [17].

Then, the last layer of the Keras neural network model, also known as output layer, is added with sigmoid activation function so to ensure the output of the network is in binary format because there are only 2 possible outcomes from this project such that "0" refers to real news and "1" refers to fake news. Besides that, early stopping is also implemented in the models so to decrease overfitting for better generalization [18]. 80% of the news dataset is used as training set and the remaining 20% are used as test set. Also, 20% of the training set will be used as validation set so to validate the model before proceeding to test or evaluate the model with unseen data.

E. Model Evaluation

After training the Keras neural network models, it will be tested on the unseen test dataset. The metrics that will be used to evaluate these models are accuracy, recall and computational time. The reason of using accuracy metric in this project is to determine how accurate the models can be used to classify real or fake news. With higher accuracy, the model will most likely to be able to classify the real and fake news correctly.

However, in some cases that accuracy may not be as high as desired, but the model can still be used as it can predict the fake news correctly which is our key purpose of project. Therefore, recall metric is also used to determine how likely the model can predict the fake news correctly and with higher recall, it can show that the model is more likely to classify the fake news correctly. In this project, recall is more important because when the recall rate is high and actual fake news are correctly detected, the spreading of fake news can be controlled in social media.

F. Model Deployment

As this project is meant to be deployed in actual real-world social media, therefore model deployment can be ignored in this project. However, it is somehow good to suggest this solution to social network companies so to control the spreading of fake news around the social media which can bring about negative consequences.

IV. Results and Discussion

In this section, outcomes for each stage of the machine learning workflow are shown except for model deployment which has no outcome to be shown as the model is not deployed. Besides, the output performance of 4 Keras neural network models trained will also be shown.

A. Data Collection and Cleaning

After the 2 datasets have been collected, combined, and cleaned, the final news dataset will then have only 3 attributes, which are title, content and label representing news title, news content and news label respectively as shown in Fig 2.

	Title	Content	Label
0	Why These Democrats Flipped To Defeat Wall Str...	WASHINGTON -- Anyone who wants to understand t...	0.0
1	Teen suspect in Arizona woman's slaying in cus...	A 14-year-old boy sought in the shooting death...	1.0
2	Retiring TSMC founder predicts fewer doctors, ...	Taiwan Semiconductor Manufacturing Co (TSMC) c...	1.0
3	Trump Threatens WW3 "Reviewing Plans To Send W...	By Aaron Kesel\nPresident Donald Trump is "act...	0.0
4	Box Office: 'Blade Runner 2049' fades to $31.5...	Harrison Ford, Ryan Gosling, Ana de Armas and ...	1.0

Fig. 2. 5 sample rows of collected, combined and cleaned news dataset.

B. Data Exploration

In data exploration stage, the dataset label ratio is first to be examined so to know whether the dataset is imbalanced which could possibly affect the outcome of the model training in the later stage. By getting the total count for each label and visualize using a bar chart in Fig 3, the dataset can be considered as a balanced dataset such that the ratio of real news to fake news is nearly 50:50.

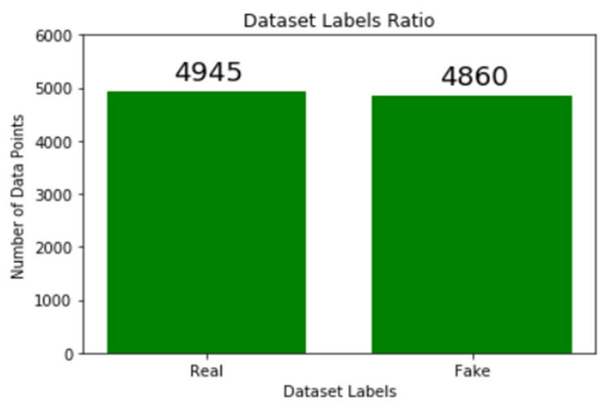

Fig. 3. Dataset Labels Ratio.

Then, word counts in ratio format are plotted for each news label to determine if there are any significant words that contribute to the real or fake news used. Fig 4 shows the word cloud and the top 20 words for real news respectively while Fig 5 shows the word cloud and top 20 words for fake news respectively.

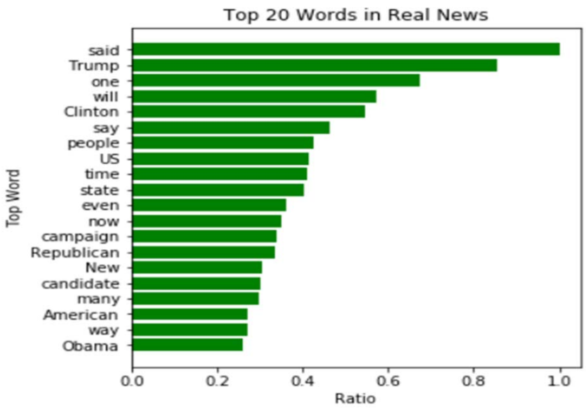

Fig. 4. Word count for real news in ratio form.

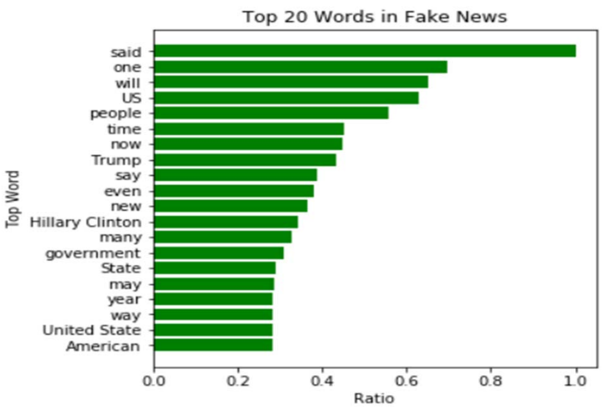

Fig. 5. Word count for fake news in ratio form.

As could be observed from both the word cloud and top occurring words, many words like “said”, “Trump”, “one” and more are appearing frequently in both real and fake news. This indicates that all these words or terms are common in many news regardless of the labels and can possibly be ignored or given a smaller weight when training the model.

C. Data Preprocessing

In this stage, news title and news content will be going through 2 pre-processes before feeding to the neural network model. These pre-processes are text preprocessing and vectorizations. In text pre-processing, regular expression, tokenization, lemmatization and stop words removal are done to every rows of news title and news content to obtain the root keywords. For example, Fig 6 shows one of the news titles from the dataset before and after text pre-processing.

```
In [31]: title[4]
Out[31]: "Box Office: 'Blade Runner 2049' fades to $31.5 million opening weekend"

In [32]: cleaned_title[4]
Out[32]: 'box office blade runner 2049 fade 315 million opening weekend'
```

Fig. 6. An example of news title before and after text pre-processing.

Then, after text-preprocessing, the news title and content are converted into numerical vectors of N-gram vectors and sequence vectors and then combined into a single matrix respectively. An example of matrix outcome with the word vectors combined is shown in Fig 7.

```
X_tfidf_train_title

array([[0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       ...,
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.]])
```

Fig. 7. An example outcome of matrix with the word vectors combined.

D. Model Training

4 similar neural network models are trained but fed with different word vectors as mentioned in Section III where

1. Model 1: Fed with N-gram vectors of news title
2. Model 2: Fed with N-gram vectors of news content
3. Model 3: Fed with sequence vectors of news title
4. Model 4: Fed with sequence vectors of news content

One example output from the model during the training stage is as shown in Fig 8. With accuracy and recall metric specified, the training and validation accuracy and recall are also shown in Fig 8 as well.

```
Train on 6275 samples, validate on 1569 samples
Epoch 1/10
6275/6275 [==============================] - 1s 186us/sample - loss: 0.6913 - accuracy: 0.5461 - Recall:
0.5202 - val_loss: 0.6864 - val_accuracy: 0.5809 - val_Recall: 0.8674
Epoch 2/10
6275/6275 [==============================] - 1s 97us/sample - loss: 0.6744 - accuracy: 0.6316 - Recall: 0.
9331 - val_loss: 0.6673 - val_accuracy: 0.6150 - val_Recall: 0.9681
Epoch 3/10
6275/6275 [==============================] - 1s 97us/sample - loss: 0.6391 - accuracy: 0.6993 - Recall: 0.
9791 - val_loss: 0.6400 - val_accuracy: 0.6670 - val_Recall: 0.9598
Epoch 4/10
6275/6275 [==============================] - 1s 99us/sample - loss: 0.5905 - accuracy: 0.7615 - Recall: 0.
9847 - val_loss: 0.6154 - val_accuracy: 0.6899 - val_Recall: 0.9465
Epoch 5/10
6275/6275 [==============================] - 1s 98us/sample - loss: 0.5414 - accuracy: 0.8163 - Recall: 0.
9857 - val_loss: 0.5989 - val_accuracy: 0.7135 - val_Recall: 0.9395
Epoch 6/10
6275/6275 [==============================] - 1s 92us/sample - loss: 0.4980 - accuracy: 0.8655 - Recall: 0.
9874 - val_loss: 0.5910 - val_accuracy: 0.7326 - val_Recall: 0.9299
Epoch 7/10
6275/6275 [==============================] - 1s 92us/sample - loss: 0.4617 - accuracy: 0.8958 - Recall: 0.
9874 - val_loss: 0.5913 - val_accuracy: 0.7409 - val_Recall: 0.9229
Epoch 8/10
6275/6275 [==============================] - 1s 95us/sample - loss: 0.4361 - accuracy: 0.9185 - Recall: 0.
9904 - val_loss: 0.5975 - val_accuracy: 0.7467 - val_Recall: 0.9127
```

Fig. 8. An example output of Keras neural network model trained.

E. Model Evaluation

The performance comparisons between all 4 Keras neural network models are shown in Fig 9, Fig 10, and Fig 11 for accuracy, recall and computational time respectively. Table I shows the overall performances for all 4 models.

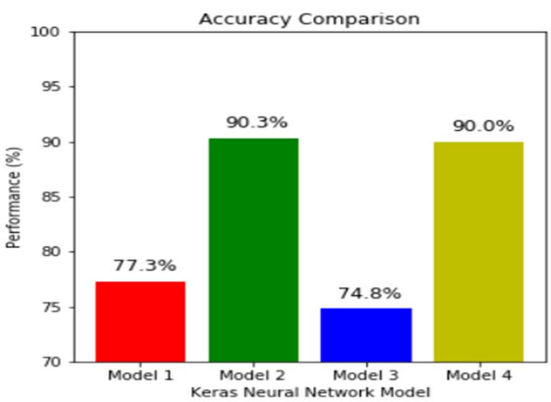

Fig. 9. Accuracy of all 4 Keras models.

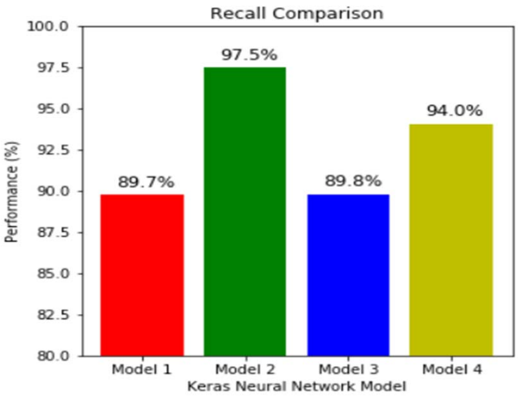

Fig. 10. Recall of all 4 Keras models.

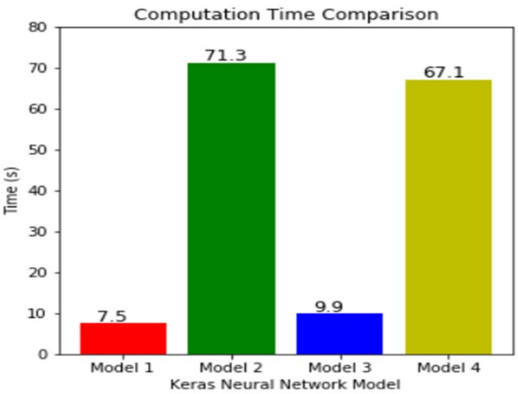

Fig. 11. Time required to train the each Keras neural network model.

TABLE I. PERFORMANCES OF 4 KERAS NEURAL NETWORK MODELS TRAINED.

Model	**Performance Metric**		
	Accuracy (%)	***Recall (%)***	***Computation time (s)***
1	77.3	89.7	7.5
2	90.3	97.5	71.3
3	74.8	89.8	9.9
4	90.0	94.0	67.1

As observed from the Fig 9 and Fig 10, models trained with news content (Model 2 and Model 4) are performing better than the models trained with news title (Model 1 and Model 3). This is because of the larger size of vocabulary in news content, leading to more detailed keywords when compared to news title which the keywords are more general overall. Model 2 and model 4 are having nearly similar accuracy but model 2 which is trained with news content using N-gram vectorization has higher recall rate of 97.5% such that it can classify the actual fake news correctly with lower mistakes.

However, even though the models trained with news content could possibly achieve such a better performance, but it also comes with a downside that it requires higher computational time due to the greater size of vocabulary as shown in Fig 11 that model 2 and model 4 requires 71.3 seconds and 67.1 seconds to complete computation respectively. On the other hand, model 1 and model 3 that are having lower accuracy of 77.3% and 74.8%, can achieve considerably high recall rate with 89.7% and 89.8% respectively.

Besides, models trained with news title require relatively less computation time when compared to models trained with news content. This is also key reason why training models with news title instead of only training with news content which can yield higher accuracy and recall rate.

In social media applications such as WhatsApp, WeChat and more which can be used for chatting and social networking, users tend to response faster to incoming messages and fake news can be spread faster if anyone intended to do so. In this case, models with low computation time and high recall rate are much required. Therefore, models trained with news title can eventually come in place to detect the any fake news generated. Despite of longer computation time, models trained with news content can also be used in social media applications such as Facebook and Twitter where the fake news is coming from feeds. The fake news may not need to be detected as soon as it is published because users' feeds may not be updated so soon. But with higher accuracy and recall model which can accurately detect the fake news, the feeds can be removed as soon as the model detects the fake news.

V. CONCLUSION

In conclusion, the best neural network model trained in this project work can achieve up to 90.3% accuracy with 97.5% recall that can accurately detect fake news with very low mistakes. Neural network models trained with N-gram vectors are performing slightly better than models trained with sequence vectors mainly because of N-gram vectors using TF-IDF which will not rely only on term frequency, but also a weight score that emphasize on more important terms. Models trained with news title are suitable to be used in social media applications that users would response fast on any updates or incoming messages due to its fast computation time and high recall rate (low mistake rate). With fast computation, any message that sends out fake news can be stopped. On the other hand, for social media applications are having feeds updated from time to time, fast computation time is not very crucial and therefore, models trained with news content that would have higher accuracy and recall will be of better choice to accurately detect the fake news and stop users from spreading it.

For future improvement, the Keras neural network models can be further improved by tuning the parameters to achieve even higher accuracy and recall. Recurrent Neural Network (RNN) with long short-term memory algorithm (LSTM) can also be used to further enhance fake news detection performance with NLP. Besides, further research can also be done on the images, videos, texts on images of the news to further improve the models in future. Besides that, to further implement this solution in Malaysia, similar approaches or techniques can be used to train the models with news dataset collected in Malaysia. Further research and experiment can also be done in Malay and Chinese news.

Acknowledgment

This research is supported by Fundamental Research Grant Scheme (FRGS), Ministry of Education Malaysia (MOE) under the project code, FRGS/1/2018/ICT02/USM/02/9 (PKOMP6711634) with title "Automated Big Data Annotation for Training Semi-Supervised Deep Learning Model in Sentiment Classification".

References

[1] N. H. Hassan, "New Straits Times," New Straits Times, 21 February 2019. [Online]. Available: https://www.nst.com.my/opinion/columnists/2019/02/462486/anti-fake-news-act-irrelevant. [Accessed 24 November 2019].

[2] J. Vasandani, "Towards Data Science," Medium, 9 April 2019. [Online]. Available: https://towardsdatascience.com/i-built-a-fake-news-detector-using-natural-language-processing-and-classification-models-da180338860e. [Accessed 24 November 2019].

[3] M. B. a. M. O. Ms. Tardáguila, "The New Work Times," The New Work Times, 17 October 2018. [Online]. Available: https://www.nytimes.com/2018/10/17/opinion/brazil-election-fake-news-whatsapp.html. [Accessed 24 November 2019].

[4] "Machine Learning," Google Inc, [Online]. Available: https://developers.google.com/machine-learning/guides/text-classification. [Accessed 22 November 2019].

[5] F. T. Asr and M. Taboada, "Big Data and quality data for fake news and misinformation detection," 23 May 2019. [Online]. Available: https://doi.org/10.1177/2053951719843310. [Accessed 22 November 2019].

[6] Jruvika, "Kaggle," 8 December 2017. [Online]. Available: https://www.kaggle.com/jruvika/fake-news-detection. [Accessed 15 November 2019].

[7] Nop.AI, "Kaggle," 1 November 2019. [Online]. Available: https://www.kaggle.com/nopdev/real-and-fake-news-dataset. [Accessed 15 November 2019].

[8] P. V. F. and D. H. C., "Exploratory data analysis," America Physcology Association, 2012.

[9] W. Badr, "TowardsDataScience," 22 February 2019. [Online]. Available: https://towardsdatascience.com/having-an-imbalanced-dataset-here-is-how-you-can-solve-it-1640568947eb. [Accessed 21 November 2019].

[10] S. Singhal and M. Jena, "A Study on WEKA Tool for Data Preprocessing, Classification and Clustering," *International Journal of Innovative Technology and Exploring Engineering (IJITEE),* vol. 2, no. 6, pp. 250-253, 2013.

[11] D. J. &. J. H. Martin, Regular Expressions, Text Normalization, Edit Distance, Jurafsky Ed3, 2018.

[12] S. Vijayarani, J. Ilamathi and Nithya, "Preprocessing Techniques for Text Mining - An Overview," *International Journal of Computer Science & Communication Networks,* vol. 5, no. 1, pp. 7-16, 2015.

[13] "Machine Learning," Google Inc., [Online]. Available: https://developers.google.com/machine-learning/guides/text-classification/step-3#n-gram_vectors_option_a. [Accessed 22 November 2019].

[14] "Medium," Medium, 8 January 2019. [Online]. Available: https://medium.com/@pemagrg/one-hot-encoding-129ccc293cda. [Accessed 25 November 2019].

[15] F. A. Ozbay and B. Alatas, "Fake news detection within online social media using supervised artificial intelligence algorithms," *Physica A: Statistical Mechanics and its Applications 540,* no. 12317417, pp. 1-17, 2020.

[16] K. Poddar, G. B. A. D and U. K. S, "Comparison of Various Machine Learning Models for Accurate Detection of Fake News," in *Innovations in Power and Advanced Computing Technology (i-PACT)*, Vellore, India, 2019.

[17] J. Brownlee, "Machine Learning Mastery," 20 June 2016. [Online]. Available: https://machinelearningmastery.com/dropout-regularization-deep-learning-models-keras/. [Accessed 22 November 2019].

[18] J. Brownlee, "Machine Learning Mastery," 7 December 2018. [Online]. Available: https://machinelearningmastery.com/early-stopping-to-avoid-overtraining-neural-network-models/. [Accessed 23 November 2019].

Appendix

The source code of this project can be obtained from: https://github.com/ShengHow95/FakeNewsDetection.git