Bangla Fake News Detection

Md. Nahidul Islam
Department of computer science and engineering
Daffodil International University, Bangladesh
nahidul15-12448@diu.edu.bd

Md. Hanzala Ahmed Hadi
Department of computer science and engineering
Daffodil International University, Bangladesh
hanzala15-12404@diu.edu.bd

***Abstract*—Fake news has been coming into sight in significant numbers for numerous business and political reasons and has become frequent in the online world. Because in this recent time smartphone is everywhere and a person can do anything with it. Some fraud people sometimes spread a fake news. They represent it as it is authentic. Significant number of research has been published on Bangla fake news. We also tried to detect the fake bangla news in most accurate way. We have used 3 classifier in this paper and got good result. Our 3 classifier shows respectively 77Percentage 55Percentage and 60Percentage accuracy . we expected more but data was little unbalanced..**

***Index Terms*—Fake news, Bangla Fake News, Classifier, Accuracy**

I. Introduction

An article that can potentially mislead or deceive readers by providing wrong information are known as fake news. Usually fake news are written and published with the intention to damage the reputation of an agency, entity, or person. Sometimes people share a non based news in their social media but they don't know how it will impact in the real world how people will take it. The popularity of social media , easy access to online advertisement revenue, increased political divergence have been the reasons for the spread of fake news. Fake news websites and channels use their fake news content with the intention to misguide reader of the content and spread misinformation through social networks and if its go viral then it's spread world wide. It has bad impacts on millions of people and the surrounding environment. Now a days developer have published some website where we can check that a news is fake or not but uploading a news. But it works manually. Some conscious people don't read news without authorised web site or who published. Because they don't trust unauthorised web site, journal, page news because that might be fake. But there are still a huge amount of people use smartphones , computer or laptop they don't know which site , journal, page news if fake. When they read any news online they don't even check the news properly what is that's news source. They read a news and spread it, this is how news goes viral. Without knowing the news is even fake or not. Based on our research we found no highly computer base approach to handle this things. Because it's a risky thing. It's can spread a negative impact in our society. We have collected some data from internet and applied different algorithms to check which algorithm can give a better accuracy to predict whether a news is or not. In this paper we have used a data-set of 3 levels 0,1,2 and applied Logistic regression,DecisionTreeClassifier And RandomForestClassifier algorithm

Identify applicable funding agency here. If none, delete this.

II. Literature Review

Farzana Islam , Mohammad Minhazul Alam collected data and proposed Naive Bays and Random Forest classifier to detect fake news. They split the dataset into 75:25 ratio. Where 75 Percentage are training data and 25 Percentage are testing data. In Naïve Bays they got highest 52 Percentage accuracy and in Random Forest they got 85 Percentage accuracy, where error was 18.01 Percentage.

Md Gulzar Hussain , Md. Rashidul Hasa et al. took a dataset with 60.92 Percentage real data and 39.08 Percentage fake data and splits the dataset into 70 Percentage training and 30 Percentage test data. They have applied Naïve Bays and Support Vector Machine Classifier. In Naïve Bays the got highest 93 Percentage accuracy and in SVM they got over 97 Percentage accuracy.

Md Zobaer Hossain and Md ashraful Rahman used 4 dataset where one dataset contains fake news another ones contains authentic news other two contains the mix news. Where almost 48000 data is real and 1300 data is fake. In the case of news embedding SVM ,LR scores 46 Percentage,53 Percentage. They had applied SVM and got the accuracy of 91 Percentage.

Arnab Sen Sharma , Maruf Ahmed Mridul and Md Saiful Islam have ued CNN. First they split the data into two parts . 70 Percentage of the data was training data and 30 Percentage data was test data. After applying CNN they got the accuracy of 96.4 Percentage.

Some approaches simply used a naïve bays classifier to classify the news . After a little bit processing on grams from the news context , words are fed to a naïve bayes classifier .

Ruchansky , et al. proposed a Capture Score and Integrate model . They took words from twitter and weibo and the applied Naïve bays on it and got the accuracy 95.Percentage.

Shafayat Bin Shabbir Mugdha, Marian Mohammad et al. thy first tokenize the data and then applied SMV, LR, RF . SMV provide the accuracy of 54.12 Percentage . LR gives about 69.53 Percentage and RF gave about 63.14 Percentage accuracy.

III. Result

we have used 30 Percentage test data to predict out fake news with the use of 70 Percentage train data from our dataset.

We applied several classifire to predict the best accuracy for the taken dataset. As for Logistic Regression the train test data accuracy is 77 Percentage . For Decission Tree Classifier the test accuracy is 55 Percentage And RandomForest Classifier provied the accuracy of 60 Percentage.

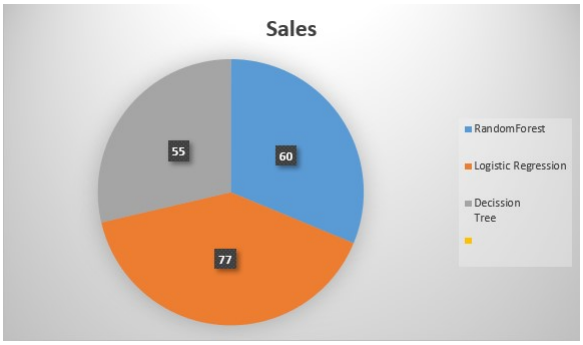

Fig. 1. For this dataset the best accuracy we get is 77 Percentage which is Logistic Regression and the worst accuracy 55 Percentage is provided Decision Tree Classifier.

IV. Conclusion

We took a labeled dataset. In this paper we have used three classifier Logistic Regression, Random Forest and Decision Tree Classifier. Logistic Regression provides highest 77Percentage accuracy. Here we worked with sentence level based but we think if we it will be more specific if we works with word level. If we apply more classifier then it will helpful to take which classifier is most accurate. And which Classifier is best for fake news detection.

References

[1] J. Frost, The Importance of Statistics - Statistics By Jim. Retrieved 4 January 2020, from https://statisticsbyjim.com/basics/importancestatistics/'

[2] One Sample T-Test - Statistics Solutions. (2020). Retrieved 4 January 2020, from https://www.statisticssolutions.com/manova-analysis-onesample-t-test/

[3] Ahmad, A. and Amin, M. R. (2016). Bengali word embeddings and it's application in solving document classification problem. In 2016 19th International Conference on Computer and Information Technology (ICCIT), pages 425–430. IEEE.

[4] R. K. Nielsen, R. Fletcher, N. Newman, J. S. Brennen, and P. N. Howard, "Navigating the 'infodemic': How people in six countries access and rate news and information about coronavirus," Apr 2020.

[5] De Sarkar, Sohan, Fan Yang, and Arjun Mukherjee. "Attending sentences to detect satirical fake news." In Proceedings of the 27th International