Detecting Fake News using Machine Learning and Deep Learning Algorithms

Abdullah-All-Tanvir
Dept. of Computer Science and Engineering
East West University
Dhaka, Bangladesh
abdullahalltanvir@gmail.com

Ehesas Mia Mahir
Dept. of Computer Science and Engineering
East West University
Dhaka, Bangladesh
ehesasmm@gmail.com

Saima Akhter
Dept. of Computer Science and Engineering
East West University
Dhaka, Bangladesh
saima.maloncho24@gmail.com

Mohammad Rezwanul Huq
Dept. of Computer Science and Engineering
East West University
Dhaka, Bangladesh
mrhuq@ewubd.edu

***Abstract*— Social media interaction especially the news spreading around the network is a great source of information nowadays. From one's perspective, its negligible exertion, straightforward access, and quick dispersing of information that lead people to look out and eat up news from internet-based life. Twitter being a standout amongst the most well-known ongoing news sources additionally ends up a standout amongst the most dominant news radiating mediums. It is known to cause extensive harm by spreading bits of gossip previously. Online clients are normally vulnerable and will, in general, perceive all that they run over web-based networking media as reliable. Consequently, mechanizing counterfeit news recognition is elementary to keep up hearty online media and informal organization. This paper proposes a model for recognizing forged news messages from twitter posts, by figuring out how to anticipate precision appraisals, in view of computerizing forged news identification in Twitter datasets. Afterwards, we performed a comparison between five well-known Machine Learning algorithms, like Support Vector Machine, Naïve Bayes Method, Logistic Regression and Recurrent Neural Network models, separately to demonstrate the efficiency of the classification performance on the dataset. Our experimental result showed that SVM and Naïve Bayes classifier outperforms the other algorithms.**

Keywords—: Fake News, Twitter, Social Media, Data quality, counterfeit, Machine Learning.

I. INTRODUCTION

An intense inspection the present tweets demonstrates that false news spreads frequently through human than a genuine news does . Lie gets traveled around us quicker, and more extensively than reality in all spheres of information, and the effects were more dangerous and horrifying. There are several kinds of tweets like issues on a government, trending topics around the world, mental abuse, urban legends, occasions in calamities. What's more notifying is that it's not just bots that is outpouring the majority of the misrepresentations, studies claimed. It was some specific individuals performing a large share of this crime. Normally general users, as well, they explained. In this case, verified users and those with numerous fans were not more often the center in spreading misinformation of the corrupted posts. Fake news on social media which got viral like a rocket in no time can cause much havoc to our society human and country.

Measuring precision and validity in news contents are all around well-examined points in every sphere of daily life. The multiplication of huge scale internet-based life information and its expanding use as an essential news cue, be that as it may, is compelling a reconsideration on these specific areas. Strategies in the past that relied upon the arrangements made by the journalists "watchmen" to channel out below average substance are no increasingly proper as engagements into social networking's volume has promptly overshoot our ability to control the standard physically. Rather, stages like Facebook and Twitter have permitted sketchy and incorrect "news" substance to contact broad gatherings of people without survey. Online life clients' inclination toward accepting what their companions share and what they read despite precision enables these fake stories to proliferate broadly through and over numerous platforms. In spite of examination into gossip engendering on Twitter, gossip and unauthenticated news are getting to be progressively tricky. Computational strategies have demonstrated helpful in comparable settings where data volumes overpower human examination abilities. Moreover, regularities in bot conduct and monetarily spurred sentimentalists recommend automated machine learning methodologies could encourage to specify these issues.

This work makes the following contribution. It has manufactured a fake news framework for news shared on Twitter utilizing five unique calculations. We consider every explanation which is posted by the clients. We utilize a rundown of highlights for a bit of news (a URL after canonicalization) the individual clients who shared the URL by tweeting or retweeting it, close by with the words that show up in the title of the news article itself. Five different types of classification models including Support Vector Machine (SVM), Naïve Bayes, Logistic Regression, Long short-term memory, Recurrent Neural Network are implemented for this task. A combination of these classification models are also tested to further enhance the accuracy of prediction. Using sci- kit learn [10], these models are implemented to prepare from the training dataset using k-fold (k=2) cross-validation, and then predict using the data set.

II. RELATED WORK AND DATASETS

"Fake" news detection on twitter has been initially researched by many authors in the past. But during the presidential election of US, this issue became more popular and everyone was give their best to find out some better solutions for this classification. The previous history was briefed on "The Atlantic" news [1]. Different papers and journals explained the core of the problem and its sections [2][3]. Some proposed data mining approaches [4], some uses the Stance Detection method by trained the machine using stances [5]. Some authors also used a machine learning algorithm implemented as a software system for detection [6]. We mainly focused on a paper [7] which proposed a methodology for characterizing trending twitter threads.

978-1-7281-1557-3/19/$31.00 ©2019 IEEE

In the modern timeline, anoutstanding science news site "SCIENCE" distributed an article where they ordered twitter news as evident or fictitious utilizing data from six free reality checking associations that showed 95 to 98% concession to the orders. [8]

Out of all the studies in this specific area, one study proposed a survey on the field of NLP for false news detection. This paper displays an overview of fake news recognition. Their review presents the difficulties of the methodology in the detection of news or twitters like this. They deliberately audit the data and NLP arrangements that have been produced for this whole study. They likewise examine the points of confinement of those twitters and issue definitions, our bits of knowledge, and suggested arrangements. [9]

One paper used the "Crowdsourcing a large corpus of clickbait" to address the urging task of clickbait detection, they constructed a new corpus of 38,517 annotated Twitter tweets, the Web is Clickbait Corpus 2017. [11].

A. Chile earthquake 2010 Datasets

As we already told that we mainly focused on a paper which worked with Twitter threads. Specifically, this data includes the information about the Chile earthquake 2010 and an earlier inspection into some popular threads resulted in this output [10], where Twitter played a significant role both in coordination and misinformation. We used this dataset for our proposed model. We collected this dataset by sending a mail to the Cody Buntain, who is a Postdoctoral fellow in University of Maryland and by agreeing some terms and conditions he mailed us back the dataset for use. There are 20,360 data in the dataset. This dataset includes the Label class as H and N where 'H' stands for harassment indicating Fake and 'N' stands for non-harassment indicating Real or Fact. The dataset also includes one column for statement where all the Twitter threads or twits including the links and mentions.

III. METHODOLOGY

From the above discussion we already came to know that there are two classes in a news, whether it's real or fake one. To classify a news, we need to understand the problem definition first, then we go for our model and evaluate the result. Machine Learning is replete with its algorithms but some of them are really good for "Fact or Fiction" detection and some are on an average scale.

Our main focus was on the feature engineering that if we could tune-up the features or add some other features the accuracy of detecting news can be much efficient. From the idea of psychological research on false news, we find out the word lengths in a tweet statement can be a great feature as unauthenticated news content a lot of title, words and fictional statements. So, we added a new feature word length which is actually the count of words in a tweeted statement without any links, date or any indications.

A. Algorithms

In our model, we used 5 different types of machine learning algorithms and for the implementation work, we used Python 3.6.5 as our programmable language. The classification models that we implemented using the above- mentioned dataset are Bayesian Model, Logistic Regression & also Support Vector Machine - two most famous deep learning methods RNN Recurrent Neural Network and Long Short-Term Memory were also implemented to see how well our data fit into the model. These algorithms are good for different classifications and they got their own properties and performance based on different datasets. As we said earlier Naïve Bayes, Logistic Regression and SVM are more commonly used algorithms for classification problems.

B. Preparing The Datasets

To apply these algorithms, first, we need to process our dataset. We used Pandas, Numpy, Scikit-learn, keras library, and visualization, we use matplotlib. As we are dealing with text data, we will implement the following different ideas in order to process the dataset. Those are as follows,

- ✓ Count Vectors
- ✓ TF-IDF
  - Word Level
  - N-gram Level
  - Character Level
- ✓ Word Embedding

Count Vectors: Count Vector represents a notation in the form of a matrix data set matrix notation in which corpus document is represented by each row, each column represents a corpus term, and each cell represents the frequency count of a particular term in a particular document.

TF-IDF: TF-IDF represents how frequent a term is in an entire document. It tries to assign a metric value to represent the presence of that term. This is widely and frequently utilized in text mining. This weight is a factual measure used to assess how essential a word is to a report in a gathering or corpus.

$$TF(t) = \frac{Number\ of\ times\ term\ t\ appears\ in\ a\ document}{Total\ number\ of\ terms\ in\ the\ document}$$

$$IDF(t) = log_e^{\frac{Total\ Number\ of\ documents}{Number\ of\ dcument\ with\ term\ t\ in\ it}}$$

Based on several types of input tokens like words, characters, n-grams, TF-IDF Vectors can be generated

- Word Level TF-IDF: TF-IDF value of each term represented in a matrix format.
- N-gram Level TF-IDF: Combination of N termsis represented by N gram level. This matrix depicts TF-IDF N-gram scores
- Character Level TF-IDF: TF-IDF values of the n-grams character level in corpus represented in a matrix.

Word Embedding: A word embedding is a form of representation of words and documents with a dense representation of vectors. The text learns the where the word is located in the vector space and is based on the words used around the word. There are four key steps:

- Embedding of the pretrained word
- Create an object tokenizer
- Transform text documents into tokens ' sequence and pad them
- Create a token mapping and its embedding

We split our dataset into 60% train data and 20% test data including 20% validation LSTM, RNN model and we use k-fold cross validation for our other models where k = 2.

In the LSTM, as it is a part of the neural network, we used the embedding layer, hidden layer, an input layer and output layer finally. We used 'Sigmoid' activation, 'glorot_normal' kernel initializers, 'categorical_crossentropy' loss and 'RMSprop' as an optimizer.

RMSprop: The RMSprop optimizer is similar to the downward momentum gradient algorithm. The RMSprop optimizer limits vertically oscillations. Learning rate can be increased and to converge in the horizontal direction, any algorithm can take more steps more quickly. Beta indicates the momentum value and is usually set to 0.9.

$$v_{dw} = \beta . v_{dw} + (1 - \beta) . dw \quad (1)$$

$$v_{db} = \beta . v_{dw} + (1 - \beta) . db \quad (2)$$

$$W = W - \alpha . v_{dw} \quad (3)$$

$$b = b - \alpha . v_{db} \quad (4)$$

Sigmoid activation: The Sigmoid curve looks like a function in S shape as shown in Figure 1. Sigmoid is used because it exists between 0 and 1. It is therefore particularly used for models in which the probability as an output must be predicted. Since there's only a chance of anything from 0 to 1, Sigmoid's the right decision.

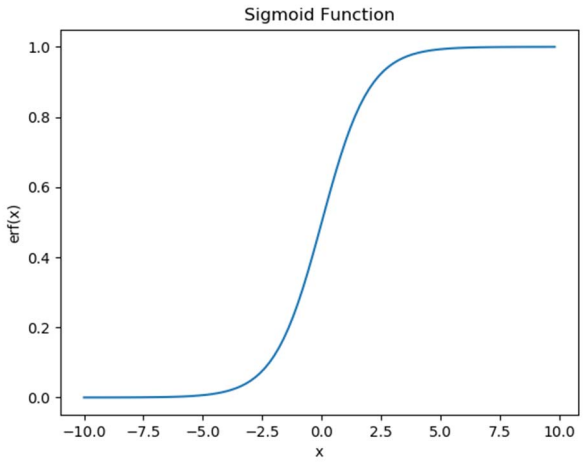

Figure 1: Sigmoid Function

IV. Implementation

We have implemented inter and intra comparison on five different classification model. Based on the comparison we have also considered the precision and recall value of the models. A combination of five different models is tested in accuracy using N-Gram and Character level vectors in this order to improve the accuracy of prediction. Using scikit learn these models are implemented to learn from the training data using k-fold (k=2) cross-validation, and then predict using the data set. Then we have evaluated the performance of the models.

A. Inter and Intra Comparison of the Classifiers

The inter-class and intra-class cluster similarity is a crucial part of Classifying data. The intra-class cluster shows the distance between data point with the cluster center, meanwhile, inter-class cluster shows the distance between the data point of one cluster with the other data point in another cluster.

As mentioned earlier, several features were chosen for observing the performance using different supervised and deep learning methods. There are basically four feature vectors that was extracted from our text dataset, which are

- Count Vector
- Word Level Vectors
- N-gram vectors
- Character level vectors

V. Results

Firstly, Naïve Bayes Model was tested on each of the feature vectors mentioned above. It gives 73% accuracy on count vector feature, 75% on Word Level TF-IDF, N - gram vector, and character vector as well.

Then Logistic regression model was performed. This time the performance was slightly optimized than before, predicting 74% and 76% respectively on count and word level vectors. Whereas 75% and 76% was the results of Logistic Regression stage.

Thirdly Support Vector Machine was performed for observing any further enhancement over the previous result so far. But no enhancements were observed at this stage as testing accuracy from SVM is 74% in all the four mentioned features vectors in our study. Figure 2 shows the above mentioned findings.

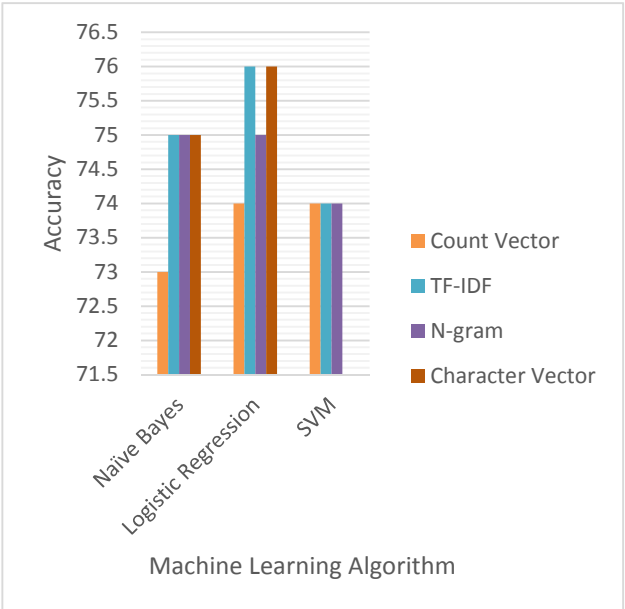

Figure 2: Accuracy Comparison

Figure 3 shows the result after performing cross-validation. We achieved 62.47% for Logistic Regression, 84.56% for Naïve Bayes and 89.34% for SVM in Count Vector feature. And for TF-IDF, we got 69.47% for Logistic Regression, 89.06% for Naïve Bayes and 89.34% for SVM.

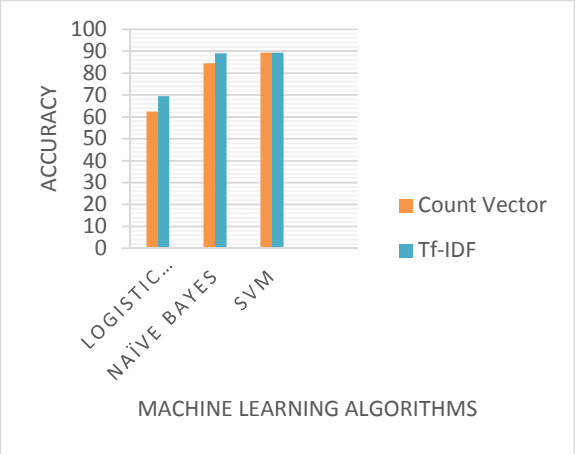

Figure 3: Accuracy Comparison After Cross Validation

After testing the data with several supervised methods, deep learning model was chosen as it always gives an efficient output layer based on the weight of input layers. For this task of classifying fake news, Recurrent Neural Network and Long Short-Term Memory Model was taken as a method at this point.

After feeding the Chile Earthquake Dataset 2010 into the RNN model, we achieved 74% accuracy which is considered the same as what we got in supervised methods and by using LSTM, we achieved 76% accuracy as reported in Table 1.

Our proposed model works fine in case of the huge dataset as it deals with 20360 data, also it takes a comparatively shorter time than what it would take before.

Table 1: Accuracy Comparison After Cross Validation

Classifiers	ACCURACY	
	Count Vector	TF-IDF
Logistic Regression	62.47	69.47
Naive Bayes	84.56	89.06
SVM	89.34	89.34

Table 2: Accuracy achieved for Deep Learning Algorithms

Classifiers	ACCURACY	
	Default	With Kernel initialization
Long Short-Term Memory (LSTM)	73%	76%
Recurrent Neural Network (RNN)	74%	73%

A. Comparison of the Inter Accuracy Prediction:

Figure 4 shows the overall comparison among all algorithms. SVM Shows the highest accuracy with TF-IDF feature among all the classifier we have implemented in our dataset. Logistic regression attempts to predict outcomes based on a set of independent variables, but logistic models are vulnerable to overconfidence. It requires that each data point be independent of all other data points. In our dataset, the feature word length depends on the news statement. If observations are related to one another, then the model will tend to overweight the significance of those observations.

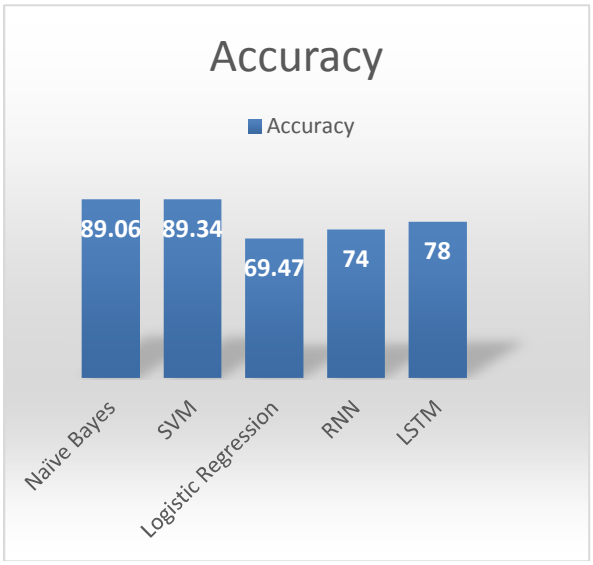

Figure 4: Overall Accuracy Comparison of Machine Learning and Deep Learning Algorithms

Table 3 shows the precision, recall and F1-score for machine learning algorithms such as Naïve Bayes, Logistic Regression and SVM. All three algorithms have same level of precision. SVM outperforms the other algorithms considering recall. F1-score of both Naïve Bayes and SVM is 0.94 which is the highest here

Precision:
Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Recall:
Recall is the ratio of correctly predicted positive observations to the all observations in actual class - yes.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

F1 Score:
F1 Score is the weighted average of Precision and Recall.

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

Table 3: Precision, Recall and F1-Score for Naïve Bayes, Logistic Regression, SVM

	Precision	Recall	F1-Score
Naïve Bayes	0.89	0.99	0.94
Logistic Regression	0.89	0.75	0.81
SVM	0.89	1.0	0.94

V. CONCLUSION

In this paper, we analyzed a computerized model for checking the verification of news extracted from Twitter which gives general answers for information accumulation and expository demonstration towards fake news recognition. After having an idea from the supervised models, a deep learning-based model is proposed to identify fake news.

The accuracy metric presumably would be altogether improved by methods for utilizing progressively complex model. It is worth noting, that even with the given dataset, only part of the information was used. The current project did not include domain knowledge related features, such as entity-relationships. Future studies could extract name entities from each pair of news headline and news body and analyze their relationships through a knowledge base.

The study demonstrated that even the very basic algorithms on fields like AI and Machine Learning may find a decent outcome on such a critical issue as the spread of fake news issues worldwide. Accordingly, the aftereffects of this examination propose much more, that systems like this might come very much handy and be effectively used to handle this critical issue.

This work exhibits a programmed model for identifying fake news in well-known Twitter strings. Such a model could be important to a huge number of social media users by expanding their own credibility decisions.

The dataset in this examination is relied upon to be utilized for arrangements which utilized machine learning based statistical calculations, for example, Support Vector Machines (SVM), Naive Bayes (NB), Recurrent Neural Network (RNN), Logistic Regression (LR), Long Short Term Memory (LSTM). In this investigation, SVM performs best for characterization technique.

REFERENCES

[1] Meyer. The Grim Conclusions of the Largest-Ever Study of Fake News.

[2] Tandoc Jr. Lim. Ling, (2017, August 30). Defining "Fake News" A typology of scholarly definitions.

[3] D Horne. Sibel Adali. Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News, Benjamin.

[4] Shuy. Silva. Wangy.Tang. Huan, Fake News Detection on Social Media: A Data Mining Perspective.

[5] Machine Learning (CS229), Fake News Stance Detection.

[6] Sharf. Zareen. Saeed. Twitter News Credibility Meter

[7] Cody Buntain, 2017 IEEE International Conference on Smart Cloud. Automatically Identifying Fake News in Popular Twitter Threads

[8] Soroush Vosoughi, Roy, Aral, (March 2018), The spread of true and false news online

[9] Oshikawa.Qian.Wang, (November 2018), A Survey on Natural Language Processing for Fake News Detection

[10] Golbeck, Jennifer, et al. "A Large Labeled Corpus for Online Harassment Research" Proceeding of the 2017 ACM on Web Science Conference. ACM,2017

[11] Post, Gollub. Komolossy. 2017, Crowdsourcing a Large Corpus of Clickbait on Twitter.

[12] S. R. Maier, "Accuracy Matters: A Cross-Market Assessment of Newspaper Error and Credibility," Journalism & Mass Communication Quarterly, vol. 82, no. 3, pp. 533–551, 20

[13] B. Liu, J. D. Fraustino, and Y. Jin, "Social Media Use during Disasters: A Nationally Representative Field Experiment," College Park, MD, Tech. Rep., 2013.

[14] K. Starbird, J. Maddock, M. Orand, P. Achterman, and R. M. Mason, "Rumors, False Flags, and Digital Vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing," conference 2014 Proceedings, pp. 654–662, 2014.