Fake News Detection Using Deep Learning Techniques

Chaitra K Hiramath [1]
M.Tech in CSE
KLS Gogte Institute of Technology,
Belagavi,Karnataka,India
chaitrakh1992@gmail.com

Prof. G.C Deshpande [2]
Assistant Professor
KLS Gogte Institute of Technology,
Belagavi,Karnataka,India
gcdeshpande@git.edu

***Abstract*— News is crucial part of our life. In day to day life current news are helpful to enhance knowledge what happen around the world. So most of peoples prefer watching news most of the peoples generally prefer reading newspaper early in the morning enjoying with cup of tea. If news is fake that will mislead peoples sometimes fake news utilized to spread rumors about things or it will affect some political leader positions just because of fake news. So it's crucial to find the fake news. So we proposed system to detect fake news but now a day's data on web or social media is increasing vastly and it is so hectic to detect news is fake or not by looking all data and it is time consuming so we utilize classification techniques to classify huge data. Here we proposed fake news detection system based on classification such as Logistic regression (LR), Naïve bayes (NB), Support vector machine (SVM), Random forest (RF) and deep neural network (DNN). We compare all machine learning techniques for detecting fake news.**

Keywords: Fake news, Logistic regression, Support vector machine, Naïve Bayes algorithm, Random Forest algorithm, Deep neural network.

I. INTRODUCTION

The term of false news is regularly connected with misdirection, gossip, fraud, deception and so on. Related work may be for gossip, rumor, fraud and fabrication. Issues identified with such subjects are frequently been seen depending on classification. Likewise, a large portion of printed matters have seen false information location as a binary classification issue. Some organization uses numerous deep learning techniques on databases made out of false information articles and genuine information articles mined from media news database and saw classifiers is great to classify huge data.

There are various classifiers existed. From LR, RF and NB to DNN. Classification categorized into types LR, NB, SVM, RF and DNN. These are most commonly utilized classifiers. Classification is termed as "Supervised learning". Linear Classifiers are LR, NB, SVM, Decision Trees, RF, Neural Networks (NN) are classifiers in Machine Learning (ML).

Bits of gossip or spam recognition are considered issue in Natural Language Processing (NLP) & specialists discovered these issues broadly. Issue of recognizing or finding false information in regular day to day existence, although very much similar to deception finding, but it is critical to detect due to the news body frequently comprise a less and small proclamations.

Everyday access of news sources, for example, web based life channels, news sites, and online papers have created demanding to verifying reliable news sources due to enhancement of misleading information. We center around the ID of phony contents or articles in news sites. First, we present database for the phony news discovery task, using numerous news spaces and depict the accumulation, explanation, and approval process in detail and present a few exploratory examinations on the acknowledgment of etymological varieties in phony and real news content.

Now a day's fake news identification has gained a developing enthusiasm from the overall population and analysts as the spreading of deception online increments, especially in news sources like internet based life channels, news web journals, and online papers.

The section I gives introduction of fake news detection system utilizing RF, NB, SVM, DNN, and LR classification techniques. Section II introduces the writing survey of present frameworks and proposed framework design is given in Section III subtleties area IV givess test examination, results and talk of new framework. Area V finishes up our proposed framework.

II. LITERATURE SURVEY

Rumor play a complex part in clashes of numerous types such as intergroup and global [1]. Their vitality is considered about the way that in current day, the society ability of clashing gatherings to influence each other are very constrained with administrative cutoff points and famous assessment [2].

Features of society engage improvement of significance of those methodologies for affection that are considered about difference in greater part's musings in regards to battle toward

a way that is gainful for one of conflicting get-togethers. Such changes are possible on account of exceptionally picked data spreading first through the expansive communications and through channels of easygoing correspondence. Bits of gossip in this case raises genuine weapon in clashes [3].

V. Subrahmanian et al [4] C. Shao et al. [5] says Because of the promotion of artificial intelligence (AI) and related territories of intellectual processing, the number of bots has detonated all through the system. In this segment, author will investigate their job in the gossip rumors and misinformation spreading.

William Y Yang gives example, there can be sure misrepresented or deceiving data joined to a genuine articles or news. In this way, the whole news or article cannot be accepted as totally obvious or cannot be disposed of as completely false. This issue was tended to in [6] where it is presented Liar dataset including a generous size of small political articles distinctive category comments deciding measure of false substance of every declaration.

Aswini Thota et al [7] introduced automated detection system for fake news detection he said is difficult work to complete as it needs representation to recognize nuances in natural language. To address the gaps between related and unrelated news, they introduce neural network architecture to exactly guess the posture that is to be provided headline and article body.

Wang gives the LIAR dataset [8], consisting of statements made by public figures, annotated with its veracity, extracted from the site polifact.com for research focused on rumors, there is the PHEME dataset, by Zubiaga et al. [10]. This dataset groups a number of tweets in rumor threads, and associate them with news events.

N. J. Conroy et al [9] enroll numerous methodologies that appear to be encouraging close to goal of effectively arranging deluding news. Additionally, noted basic substance related n grams and shallow grammatical feature known as POS labeling have shown not sufficient for the grouping task.

Veronica Perez-Rosas et al. [11] use dataset to detect fake news from different domains. Generally they obtained news from an assortment of standard news sites in the US, for example, CNN, ABCNews, USAToday, FoxNews, NewYorkTimes, Bloomberg, and CNET.

M. Granik et al. [12] present naïve bayes to detect fake news here they utilizes Facebook news post dataset also he check accuracy of the system and it gives 74% accuracy.

In the other hand WhatsApp in Brazil [13] constrains the number of messages with the same content that can be shared by the same user, is using a AI to find abuses and harass messages and like Facebook using third-party agencies (TPA) to check and classify news. Also, the WhatsApp team trained and showed the capabilities of their app to the current president candidates and their communication team in an attempt to avoid possible use of the app for fake news spread.

III. SYSTEM ARCHITECTURE

A. System Architeure

At first we get dataset as news or articles list form. At that point on that dataset processing is done for feature mining and determination. Data should be pre-processing. It incorporates expulsion of punctuations, URL's, images, stemming and stop words. At that point NLP processing for extricating highlights and dependent on that training file is made and afterward we classify that information utilizing classifiers, for example, LR, SVM, NB, RF and DNN. News Dataset: Dataset of news is taken from online.

Processing: Content information needs processing to execute AI on them. There are kinds of methods generally utilized to change over content information into a structure that is prepared for demonstrating. The information processing steps that are applied on headlines and articles.

Stemming and stopwords removal: Stemming technique utilized to detach suffixes or prefixes from a word. And Stops Words may be filtered and processed from content because are increasingly normal and hold less significant data. Stop words generally utilized as a linking element of the sentences, for instance, “and”, “or”, “but”, “of”, “in”, “from”, “to”, “a”, “an”, and “the” etc. This kind of stop words that are of not vital and it can utilize important processing time, since to take away stop words is crucial task.

Natural Language Processing: there is huge amount of information stored in file but it is inaccessible to computer assisted analysis NLP allows to analysts to find crucial information.

Feature Extraction: Here important features are extracted to generate training file.

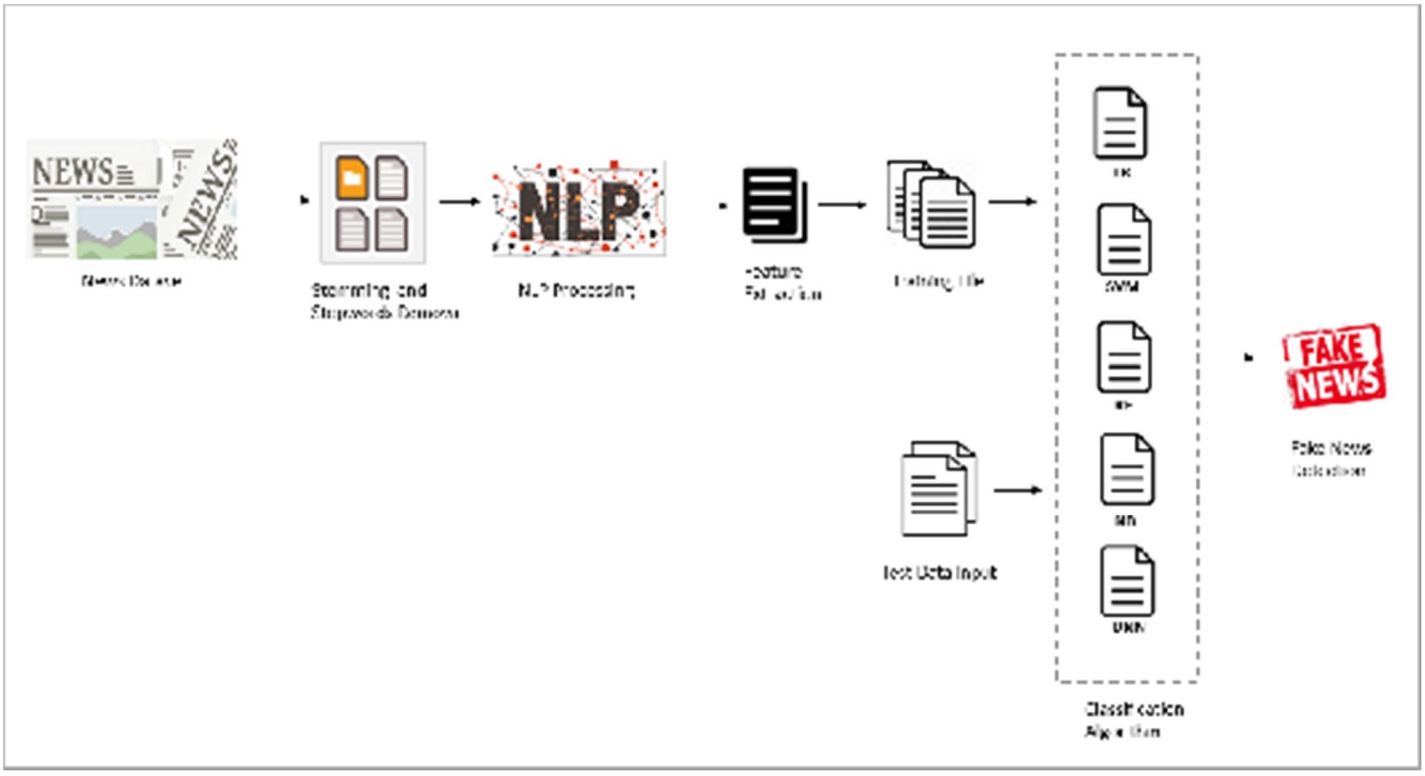

Fig 1. System Architecture

B. Algorithm

Deep Neural Network

Back propagation Method
{I1, I2,…., In} are given Inputs

Where,
Ii is Input Layer input in, and i=1, 2, n.
J is hidden layer where Sigmoid Transfer function is utilized to estimate output of every neuron in hidden layer.
O is output layer.
and are weights for hidden and output layer.

Sigmoid transfer function is as follows:

The steps are as follow:
1: Execute network forward with input records to obtain network yield.

2: Calculate error value.

3: Error signal vectors and of both layers are calculated. Vector is for output layer, is for hidden layer.

4: Adjust output layer weights.

5: Adjust hidden layer weights.

6: Go to step 1

7: The training cycle is completed.

C. Experimental Setup

I. Framework expands on Java system. For advancement, the NetBeans (adaptation 8.1) instrument is used MySQL as backend. Classification algorithms along with numerous extraction techniques. Framework doesn't require a specific equipment to run any standard machine can run the application.

II. Database Description: News dataset is downloaded from web it consists of news from different sites.

D. Comparison Results

This part represents performance of LR, RF, SVM NB and DNN algorithms. Figure 2 exhibit Memory Comparison of five classification algorithms for numerous Threshold. X-axis represent Algorithms & Y-axis present Memory in bytes. According to graph DNN uses more memory that NB and NB uses more than SVM also memory required for LR is less than RF and SVM.

Table 1 shows memory comparison in tabular form LR requires less memory than other algorithms.

TABLE I. MEMORY COMPARISON

SR No.	Algorithm	Memory in bytes
1	LR	920000000
2	RF	950000000
3	SVM	1100000000
4	NB	1100000000
5	DNN	1120000000

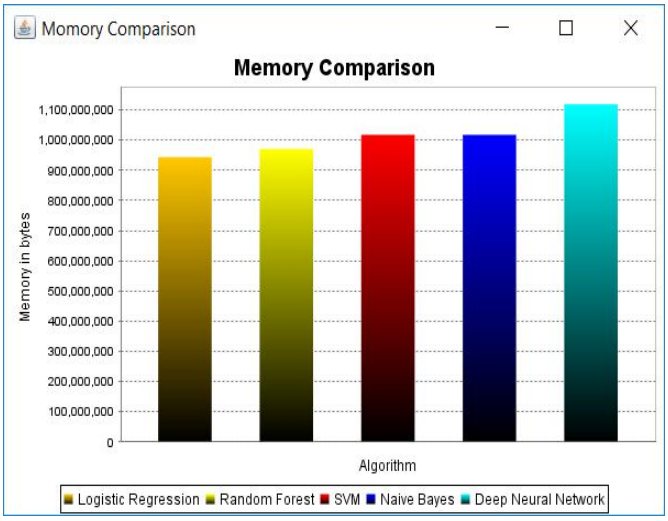

Fig. 2: Memory Comparison Graph

Figure 2 depicts comparison of LR, RF, SVM NB and DNN classification algorithms for different byte value. X-axis present algorithms and Y- axis present memory in bytes. LR algorithm utilizes less memory compared to all other methods.

Table 2 shows Time comparison table where DNN requires less time than other algorithms.

TABLE II. TIME COMPARISON

SR No.	Algorithm	Time in ms
1	LR	3750
2	RF	1800
3	SVM	2800
4	NB	1900
5	DNN	400

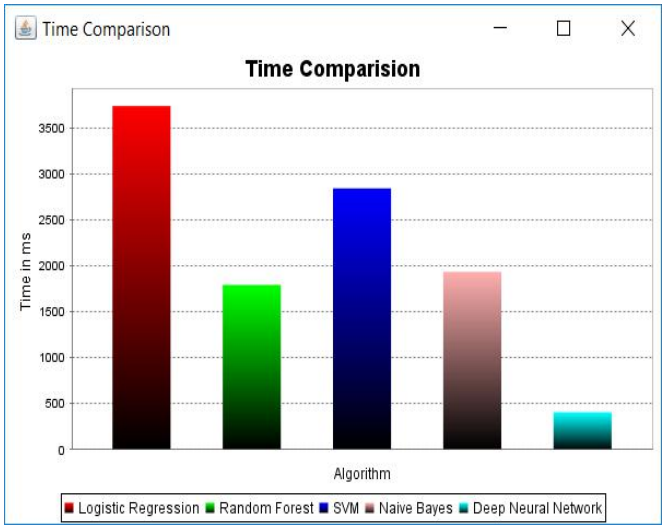

Fig. 3: Time Comparison Graph

Figure 3 depict Time comparison of LR, RF, SVM NB and DNN classification algorithms for different time in ms. X-axis present algorithms and Y- axis present Time in ms. DNN requires less time than other classification algorithms. DNN algorithm requires 400ms which is much less than all other methods.

Table 3 shows accuracy comparison of algorithms where accuracy of DNN is greater than other four algorithms.

TABLE III. ACCURACY COMPARISON

SR No.	Algorithm	Accuracy in %
1	LR	75
2	RF	77
3	SVM	79
4	NB	89
5	DNN	91

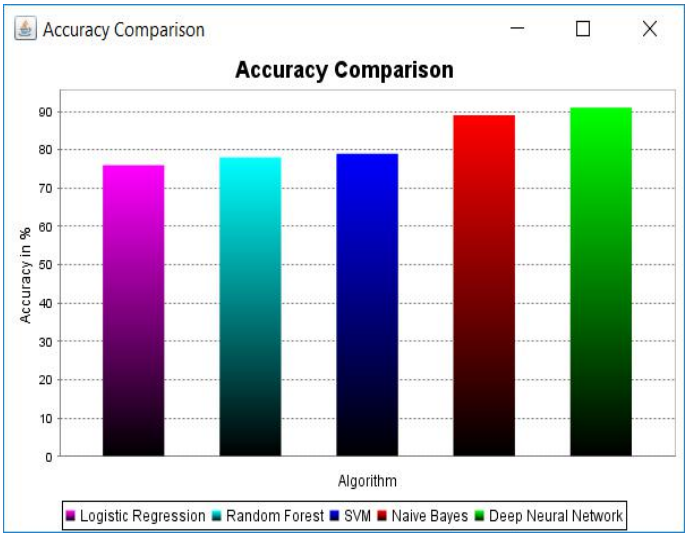

Fig. 4: Accuracy Comparison Graph

Figure 4 represent the Accuracy comparison of LR, RF, SVM NB and DNN classification algorithms for different percentage. The X-axis display algorithms and Y- axis display accuracy %. Accuracy percentage of DNN is more than rest of other classifiers.

IV. CONCLUSION

Here we, present different algorithms for classifying statements made by public figures were implemented. In proposed system LR, RF, SVM NB and DNN classification techniques are utilized that will help to detect fake news. Classification techniques like LR, RF, SVM NB and DNN for feature selection and extraction utilized, DNN will work fine in execution time and accuracy cases but it needs large memory than other. Then we compare NB, RF, SVM, LR and DNN on basis in terms of time and memory and accuracy, according to comparison results it exhibit that DNN Algorithm is improved than rest algorithm in accuracy and time kind because rest classifiers requires more time and gives less accuracy hence DNN is more crucial to detect the fake news.

REFERENCES

[1] Arif, A., Robinson, J. J., Stanek, S. A., Fichet, E. S., Townsend, P., Worku, Z., &Starbird, K. (2017, February). A Closer Look at the Self-Correcting Crowd: Examining Corrections in Online Rumors. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (pp.155-168). ACM.

[2] Louni, A., Santhanakrishnan, A., &Subbalakshmi, K. P (2015). Identification of source of rumors in social networks with incomplete information. arXiv preprint arXiv:1509.00557.

[3] Pasquini, C., Brunetta, C., Vinci, A. F., Conotter, V., &Boato, G. (2015, June). Towards the verification of image integrity in online news. In Multimedia & Expo Workshops (ICMEW), 2015 IEEE International Conference on (pp. 1-6). IEEE.

[4] V. Subrahmanian, A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, A. Flammini, and F. Menczer, "The DARPA Twitter Bot Challenge," Computer, vol. 49, pp. 38–46, June 2016.

[5] C. Shao, G. L. Ciampaglia, O. Varol, A. Flammini, and F. Menczer, "The spread of fake news by social bots," arXiv preprint arXiv:1707.07592, 2017.

[6] William Yang Wang." Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. ACL 2017. arXiv:1705.00648

[7] Aswini Thota1, Priyanka Tilak1, Simeratjeet Ahluwalia1, Nibhrat Lohia1 ," Fake News Detection: A Deep Learning Approach " SMU Data Science Review, Vol. 1 [2018], No. 3, Art. 10

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

[9] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," Proceedings of the Association for Information.

[10] A. Zubiaga, M. Liakata, R. Procter, G. W. S. Hoi, and P. Tolmie, "Analysing how people orient to and spread rumours in social media by looking at conversation althreads,"PloS one, vol. 11, no. 3, p. e0150989, 2016.

[11] Ver´onica P´erez-Rosas, Bennett Kleinberg, Alexandra Lefevre ,Rada Mihalcea, "Automatic Detection of Fake News" Proceedings of the 27th International Conference on Computational Linguistics, pages 3391–3401 Santa Fe, New Mexico, USA, August 20-26, 2018.

[12] M. Granik, V. Mesyura, "Fake news detection using naive Bayes classifier," 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kiev, 2017, pp. 900-903.

[13] B. Capelas, "Whatsapp anuncia planos para tentarcombater 'fake news' no brasil,"Estado

[14] Cade Metz. (2016, Dec. 16). "The bittersweet sweepstakes to build an AI that destroys fake news". Available: https://www.wired.com/2016/12/bittersweet-sweepstakes-build aidestroys-fake-news/

[15] Fake news RAMP: classify statements of public figures. (n.d.) [Online]. Available: https://www.ramp.studio/problems/fake_news

[16] The Principles of the Truth-O-Meter: PolitiFact's methodology for independent fact-checking. (2018, Feb. 12) http://www.politifact.com/truth-o meter/article/2018/feb/12/principlestruth- o-meter-politifacts-methodology-i/. Accessed Mar. 24, 2018.