Journal of Theoretical and Applied Information Technology
30th April 2021. Vol.99. No 8

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

FAKE NEWS DETECTION BASED ON WORD AND DOCUMENT EMBEDDING USING MACHINE LEARNING CLASSIFIERS

[1]IBRAHIM EL DESOUKY FATTOH * [2] FARID ALI MOUSA

[1] Computer Science Department, Faculty of Computers and Artificial Intelligence, Beni-Suef University, Egypt.
[2] Information Technology Department, Faculty of Computers and Artificial Intelligence, Beni-Suef University, Egypt.
Email: [1]Ibrahim_desoky@fcis.bsu.edu.eg , [2] faly@msa.eun.eg

ABSTRACT

Fake news is a problem that has a major effect on our life. Detection of fake news considered an interesting research area that has some limitation of the available resources. In this research, we propose a classification model that is capable of detecting fake news based on both Doc2vec and Word2vec embedding as feature extraction methods. Firstly, we compare between the two approaches using different classification algorithms. According to the applied experiments, the classification based on Doc2vec model provided promising results with more than one classifier. The Support vector machine resulted the best accuracy with 95.5% followed by Logistic Regression 94.7% and the Long Short Term Memory produced the lowest accuracy. On the other hand, the classification based Word2vec embedding model results high accuracy only with Long Short Term Memory classifier with 94.3%. Secondly, the classification models based on proposed Doc2vec have shown to outperform a corresponding model that based on TF-IDF on the same dataset using Support Vector Machine and Logistic Regression classifiers.
Keywords: *Fake News Detection; Word2Vec; Doc2Vec; Machine Learning; Deep Learning*

1. INTRODUCTION

With the launch of the World Wide Web and high acquiring of social networks (e.g., Facebook and Twitter), the sharing of information has opened the way that is never seen before in human history.

In addition to other applications, newsagents have gained from extensive usage of social media sites by supplying their subscribers with an up-to-date news in nearly real time. The media expanded from newspapers, magazines and tabloids to modern media like internet news platforms, websites, social media streams, etc. This, sadly, is being used to deceive readers and easily disseminate false news which normally contains misrepresented, or even fabricated photographs.[1]

Spreading false news can have major negative consequences, and even significant public affairs may be influenced or even distorted. In recent years, fake news, particularly after the 2016 US elections, has gained more attention. Fake news is difficult for people to spot. The only way that a person can detect false news manually is by getting an immense understanding of the subject being covered. And if the information in the report is true or false, it is very difficult to identify effectively.[2]

The false news is often portrayed as factually precise, but it does not really exist. We assume in the world today that what we read on blogs or social media is true and do not seek to verify whether the information given is accurate or not. Since the modern culture, individuals, technologies and systems make all these attempts, we still see every day in some form or form some false news.[3]

The detection of fake news can be very helpful for improving the environment in many respects in a scenario of tremendous quantities of false news knowledge. When we dig at this issue of bogus news more closely, there are two broad fields of inspiration, one is the identification of the fake news and the second is the classification of fake news and their wide variety.[4]

The key goal is to identify the false news, which is a conventional text classification issue with an immediate proposal. The model is important to discern between 'real' and 'false' news. This results in implications for social networking platforms such as Facebook, Instagram, Twitter and for instant messaging apps such as, WhatsApp, where this fake news receives a huge boost and becomes viral around the country and worldwide. The method suggested helps in finding the news' credibility. If the news is not accurate, the news report would recommend the reader with the relevant news article[5].

In this paper, we propose using both word embedding (Word2vec) and document embedding (Doc2vec) as feature extraction methods for the creation of various classifications such as logistic regression, support vector machines, random forests, neural networks with multilayer perceptron and long short-term memory classifiers. The suggested classifiers used to differentiate between misleading facts and real news. The classifiers are evaluated and educated on a web-based dataset.

This paper is prepared as follow: section 2 will cover the area of related work, section 3 will discuss the proposed model phases, section 4 will cover the used dataset, applied experiments and discusses the obtained results, and section 5 will provides the conclusion and future work.

2. RELATED WORK

Mumbai released three research papers on fake news identification by three students from Vivekananda Education Society Institute of Technology in 2018. They wrote the social media era began in the 20th century in their research paper. Eventually the use of the web expands, the postings rise, the number of papers increases. They used diverse methods and tools to identify fake news such as natural language processing (NLP) techniques, artificial intelligence (AI) and machine learning (ML).[6]. Nguyen Vo of Ho Chi Minh City University of Technology (HCMUT) student in Cambodia has investigated and applied false news identification in 2017. In his project, false news identification, he used Bi-Directional GRU. Yang et al. used some deep learning algorithms besides other deep learning models, such as Auto-Encoders, CNN and GAN.[7]

Rubin et al.[8] recommended a model for the detection of news stories on satire and comedy. They reviewed and inspected 360 news articles in Satiric, primarily in four areas; civic, scientific, the market and soft news. They suggested a support vector machine classification model with five key features that were built on the basis of their satirical news study. There are five features; Absurdity, Grammar, Humor, Punctuation and Negative Affect. Its overall 90 percent accuracy has been reached with just three attribute combinations; Absurdity, Punctuation and Grammar. Researchers in [9] explored the term's uses, especially regarding fake news. Six broad definitions of the word "fake news" have been found such as; news satire, fabrication, news parody, manipulation (for instance, photography), advertisement (for example advertising presented as journalism) and propaganda. Their goals and appropriations were "the look and feel of real news".

Fallis [10] explores how disinformation has become described. He concluded that "disinformation is misleading information with a misleading function". In [8], Rubin splits false news into three separate categories: extreme manufacturing, huge falsification and humorous falsehoods. Instead of any other grouping, they do not explain why they choose these types. However, they discussed in details what will each group contain and how to discriminate between them. They further highlight the absence of a body to perform such a study and underline the 9 criteria on the creation of such an entity which are; ' 'Digital textual format entry,' 'Verifiable ground facts', 'Predefined time frame', 'Language and culture', 'The manner of delivering news', 'pragmatic concerns' and 'Homogeneity of lengths and writing matter. Testing with the classification challenge, Kai Shu, Amy Sliva, Suhang Wang and Jiliang Tang [11] suggested features such as; a number of characters per words, number of words, phrase rates and sentences, sections of speech tags (i.e., n-grams, and bag-of-word approaches).

In the research paper, the authors of [12] proposed fraud detection with a named benchmark 'LIAR' dataset, and an apparent increase in the efficiency of false messages/news identification appeared. The authors argued that corpus was used to identify places, mining views, rumor identification and analysis on NLP policies.

The researchers of [13] implemented the need to spot hoaxes. They used the ML approach by integrating advertising with approaches to social content. The researchers stated that the study performed well compared to the literature mentioned. They used Facebook messenger chatbot to execute it. There are three numerous datasets of Facebook Italian news posts included. The Boolean crowd sourcing algorithms are all content oriented approaches were implemented with social and content signals. Tabloidization in the form of Click baiting was identified by the authors in [14]. Click baiting was described as a method of fast online propagation of rumors and misinformation. As a means of dissatisfaction, scholars have explored possible methods to automatically detect clickbait.

Content metrics including the lexical and semantic theoretical standard, were implemented by the authors. In [15], the authors have studied and analyzed the respective scope and performance and values, tools and algorithms were used to distinguish falsified and manufactured news stories. The paper also described the study difficulties by the unknown features of fake news and the complex ties between news stories, writers and topics.

The authors discussed the Fake-Detector paradigm for automatically false news inference. It depends on textual identification and provides a wide-ranging network paradigm for concurrently studying the representations of news stories, writers and topics.[16]. Fake-Detector tackles two key components: the feature learning representation and reputation mark inference, which together form the profoundly diffused Fake-Detector network model.

Another research by William Yang et.al. in [17], because of a lack of fake data and its efficacy. He decided to introduce a new 'LIAR' dataset. It is a news identification, freely accessible information set for the false news. Using it utilizing a wide variety of methods as logistic regression, support vector machine and deep-learning (Bidirectional-LSTM and Convolutional Neural Networks) models. The findings showed that CNN models are indeed the best. Himank Gupta. et. a l. [18] has a system focused on a particular approach of machine learning that solves a number of problems, including precision shortages, time lag (Bot-Maker) and fast time computation for thousands of tweets in one second. First, 400,000 HSpam14 dataset tweets were obtained. Then, the 150,000 spam and 250,000 non-spam tweets are further characterized. They also extract few lighter features alongside with top 30 words which give the Bag-of-Words model the highest information gain. They reached 91.65% accuracy and exceeded the current approach by almost 18%. The author referred to Kai Shu et al. in [11], for a more detailed survey of false news identification work trying to concentrate on a lightweight click baits identification system focused on high-level title functionality.

3. PROPOSED MODEL

3.1 Data Preprocessing

Until creating a vector model, the data must be liable to some refinements such as elimination of stop - words, tokenization, lower case segmentation and removal of punctuations. This helps one to decrease the size of the real data by eliminating the unnecessary data. For each document, we built a standardized feature for handling to eliminate punctuation and non-letter characteristics.

Stop Word Removal: Stop words in a language are nonsensical words that produce noise when used as a text classification features.

There are words widely found in phrases to better link thoughts or the structure of the expression.

Stop words are known as posts, preparations, conjunctions, and pronouns. Popular terms have been removed from each document such as; about, a, an, as, are at, by, be, from, for, in, is, on, how, or, the, that, this, these, was, when, where, who, what, will, too, etc. After that, the processed documents were preserved and transmitted to the next stage.

Stemming: The next step is to change the tokens after the tokenization into a regular form of the data. Stemming essentially translates the terms to the original form and reduces the number of word types or classes within the results. The terms 'running,' 'ran,' or 'runner' shall be reduced to 'run'. Our stubble is used for easier and more effective classification. In addition, we use Porter stemmer due to its accuracy, as it is the most widely used stemming algorithms.

3.2 Feature Extraction

High dimension learning is one of the categorization challenges. text description.

A large expression number, phrases and words are contained in the documents which result in the learning process becoming highly computational.

In addition, the classifier precision and output may be influenced by irrelevant and redundant functions. It is better to reduce the feature size and prevent wide area measurements of the feature. The integrations for the majority of our models are generated with the models Word2Vec and Doc2Vec.

3.2.1 Word2vec embedding:

Word2vec was developed for "Representation of continuous vector computing of words from larger datasets" by Tomas Mikolov and Google team in 2013 [19]. They wanted to produce faster learning models, which were common at the time and clearly easier, in a different way from neural network models.

The author and his team have proposed two predictive models – the CBOW (Continuous Bag of Words) and the Skip-gram. The predictive models train their vectors to boost their predictive capacities, so that better outcomes are learned.

Continuous bag-of-words (CBOW):

Words in this model are taken for instance, then all the vectors are added and the relation between the words in a sentence to predict the missing word in a sentence by applying the same projection

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

window. The reason why they call it bag-of-words is that it is not necessary to order words.

Skip-gram: The reverse model of the CBOW model. The input layer will contain the target word and on the output layer, the word context. But how do you guess the phrase from a word? First, vocabulary words from the training documents will be needed. Second, inserting the word we want to use in one-hot vector, then we will put 1 for the word used and 0 for the rest. After that, we multiply the one-hot vector with a matrix of potential phrases and our one-hot vector will fit the correct one.

3.2.2 Do2vec embedding:

Doc2vec is a technique that uses a vector to describe a text which is a simplified version of Word2vec embedding in natural language processing. The distributed Bag of words (DBOW) and the Distributed Memory (DM) models Doc2vec suggested. DBWO is a doc2vec analogue model for the word2vec skip-gram. In the current sense or in the essay the DM mind remembers what is lost. it will reduce the BOW problems to a minimum[20]. Doc2VecC consists of an input layer, a screen layer and an output layer that forecasts the target word. Identical to Word2Vec. The embedded words of adjacent documents have a local meaning while the whole document is interpreted by the vector as a global context. Unlike Paragraph Vectors, which explicitly learns a vector unique to each article, Doc2VecC is the average of the word inlays sampled randomly in the text.

Each paragraph is mapped in a DM to a single vector described by a matrix D column and each term is linked to a special vector, which is expressed by a W matrix column. The vectors of paragraph and phrase are averaged or concatenated in order to forecast iin context the next word. Concatenation as the method for integrating vectors is used in the experiments. It can be used as another term in paragraph token. It serves as a memory that recalls what is lost – or the subject of the paragraph.

Another approach is the DBOW model to neglect the input contextual terms but push the model to predict randomly sampled words in the output from the paragraph. In particular, in each iteration of the stochastic gradient descent, a text window is taken, a random word is taken from the text window, and a ranking task is created, provided the vector in paragraphs.

3.3 Detection of Fake News

The basic steps of the proposed model for detecting fake news, and classifiers used in this research shown in figure1. The first step is text preprocessing, and then creating the features vectors using two different methods (Doc2vec and Word2vc). We used for detecting the class of the new document five classifiers. The classifiers are, Support Vector Machine (SVM), Artificial Neural Network (ANN), Long Short Term Memory (LSTM), Logistic Regression (LR), Multi-Layer Perceptron (MLP) and Random Forest (RF). Word2vec and DM model of Doc2vec are used as feature extraction step for the SVM, LSTM, MLP, LR, and RF, the extracted features are used to train the classifiers and then test each model against test set.

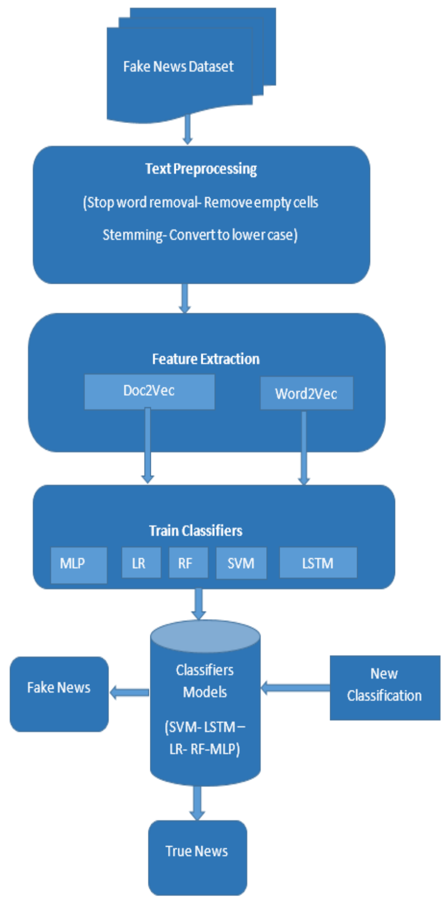

Figure 1. Fake News Detection Proposed Model

The following is a brief description for the applied classifiers

3.3.1 Logistic regression

Logistic regression is such a well-known classification method. The function variables are known to be non-random in the simplest edition. The class response is a binary random variable which takes on a value of 1 for the interest class with

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

a certain likelihood p and a value of 0 with likelihood 1 − p. The success probability "p" is a values function of the feature variables, in fact, the logs odds " ($\log[\frac{p}{(1-p)}]$) or the odds logarithm ratio of the predictor variables is a linear function.
Logistic regression includes hypothesis checking, along with other evaluations, calculations and fitness measures on the value of each variable.
The variable importance checking can be used to selection of features in the classification configuration. Modern computer implementations involve several variations of stepped variable selection (iterative).
The logistic classification might be the most commonly used data mining method due to the mathematical analogy with ordinary multiple regression and the simplicity of automated variable collection.[21]
Components of a probabilistic machine learning algorithm: Logistic regression, like the naive Bayes, is a graded probabilistic that uses supervised instruction. A training group of input/output pairs is required for machine classifiers ($x^{(i)};y^{(i)}$). We can use superscripts in parentheses for each instance in the training collection — any case could be an individual classification document:

1. The input feature representation. This is a vector of features for any input observation $x^{(i)}$, [$x_1;x_2$; :::;x_n]. We are going to list the input feature i, $x^{(j)}$ like $x_i^{(j)}$, often simplified as xi, However, we will notice the notation f_i, $f_i(x)$, or for the classification of the multiclass, $f_i(c;x)$.
2. A classification function determined by p(y|x) to Y for the approximate class. The Sigmoid and SoftMax Classification Tools are added in the next section.
3. Target learning function that typically eliminates errors in training scenarios.
4. An algorithm for objective function optimization (see Sigmoid function in Figure 2). We implement the algorithm for stochastic downward gradient.
Logistic regression has two stages: the first one is the training phase: the method is trained with stochastic downward gradient and cross-entropy loss (specifically weights w and b).
Providing test example x, p(yjx) is determined and the mark y = 1 or y = 0 is returned.

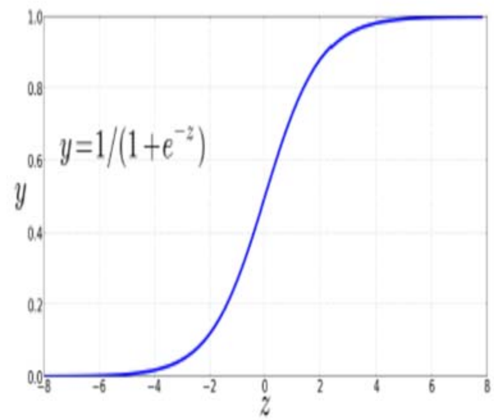

Figure 2. The Sigmoid function of Linear Regression

3.3.2 Random forest classifier

A random forest is a meta-estimator composed of several decision trees working as an ensemble. When constructing each individual tree, it uses baggage and randomness to attempt to create a forest of negatively correlated trees whose committee prediction is more reliable than any single tree. Each single tree spreads a class prevention in the random forest, and the most voted class becomes the predictor of the model.
The random forest model performs out well, because a large number of comparatively negatively correlated trees (model) that function as a committee outperforms the individual constituent models, as well as because the low association between models is the secret to this random forest model.[21]
Random forests are an important entire method based on the CART algorithm for individual trees cultivation. The leading idea is to incorporate several trees into a robust ensemble instead of only constructing one decision book for prediction. Random forests use bagging to plant numerous trees by taking a significant amount of the training data samples, that is, by sampling and substitution, samples that are the same size as the training data. The sampled data is then sent to a CART-like algorithm to construct for each bootstrap sample, a decision tree. Since the samples of this bootstrap comprise various sections of the original data, the trees will vary between samples and form a collection of separate trees.

3.3.3. Multilayer perceptron (MLP)

A multi-layered perceptron (MLP) is a feedforward artificial neural network model, where the input data sets is mapped to a collection of suitable outputs. It comprises three layers forms; the input, the output and the hidden layer. The input layer receives the processing signal.

The appropriate task is performed via the output layer including prediction and classification. The true processing engine of the MLP consists of an infinite number of hidden layers between the input and output side. Like a transmission network in an MLP, the data flow in a forward direction from the input layer to the output layer, like the feed-forward neural network.

The MLP neurons are conditioned by the back propagation neural network. MLPs have been developed to approximate any continuous function and to solve problems that cannot be isolated linearly. Pattern description, identification, estimation and approximation are the key cases of MLP application[22].

The ANN classifier consists of nodes, an input layer as $(x_1, x_2, \dots x_n)$ and an optional hidden layer as well as an output layer (y), (shown in Figure 3). The ANN's goal is to analyse a set of weights (w) (between the input nodes, hidden nodes, and output nodes) that minimize the error of the total sum square.

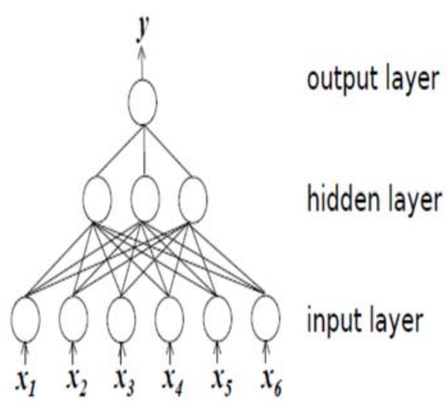

Figure 3. Artificial Neural Network

The weights w_i are attuned with respect to the learning parameter λ ∈ [0, 1] while in training, till all the outputs turn out to be consistent with the output. Too dramatic weight changes may be made if λ is high, while further iterations may be needed (called epochs) if the value is too small before the model learns enough from the training data. The specification of parameters that are learned from training data (i.e. the stock of training data) is known to be difficult to use ANN and thus poorly operate on invisible data. If there are too many hidden nodes, the system could overwrite the current data while stopping the system from accurately updating the input value if there is too few. In addition, it is important to pick a stop criterion. This can involve stopping when the network cumulative loss is below those defaults or if there are a number of epochs (iterations) done.

3.3.4 Long short-term memory (LSTM)

Long short-term memory (LSTM) [16] are a specialized type of recurrent neural networks that have the ability to record long-lasting patterns selectively. It is an excellent alternative for modeling sequential data and thus, for studying complex human behavior dynamics. The cell state is considered the long-term memory. Since the cells are recursive in nature, prior data can be stored in them. To change the cell condition, the forget gate below the cell state is used. The forgot gate releases values determine which data to forget by multiplying zero to a position in the matrix. The knowledge is preserved within the cell while the output of the forget gate is one. The gates input specifies the information, the cell states should enter. Finally, in the output gate, state the information should be shifted to the hidden secret state.

LSTM is commonly used in multi-model applications including photo subtitling and It also enhances the vanilla recurrent neural networks (RNNs). It is designed to ensure long-lasting reliance and substantially reduce the issue of progressive disappearance in RNNs. LSTM reads these words one-on-one and retains a memory state mt in dimensions D and a hidden state ht in D, given a series of word {x1,x2,...,xn}.[23] Figure 4 shows the LSTM diagram

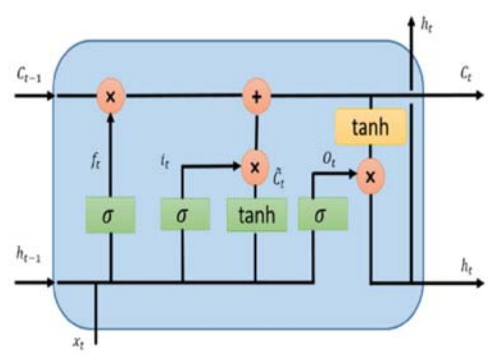

Figure 4. Diagram for Long Short-Term Memory (LSTM)

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

3.3.5 Support vector machine (SVM)

Due to support vector machines' capabilities, high-dimensional data are handled effectively. SVM is a description of the discrimination described by a hyperplane separation. And if SVM itself is a two classifier, it can also be used for multi classifying.
It operates for many classes by applying one technique versus one for each pair of classes. The key purpose of the algorithm is to identify instances dependent on a linear equation. It can also perform a non-linear rating with a kernel function.
The classification is supplied with pre-marked instances and the SVM looks for a hyper-plane that maximizes the margin by choosing points as support vectors.
Support Vector Machines (SVM), which are a basic but effective concept, are one of the most common techniques of linear discrimination. The following works SVM: mapping samples from the input field to a high-dimensional function space in order to locate the 'right' hyperplane while choosing the samples. If its margin is greater, a hyperplane dividing H, is biggest. The margin of the two hyperplanes on both sides parallel to H is the mean distance without sample spans. From the theory of risk avoidance (an estimation of the predicted failure function, i.e., the samples of the misclassification), the greater the margin, the better the classifier generalization error. **Error! Reference source not found.** illustrate this notion. It is apparent that the same training set has different splitting hyperplane. The hyperplane that divides samples from the nearer classes at greater distance from their chosen ones, provided that the two sample groups are separated by the greatest margin of each other and, thus, are least prone to small failures in the direction of the hyperplane. [24, 25]

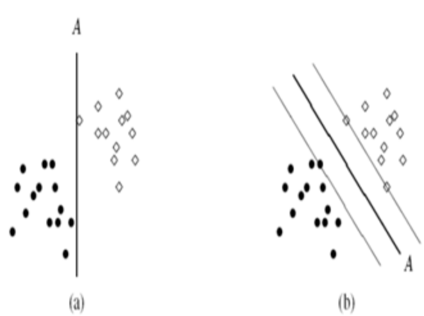

Figure 5. Support Vector Machine

4. DATASET, EXPERIMENTS, AND RESULTS

4.1 Dataset

Two different articles contain false and real news of the dataset used in this study. The dataset collection was made from sources of real-world; true articles from the news website Reuters.com were collected. While the articles of the fake news have been compiled from numerous outlets such as unreliable websites flagged by PolitiFact (a US-fact checking organization) and Wikipedia. The dataset includes various kinds of documents. However, the majority focuses on world and political news issues. More information about the dataset is described in [26, 27]

4.2 Evaluation Metrics

Table 1 shows the four basic evaluation metrices description used to assess the results of the applied experiments

Table 1.Evaluation mercies

	Formula	Evaluation Focus
Accuracy (acc)	$\frac{tp+tn}{tp+fp+tn+fn}$	Typically calculates the percentage of accurate forecasts over the total number of measured instances
Sensitivity (sn)	$\frac{tp}{tp+fn}$	Calculate accurately classified fractions of positive patterns
Specificity (sp)	$\frac{tn}{tn+fp}$	Calculate the proportion of the negatively patterns classified correctly
F-Measure (FM)	$\frac{2\times p\times r}{p+r}$	Describes the harmony among recall and precision values

Where tp represents the True Positive, fp represents the false positive. while tn represents the True Negative and fn represents the false negative.

4.3 Applied Experiments And Evaluation Results

We run the mentioned algorithms on the dataset. The dataset were divided into 80% and 20% for training and testing respectively. We started by examining the effects of the Doc2ve then the Word2vec on the different algorithms. The algorithms used for creating learning models then using them for predicting the labels for test data. Tables 2 and 3 shows the results obtained from the different classifiers based on Doc2vec and Word2vec.

Table 2. Doc2vec Results

Evaluation	LR	RF	MLP	SVM	LSTM
Sensitivity	0.946	0.876	0.941	0.948	0.624
Specificity	0.948	0.942	0.951	0.961	0.565
ccuracy	0.947	0.909	0.946	0.955	0.584
F1 Score	0.943	0.907	0.943	0.952	0.493

Table 3. Word2vec

Evaluation	LR	RF	MLP	SVM	LSTM
Sensitivity	0.622	0.691	0.589	0.649	0.965
Specificity	0.656	0.769	0.559	0.707	0.922
Accuracy	0.639	0.724	0.570	0.674	0.943
F1 Score	0.634	0.743	0.511	0.694	0.941

Figures from 6, 7, 8, 9, and 10 represents the results of each algorithm individually based on the two models Doc2vec and Word2vec to show the effects of the two feature extraction methods on the different classifiers.

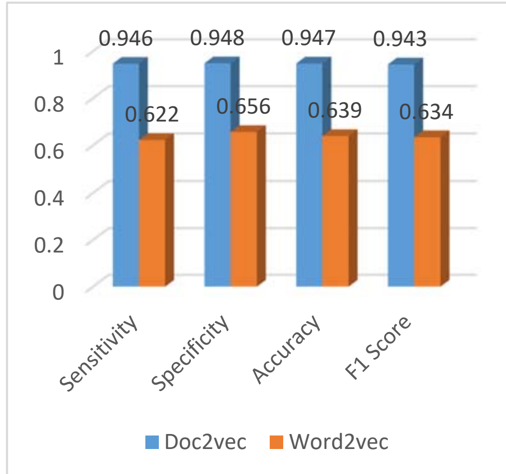

Figure 6. LR Results

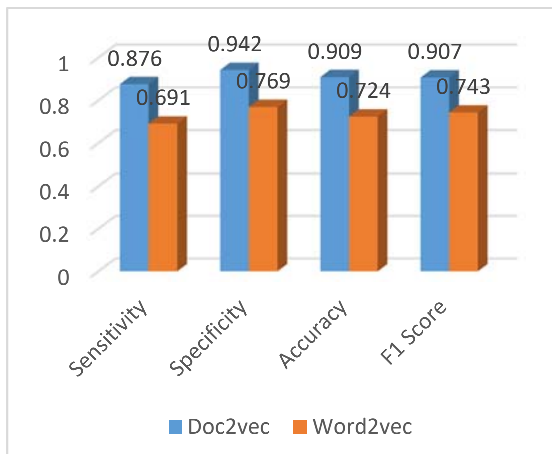

Figure7. RF Results

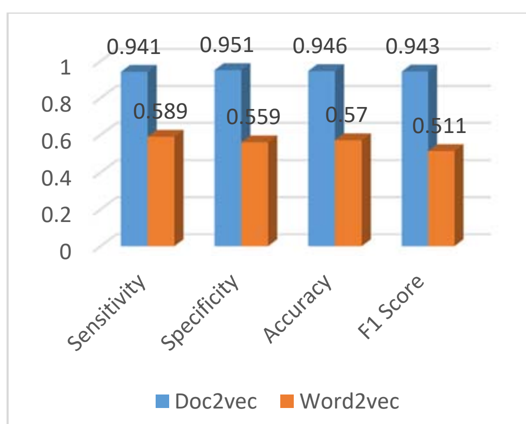

Figure 8.MLP Results

Journal of Theoretical and Applied Information Technology
30th April 2021. Vol.99. No 8

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

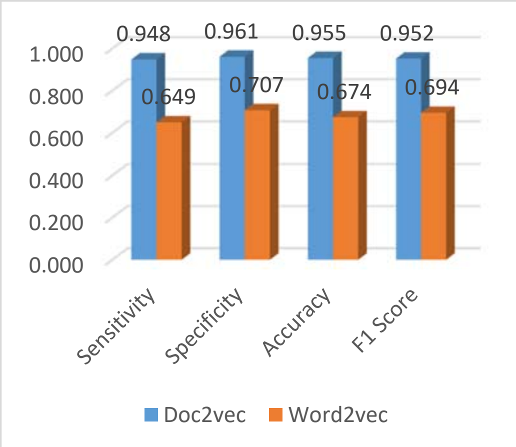

Figure 9.SVM Results

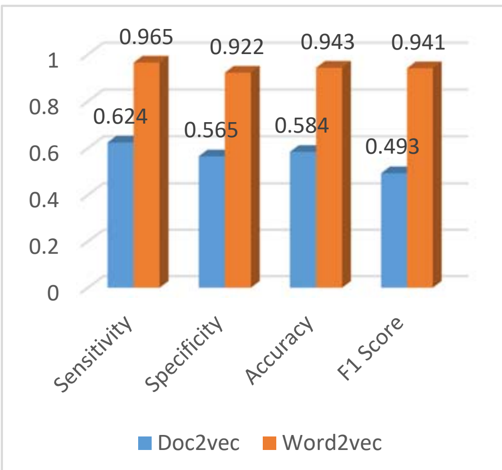

Figure 10. LSTM Results

From the above results, the Doc2vec embedding results a good accuracies in all algorithms except LSTM, while the Word2vec embedding results good accuracy only with the LSTM classifier.
Doc2Vec expands the Word2Vec by providing a document vector to the representation of the output that includes several information about the document; it also enables the model to learn some details about the order of words. The information kept about word order in Doc2Vec useful for SVM, LR, MLP, and RF and the four classifiers are learned from it and yielded acceptable and promising classification accuracy. The resulting accuracy for SVM and LR are 59.5 % and 94.7% respectively. While the way for keeping the order information does not give acceptable results with LSTM. Therefore, we applied the Word2vec also, and ordered the most common word on the training dataset, (the number used is based on the total words in the training after preprocessing steps). Each word was transferred into 32-dimension vector, and each word vector then trained by word embedding based on similarity of words. This representation of Word2vec was fed to the different classifiers as a feature selection and the ordering information are kept in this way were more suitable for LSTM and resulted accuracy was 94.3% and it is a promising accuracy.
Another experiment was applied to compare the classifiers based on Doc2vec model with the work on [27] on the same dataset, in their research they applied many classifiers, and the Linear Support Vector Machine resulted the best accuracy 92% while they used TF-IDF as feature extraction. They applied their experiments on a subset of 2000 documents of the same dataset that focus only on politics. when comparing the results of SVM and LR algorithms based on the Doc2vec model we noticed that the results obtained outperforms the results in [27] as shown in table 6.

Table 6.Comparing Doc2vec with TF-IDF [27]

Evaluation	Doc2vec	TF-IDF
SVM	95.50%	92.00%
LR	94.70%	89%

In table 6, we show the results obtained from Doc2vec with SVM and LR and the results obtained in [27] for the same classifiers. Although they concentrated their model on a specific topic. The accuracy results obtained from the classifiers that using feature extraction based on Doc2vec outperforms the results of the same classifiers based on TF-IDF.

5. CONCLUSION AND FUTURE WORK

In this research, we described two different approaches that can be used as feature extraction methods for the fake news detection problem. The problem that become very important and sensitive in our social life in last few years. The proposed two approaches are Word2vec and Doc2ve. The primary goal was applying Doc2vec and it proved with many classifier that it can be used to train more than one classifier and resulted a very promising accuracy with SVM and LR. While the results of Doc2vec with LSTM was not acceptable. So we introduced the Word2vec with a way suitable for the LSTM to preserve the order information of words and the resulting accuracy was 94.3% that can be reasonable and promising. In addition, the model based on Doc2vec compared with another model. The model based on TF-IDF as feature extraction with N-gram. The classifiers SVM and LR that based on Doc2vec model resulted higher accuracy than when they were based on TF-IDF. In the future, the models can be applied on different dataset with more documents, and hybridization between more than one classifier can enhance the accuracy of classification

REFERENCES:

[1] Welbers, K. and M. Opgenhaffen, Social media gatekeeping: An analysis of the gatekeeping influence of newspapers' public Facebook pages. new media & society, 2018. 20(12): p. 4728-4747.

[2] Jankowski, N.W., Researching fake news: A selective examination of empirical studies. Javnost-The Public, 2018. 25(1-2): p. 248-255.

[3] Carlson, M., Fake news as an informational moral panic: the symbolic deviancy of social media during the 2016 US presidential election. Information, Communication & Society, 2020. 23(3): p. 374-388.

[4] Sabeeh, V.T., Machine Learning and Semantic Knowledge Assisted Fake News Detection Models. 2020, Oakland University.

[5] Molina, M.D., et al., "Fake news" is not simply false information: a concept explication and taxonomy of online content. American behavioral scientist, 2019: p. 0002764219878224.

[6] Dubey, A.K., A. Singhal, and S. Gupta, Rumor Detection System using Machine Learning. International Research Journal of Engineering and Technology (IRJET) Vol. 7, No. 5(2020).

[7] Lan, N.P.H., et al., Phenotypic and genotypic characteristics of ESBL and AmpC producing organisms associated with bacteraemia in Ho Chi Minh City, Vietnam. Antimicrobial Resistance & Infection Control, 2017. 6(1): p. 1-9.

[8] Rubin, V.L., et al. Fake news or truth? using satirical cues to detect potentially misleading news. in Proceedings of the second workshop on computational approaches to deception detection. 2016.

[9] Tandoc Jr, E.C., Z.W. Lim, and R. Ling, Defining "fake news" A typology of scholarly definitions. Digital journalism, 2018. 6(2): p. 137-153.

[10] Fallis, D., What is disinformation? Library trends, 2015. 63(3): p. 401-426.

[11] Shu, K., et al. defend: Explainable fake news detection. in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019.

[12] Meel, P. and D.K. Vishwakarma, Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. Expert Systems with Applications, 2019: p. 112986.

[13] Ozbay, F.A. and B. Alatas, Fake news detection within online social media using supervised artificial intelligence algorithms. Physica A: Statistical Mechanics and its Applications, 2020. 540: p. 123174.

[14] Collins, B., et al., Trends in combating fake news on social media–a survey. Journal of Information and Telecommunication, 2020: p. 1-20.

[15] Mackey, T.K. and G. Nayyar, A review of existing and emerging digital technologies to combat the global trade in fake medicines. Expert opinion on drug safety, 2017. 16(5): p. 587-602.

[16] Halder, M., M. Suresh, and S. Rao, Impact of Discourse in Modern AI Based NLP Applications. Smitha Rao, Impact of Discourse in Modern AI Based NLP Applications (July 31, 2020). Institute of Scholars (InSc), 2020.

[17] Wang, W.Y., " liar, liar pants on fire": A new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648, 2017.

[18] Gupta, H., et al. A framework for real-time spam detection in Twitter. in 2018 10th International Conference on Communication Systems & Networks (COMSNETS). 2018. IEEE.

[19] Al-Ansari, K., Survey on Word Embedding Techniques in Natural Language Processing.

[20] Le, Q. and T. Mikolov. Distributed representations of sentences and documents. in International conference on machine learning. 2014. PMLR.

[21] Pranckevičius, T. and V. Marcinkevičius, Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification. Baltic Journal of Modern Computing, 2017. 5(2): p. 221.

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

[22] Crabbé, A., et al., Neural Network MLP Classifier. 2020.
[23] Demotte, P., et al. Sentiment Analysis of Sinhala News Comments using Sentence-State LSTM Networks. in 2020 Moratuwa Engineering Research Conference (MERCon). 2020. IEEE.
[24] LIN, H., L. LI, and H. WANG, Survey on research and application of support vector machines in intelligent transportation system. Journal of Frontiers of Computer Science and Technology, 2020. 14(6): p. 901-917.
[25] Alam, S., et al., One-class support vector classifiers: A survey. Knowledge-Based Systems, 2020. 196: p. 105754.
[26] Ahmed, H., I. Traore, and S. Saad, Detecting opinion spams and fake news using text classification. Security and Privacy, 2018. 1(1): p. e9.
[27] Ahmed, H., I. Traore, and S. Saad. Detection of online fake news using n-gram analysis and machine learning techniques. in International conference on intelligent, secure, and dependable systems in distributed and cloud environments. 2017. Springer.