International Journal of Advanced Nursing Education and Research
Vol.8, No.1 (2020), pp.15-26
http://dx.doi.org/10.21742/ijriar.2020.8.1.03

Experiments on Detecting Fake News over Social Media using Machine Learning Algorithms

Harika Kudarvalli[1] and Jinan Fiaidhi[2*]

Department of Computer Science, Lakehead University, Canada
[1]hkudarva@lakeheadu.ca, [2]jfiaidhi@lakeheadu.ca

Abstract

Spreading fake news has become a serious issue in the current social media world. It is broadcasted with dishonest intentions to mislead people. This has caused many unfortunate incidents in different countries. The most recent one was the latest presidential elections where the voters were misleading to support a leader. Twitter is a popular social media platform where it represents the gateway for real time news. We extracted real time data on multiple domains through twitter and performed analysis. The dataset was preprocessed and user verified column played a vital role. Multiple machine algorithms were then performed on the extracted features from preprocessed dataset. Logistic Regression and Support Vector Machine had promising results with both above 92% accuracy. Naive Bayes and Long-Short Term memory didn't achieve desired accuracies. The model can also be applied to images and videos for better detection of fake news.

Keywords: *Fake news, Twitter, Machine learning, Detection system*

1. Introduction

Twitter data is the most comprehensive source of live, public conversation worldwide. Their REST, streaming, and Enterprise APIs enable programmatic analysis of data in real-time or back to the first Tweet in 2006. Twitter is data-rich repository. Twitter tweets are public and with simple streaming API, anyone can upload these streams at the time they are tweeted to others. This is a great feature but it holds with it the danger of misleading and false information. According to Soroush Vosoughi, a data scientist at MIT who has studied fake news since 2013, it is pretty clear that false information outperforms true information [1]. Twitter's streaming API allows you to search for tweet about a certain topic using keywords or hash tags and pulls associated followers, retweets and mentions within the Twitter general environment or within specified geolocation.

Fake news affects every corner of the society. Presidential elections in 2016 were notable for incidents of fake news, including a propaganda created to mislead readers or to generate views on websites or steer public opinion. Celebrities like Rowan Atkinson, Justin Bieber were also affected when a "hoax" stating that they had died was circulated. More recently, there are many rumors [2] about the transmission of Hantavirus which has been in existence since 1993 and due to the panic of COVID 19 in society; this virus was also propagated as a highly contagious virus. But Centre of Disease Control and Prevention [3] have squatted down fears of another pandemic.

Article history:
Received (February 24, 2020), Review Result (April 2, 2020), Accepted (May 8, 2020)

Print ISSN: 2288-8950, eISSN: 2207-5143 IJRIA

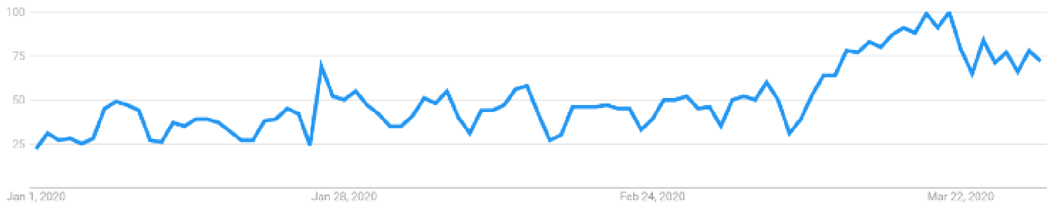


Figure 1. Google trends on fake news

Unlike other social media like Facebook, Twitter do not encourage users to actively report misinformation. In Twitter there is no option for user flagging. Users can still report such misinformation. The goal of this research is to extract real time twitter data with multiple news keywords and analyze the results using machine learning algorithms to classify them as fake or genuine news. In section II describes the related work. Section III discusses the methodology used to develop the classification models, dataset used and various machine learning algorithms. Finally, sections IV discuss the results where section VII concludes the work.

2. Related research work

2.1. Detecting fake news in social media networks

The authors in [4] suggested that most fake news contains what is called clickbaits. Clickbait's are thumbnails that attract users and generate curiosity with flashy headlines or designs to drive users to click these links to take them to the fake news. The purpose of their research work is to come up with a solution that can be utilized by users to detect and filter out sites containing false and misleading information based on clickbaits. In conclusion, they use simple and carefully selected features of the title and post to accurately identify fake posts. The experimental results show a 99.4% accuracy using logistic classifier.

2.2. Detecting fake news using machine learning and deep learning algorithms

Social media provides a platform to broadcast news to the network at an exponential rate and is a great source of information. But this does not imply that everything that we see today in the news or social media is true. Words, photos, videos are "photoshopped" into being delusional or subjective to their own propaganda [5]. The authors in [5] proposes a model for recognizing forged news messages from twitter posts by using machine learning classifiers trained on sound datasets like the FNC-I [6]. In this research authors compare the main tweet object with its child objects if they agree on the topic or not. This can easily reveal a fake tweet if one of the child object discusses a different topic. Each tweet object has a long list of 'root-level' attributes, including fundamental attributes such as id, created_at, and text. Tweet objects are also the 'parent' object to several child objects. Tweet child objects include user, entities, and extended_entities. Tweets that are geo-tagged will have a place child object. As a result, the authors performed a comparison between five well-known Machine Learning algorithms, like SVM, Naïve Bayes Method, Logistic Regression and Recurrent Neural Network models, separately to demonstrate the efficiency of the classification performance on the dataset. The different learning algorithms were compared and result showed that SVM and Naïve Bayes classifier outperforms the other classifiers.

2.3. Which machine learning paradigm for fake news detection

The authors in [7] presented a comprehensive performance evaluation of eight machine learning algorithms for fake news detection/classification. The approach is done into two stages. The first stage is about fact finding where similar raw facts are collected from the web and then in the second stage a comparison is made with the news article under investigations. The comparisons are based on classifiers that provide information on its correlation with the collected knowledge base. The second stage can be called as fact fetching stage where it can provide sense of realization of what is fake news or not.

2.4. Detection of online fake news using n-gram analysis and machine learning techniques

In [8] the authors developed a detection model for fake news using the availability of n-grams as corner stone for the features extraction processes involved in analyzing the articles under investigation. Two different features extraction techniques were employed and six different machine learning techniques. The n-gram model achieves its highest accuracy when using unigram features along with Linear SVM classifier. The highest accuracy score is 92%.

3. Describing our methodology

Learning from the previous research, we have developed a hybrid methodology for identifying fake news. It combines news style and feature extraction along with machine learning. It also involve human intelligence as it provide window of visualization. The entire research has been defined by following steps and illustrated in [Figure 2]:

- Data Retrieval
- Data Preprocessing
- Data Visualization
- Tokenization
- Feature Extraction
- Machine Leaning Algorithms
- Training & Testing Model
- Evaluation Metrics

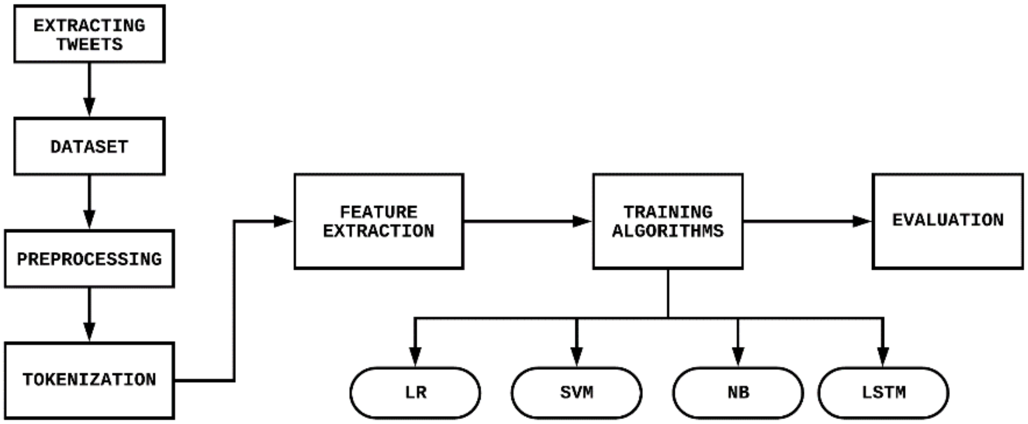


Figure 2. Flowchart of methodology

3.1. Data retrieval

Twitter is our main source of data. After creating a Twitter Developer account, we were given credentials to start scraping tweets. We used multiple keywords to download tweets like Donald Trump, politics, coronavirus. User and tweet information was also extracted like ID, username, and tweet text and if the user is verified. User verified column was our decision variable where it was assumed that if a user verified account is posting a tweet that is most likely to be reliable news. It is also assumed that people trust anything conveyed through these verified accounts [9]. Hence it is justified that Twitter goes through long process of validating the user before giving a verified blue tick to their profile. Users who wish to get their account verified must go through long process where twitter does an online background check and authenticates the user. Our dataset consists of 25117 tweets and 4 columns. To make our analysis more user friendly, we converted our user verified column to label where if the user is verified it replaces the text with 0 else 1. The dataset will go through intense preprocessing to evaluate the tweets.

3.2. Data preprocessing

Entire dataset was converted into lower case letters and punctuations were removed. As data is being extracted from Twitter, users tend to use emoticons and various punctuations which do not help in our detection process. Empty columns were detected where at least one of the four columns were missing so they had to be removed. Repeating characters like URL tags and stop words were filtered out of the dataset. Hashtags were also scrapped.

3.3. Data visualization

Once the data is cleaned and ready to be fed to the model, we can understand the data more closely with visualizing its factors in graphs or plots. Graphs give us a clear picture about the categorization of dataset. Below [Figure 3] depicts the top 25 words being used in our dataset

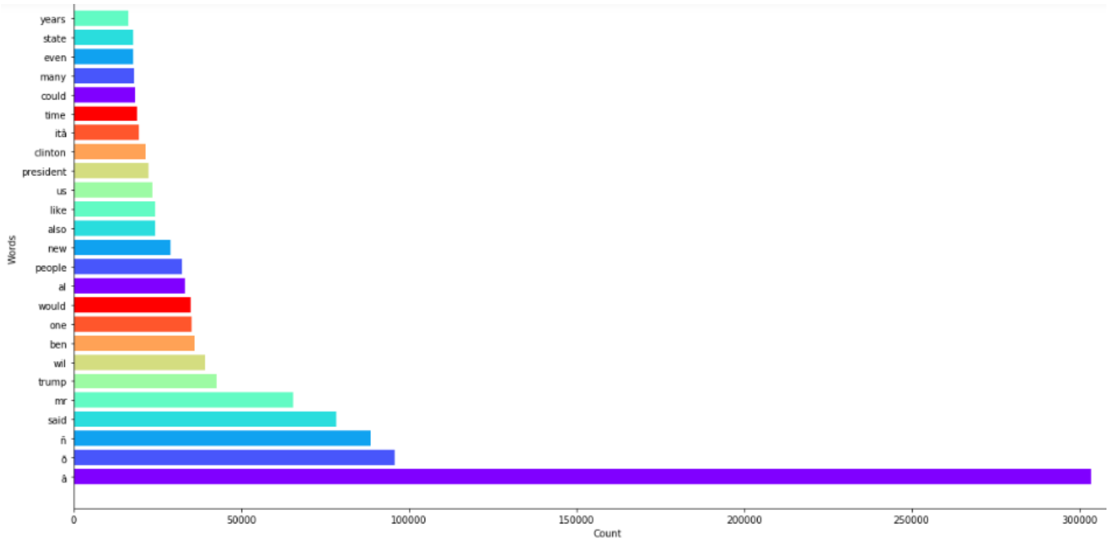
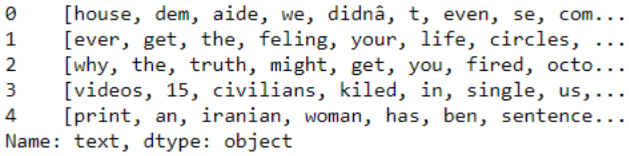


Figure 3. Top 25 words

3.4. Tokenization

Tokenization refers to splitting up a larger body of text into smaller lines, words or even creating words for a non-English language. The various tokenization functions are in-built into the nltk module itself. We have used RegexTokenizer, that extracts tokens either by using the provided regex pattern to split the text (default) or repeatedly matching the regex (if gaps are false). [Figure 4] shows the tokenized words of our dataset. Optional parameters also allow filtering tokens while minimizing the length. It returns an array of strings that can be empty.

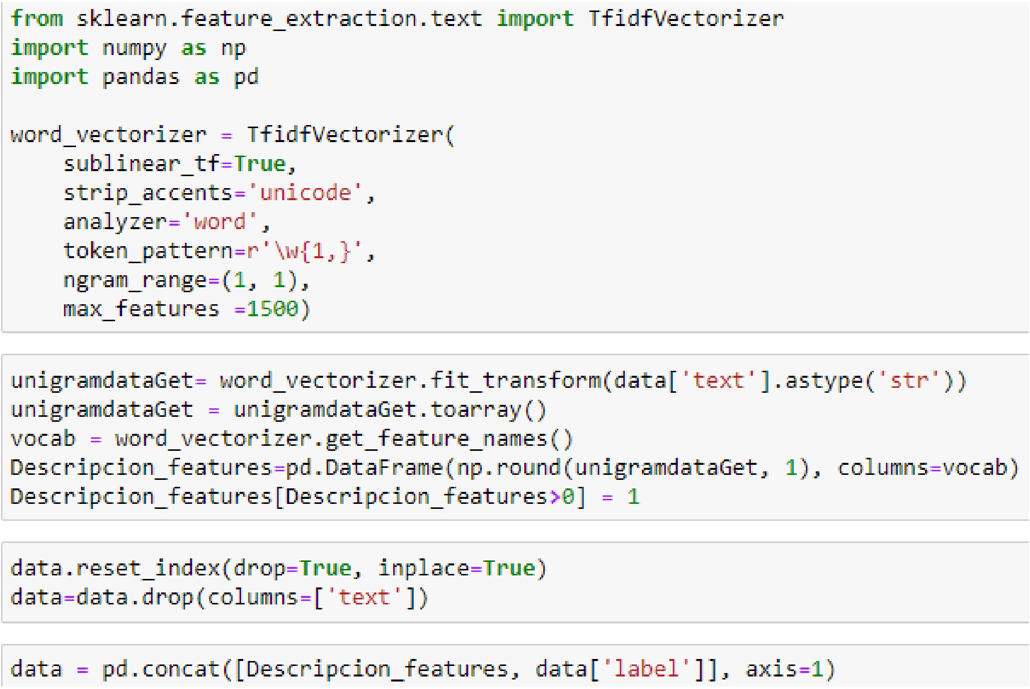


```
0    [house, dem, aide, we, didnâ, t, even, se, com...
1    [ever, get, the, feling, your, life, circles, ...
2    [why, the, truth, might, get, you, fired, octo...
3    [videos, 15, civilians, kiled, in, single, us,...
4    [print, an, iranian, woman, has, ben, sentence...
Name: text, dtype: object
```

Figure 4. Tokenization

3.5. Feature extraction

Feature extraction is a process of dimensionality reduction in which an initial set of raw data is reduced to more manageable groups for processing. By doing this we are reducing the processing time of the compiler and increasing the rate of efficiency in detecting the value of the word. A characteristic of these large data sets is many variables require a lot of computing resources to process. TF-IDF stands for “Term Frequency — Inverse Document Frequency”. This is a technique to quantify a word in documents which allows us to compute a weight to each word which signifies the importance of the word in the document and corpus. This method is a widely used technique in Information Retrieval and Text Mining. Term Frequency summarizes how often a given word appears within a document. Inverse Document Frequency downscales words that appear a lot across documents.



```
from sklearn.feature_extraction.text import TfidfVectorizer
import numpy as np
import pandas as pd

word_vectorizer = TfidfVectorizer(
    sublinear_tf=True,
    strip_accents='unicode',
    analyzer='word',
    token_pattern=r'\w{1,}',
    ngram_range=(1, 1),
    max_features =1500)
```

```
unigramdataGet= word_vectorizer.fit_transform(data['text'].astype('str'))
unigramdataGet = unigramdataGet.toarray()
vocab = word_vectorizer.get_feature_names()
Descripcion_features=pd.DataFrame(np.round(unigramdataGet, 1), columns=vocab)
Descripcion_features[Descripcion_features>0] = 1
```

```
data.reset_index(drop=True, inplace=True)
data=data.drop(columns=['text'])
```

```
data = pd.concat([Descripcion_features, data['label']], axis=1)
```

Once the features have been extracted from our dataset, we don't need the entire dataset for further analysis. Hence, we can drop the text column to increase our processing speed.

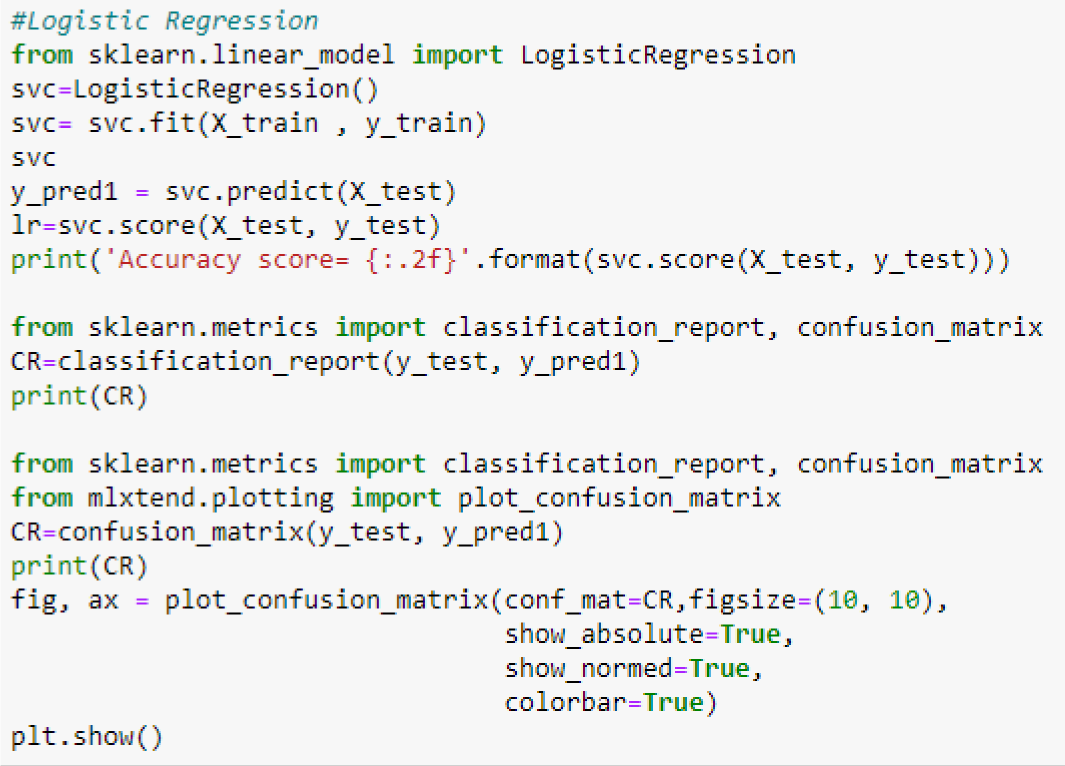
3.6. Machine learning algorithms

3.6.1. Logistic regression

This algorithm is used to predict the probability of a categorical dependent variable.

Dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.). Logistic Regression is a Linear Regression model, but it uses a more complex cost function, this cost function can be defined as the Sigmoid function. The hypothesis stated in eq. 1 of logistic regression [10] tends it to limit the cost function between 0 and 1. Therefore linear functions fail to represent it as it can have a value greater than 1 or less than 0 which is not possible as per the hypothesis of logistic regression. It is classified into binomial, multinomial and ordinal.

$$0 \leq h0(x) \leq 1 \quad \text{Equation 1: LR Hypothesis}$$



```
#Logistic Regression
from sklearn.linear_model import LogisticRegression
svc=LogisticRegression()
svc= svc.fit(X_train , y_train)
svc
y_pred1 = svc.predict(X_test)
lr=svc.score(X_test, y_test)
print('Accuracy score= {:.2f}'.format(svc.score(X_test, y_test)))

from sklearn.metrics import classification_report, confusion_matrix
CR=classification_report(y_test, y_pred1)
print(CR)

from sklearn.metrics import classification_report, confusion_matrix
from mlxtend.plotting import plot_confusion_matrix
CR=confusion_matrix(y_test, y_pred1)
print(CR)
fig, ax = plot_confusion_matrix(conf_mat=CR,figsize=(10, 10),
                                show_absolute=True,
                                show_normed=True,
                                colorbar=True)
plt.show()
```

3.6.2. Naïve Bayes

A Naive Bayes classifier is a probabilistic machine learning model which is used for classification task. The crux of the classifier is based on the Bayes theorem shown in eq. 2. Naïve Bayes [11] takes two simple assumptions into consideration a) Predictors are independent & b) All the predictors have an equal effect on the outcome. It has been classified into 3 types – multinomial, Bernoulli and Gaussian.

$$P(c|x)=\frac{P(X|C)P(c)}{P(x)} \qquad \textit{Equation 2: Bayes Theorem}$$

3.6.3. Long short-term memory

LSTMs were introduced to solve the vanishing gradient problem. They help preserve the error that can be back propagated through time and layers. By maintaining a more constant error, they allow recurrent nets to continue to learn over many time steps (over 1000), thereby opening a channel to link causes and effects remotely. LSTMs [11] as shown in [Figure 5] contain information outside the normal flow of the recurrent network in a gated cell. Information can be stored in, written to, or read from a cell, much like data in a computer's memory.

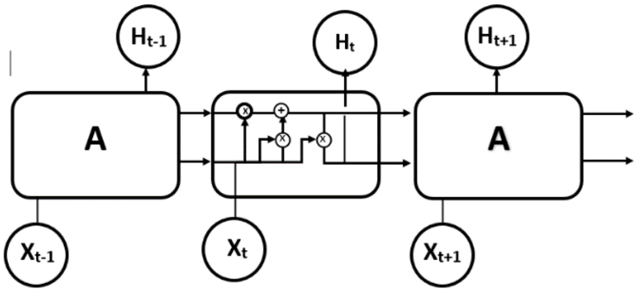


Figure 5. LSTM Architecture

3.6.4. Support Vector Machine

The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space (N representing the number of features) that distinctly classifies the data points. To separate the two classes of data points, there are many possible hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin, i.e., the maximum [12] distance between data points of both classes shown in [Figure 6]. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence.

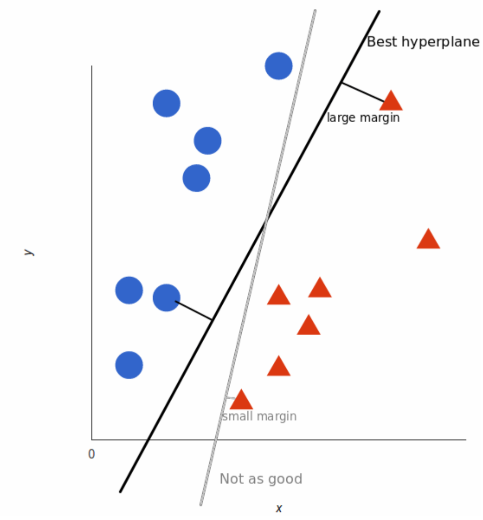


Figure 6. Support Vector Machine

3.7. Training and testing model

Once the network of each algorithm is built, the data must undergo training where we perform supervised learning. Algorithm uses trained data to learn features from dataset. Later, testing is performed to evaluate the performance of our built model using performance metrics like accuracy, precision f1 score and recall. This is also called as confusion matrix.

3.8. Results

We have performed analysis on "Tweets" dataset extracted from Twitter and the results are astonishing as we found some interesting insights. The reviews have been divided into reliable or unreliable data using features of text information. Our research started with extracting real time tweets using keywords and after preprocessing of these tweets features with importance were extracted from the dataset. These features have importance as it contains valuable characteristics that define the dataset. These features are separated to decrease processing load and the tweet information column is dropped. The first algorithm performed on dataset for analysis was Logistic Regression. It is considered as a baseline model for any machine learning analysis as it works well with big data and is less prone to overfitting. We were able to achieve 93.8% with logistic regression and it was an expected result.

Below [Figure. 7] depicts the ROC curve for our model and the curve is almost to perfect line.

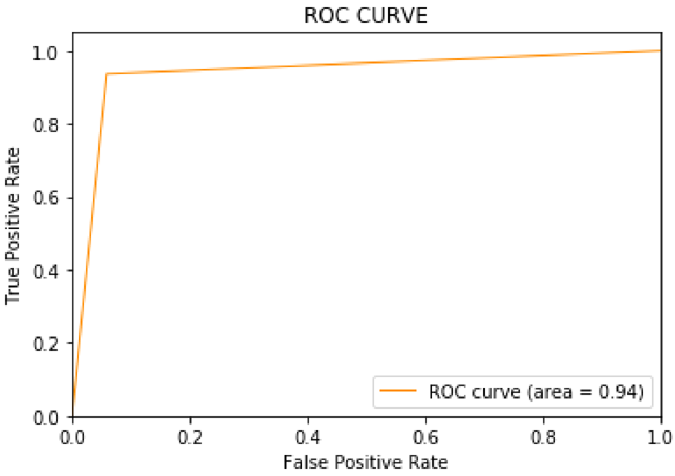


Figure 7. ROC curve for logistic regression

Next algorithm that was experimented was Naïve Bayes. As it is based on Bayes Theorem, it considers each variable independently. Our variables were independent to each other, so it was expected to perform better. But due to zero frequency [13] the classifier achieved only 72.5%. [Figure 8] is the description of our ROC curve for this model.

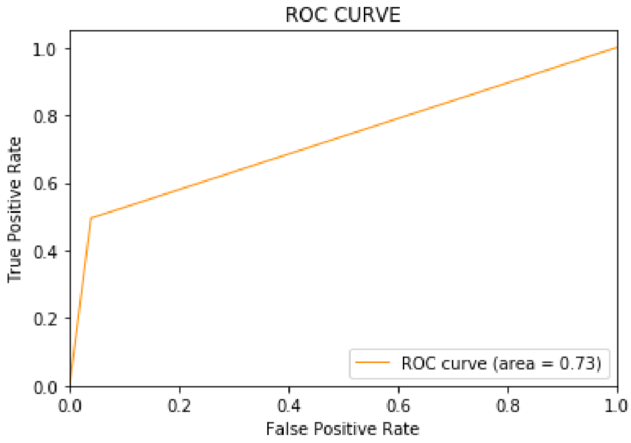


Figure 8. ROC curve for naïve bayes

As you can see, further observations were needed to classify the data therefore we performed analysis using neural networks. Neural Networks always had issues with storing large sequences of data. Therefore, we used Long Short-Term Memory to overcome this problem. It uses memory blocks and gates for its functionality. Early Stopping was also used to restrict the model from overfitting. After running for 5 epochs, we got an accuracy of 50.5% which was the least among out of all our classifiers. This gave an indication that this type of data doesn't need neural networks and it should be trained using a less complicated structure. Following is the figure that depicts the ROC curve of our LSTM model.

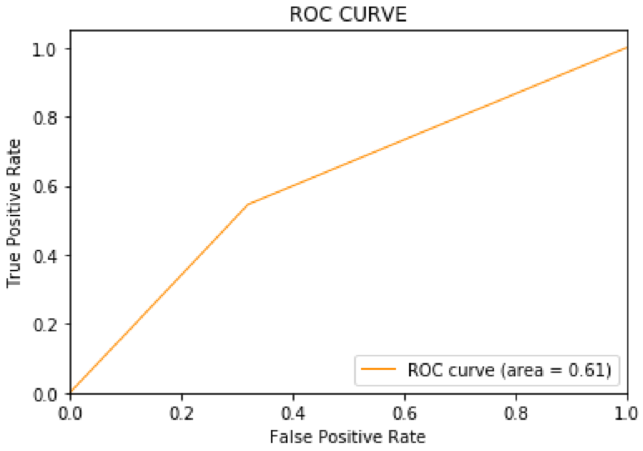


Figure 9. ROC curve for long short-term memory

The last algorithm which was performed on tweets to detect fake news was Support Vector Machine. It uses deciding boundary to separate two classes with biggest margin also called as best hyperplane. LinearSVC method was used to analyze the data extracted. SVM was able to achieve a staggering 92.5% accuracy which is the second-best accuracy achieved after logistic regression. Based on multiple researches SVM works wonders with both linear and non-linear data and is more reliable to give higher accuracy results every time. Below [Figure 10] depicts one of the best ROC curves in our analysis.

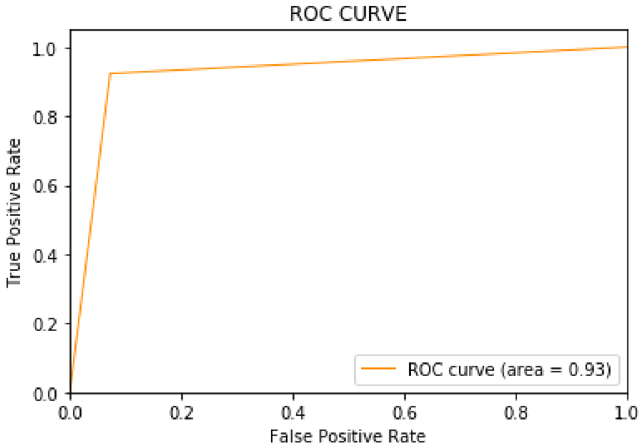


Figure 10. ROC curve for Support Vector Machine

Table 1. Comparison results

Algorithms	Precision	Recall	F1 Score	Accuracy
Logistic Regression	94%	94%	94%	94%
Naïve Bayes	79%	73%	71%	73%
Long Short-Term Memory	72%	51%	35%	51%
Support Vector Machine	93%	93%	93%	93%

4. Conclusions and future research

Due to the booming development of online social networks, the widespread expansion of fake news for various commercial and political purposes has been at an exponential rate. This rise has been constant in all the nations and the online world. It effects the socio-economic climate, the politics within and outside the country, causing pandemonium and panic among civilians. This increases the need of effective fake news detection to multi folds. Most of the models present out there have manually labelled each tweet as real or fake news. This is a tiresome and inefficient work in the modern world and can take up to long hours to confirm the originality of the news. In this research, real time tweets were pulled, and features were preprocessed to extract valuable characteristics. Four types of algorithms were analyzed under multiple evaluation metrics. Logistic Regression and Support Vector Machine presented the best results with more than 90% accuracy. Table 1 illustrates the evaluation metrics of our research. LSTM works best with unstructured data like images and videos. As there are very less absolute credible sources of information, it makes the process of detecting fake news more challenging. Although, with research and development still going on, Multimillion-dollar companies like Facebook, Twitter and Google are still rolling out their beta applications to accelerate the fight against fake news. As the complications of detecting fake news increases, more and more research and development are in order to fan the flames of those working in the field of detecting fake news. More sophisticated and efficient model can be structured by considering multiple parameters while extracting tweets. We would like to extend our research to unstructured data like photos and videos which are more dangerous forms of fake news. Analyzing non-linear data gives an opportunity to research on deep learning algorithms as well. Natural Language Processing could also be used to understand the brief meaning of tweets and compare them with multiple reliable news websites.

Acknowledgements

This research is part of the first author through her research at the Research Methodology Graduate Course (COMP5112) supervised by Dr. Jinan Fiaidhi. The authors would like also to Rochester Institute of Technology for sharing their open dataset for domestic experimentation.

References

[1] Meyer Robinson, “The grim conclusions of the largest-ever study of fake news,” The Atlantic 8 **(2018)**

[2] Joshkelliott, “One hantavirus death in China sparks ‘hysteria’ over old disease,” Retrieved from https://globalnews.ca/news/6724399/hantavirus-china-death-coronavirus/, **(2020)**

[3] CDC – Hantavirus, https://www.cdc.gov/hantavirus/index.html, **(2019)**

[4] M. Aldwairi and A. Alwahedi, “Detecting fake news in social media networks,” Procedia Comput. Sci., vol.141, pp.215–222, **(2018)**

[5] Abdullah-All-Tanvir, E. M. Mahir, S. Akhter, and M. R. Huq, “Detecting fake news using machine learning and deep learning algorithms,” 2019 7th International Conference on Smart Computing & Communications (ICSCC), Sarawak, Malaysia, Malaysia, pp.1-5, **(2019)**

[6] http://www.fakenewschallenge.org/

[7] D. Katsaros, G. Stavropoulos, and D. Papakostas, “Which machine learning paradigm for fake news detection?” 2019 IEEE/WIC/ACM International Conference on Web Intelligence (WI), Thessaloniki, Greece, pp.383-387, **(2019)**

[8] H. Ahmed, I. Traore, and S. Saad, “Detection of online fake news using n-gram analysis and machine learning techniques,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), **(2017)**

[9] Google Trends. 2020. Google Trends. https://trends.google.com/trends/?geo=US, **(2020)**

[10] Help.twitter.com, About Verified Accounts. https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts, **(2020)**

[11] O. Aborisade and M. Anwar, “Classification for authorship of tweets by comparing logistic regression and naive bayes classifiers,” 2018 IEEE International Conference on Information Reuse and Integration (IRI), Salt Lake City, UT, pp.269-276, **(2018)**

[12] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, “Truth of varying shades: Analyzing language in fake news and political fact-checking,” EMNLP 2017 - Conf. Empir. Methods Nat. Lang. Process. Proc., pp.2931–2937, **(2017)**

[13] G. Cauwenberghs and T. Poggio, “Incremental and decrementai support vector machine learning,” Adv. Neural Inf. Process. Syst., no.x, **(2001)**

Authors

Dr. Jinan Fiaidhi

Dr. Fiaidhi is a full professor of Computer Science and Professional Software Engineer as well as the Grad Coordinator of the PhD program in Biotechnology at Lakehead University. She is an adjunct research professor at the Western University and the editor in chief of IGI Global International Journal of Extreme Automation and Connectivity in Healthcare. She is also the chair of Big Data for eHealth with the IEEE ComSoc. Contact him at sabah.mohammed@lakeheadu.ca.