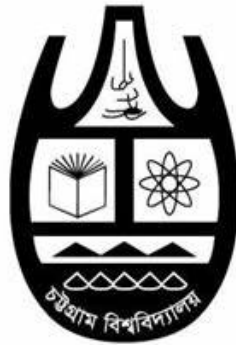# Fake News Detection in Bangla Language

A proposal submitted in partial fulfillment of the requirements for the degree of

Bachelor of Science (B.Sc.) in
Computer Science and Engineering

to

Department of Computer Science and Engineering
University of Chittagong

by

Mosharat Jahan

Student ID : 18701040

Session : 2017-2018

Under the supervision of

Mr. Rezaul Karim

Associate Professor

Department of Computer Science and Engineering

University of Chittagong

November 20, 2022

Abstract

Now is the time of internet where every incident becomes a news and every news is seen all over the world in mere seconds. Which makes the possibility of spreading a false news more prominent. In a country like ours where people tend to be very excited over a simple news, spreading of fake news can be very impactful in many ways. There are numerous examples where a fraud and falsified news has done severe harm to people. So with time many researches are being conducted to detect fake news in different languages. Though most of the researches are conducted regarding high resource languages but the impact of fake news is not confined into language. The objective of this proposed model is to detect fake news from various bangla websites as there isn't many researches conducted regarding this language.

Contents

1 Introduction

Sharing news is like an undeclared competition now a days between people. As ridiculous it may sound but it has become a reality where people judge each other based on how many likes or comments or views they get on their social media profiles. But to get these views and likes some people tend to apply unethical ways. Most common way is to share false or unverified news just to cause periodical excitement in viewers and get desired likes and views and shares as well. It may seem not harmful at first but not every news effects every people. A news which may seem normal to us may be of high significance to someone else. Other than this people use fake news for their own benefit also which is very prominent if we look at the political parties in any country. To get some votes and win elections political parties tends to spread false news about their rival parties or even about themselves to get the attention of their targeted audience.

Not only news but fake advertisements of products or anything else also has become a great threat to peoples belief in online news. Sometimes the news may not be fake but the headline is decorated in a way to attract people that has no relation with the news the headline carries inside. As now a days people tend to rely on social media based news rather than traditional news printed media, the necessity to detect false news in internet has became unavoidable. To detect false news there are some websites like www.politifact.com, www.factcheck.org, www.jaachai.com where the admins manually update false stories published in internet with logical and factual explanation [6]. But this websites can't work in real time and are not capable enough to respond quickly to any news. So many researchers apply different approaches to detect fake news efficiently.

1.1 Background

Fake news is mainly spread with the intention of misleading an incident or a story or a group of targeted people. Researches shows that the news spreading in internet leaves a long term impact on people. Which is why fake news is used to build unnecessary fear in people. Before internet people mostly tend to read news from printed media and the news wasn't so easily reachable. So the ill intended people couldn't use the media to spread false news this easily like now. With the widespread use of internet and social media news are more easily reachable to people now and so the false news too. Every fake news detection study adopts an idea that fits its research topic. In this study, we are sticking to the idea of fake news as "misinformation on the web" [2]. It includes humorous false news, serious fabrications, large-scale hoaxes, rumors, social spam etc. Also, the idea focuses on different fields such as politics, entertainment, society.

1.2 Fake News

News or article that are written with the intention of misleading and deceiving people with falsified or distorted content can be called fake news. The ease of access to internet and media has eventually increased peoples dependency on different internet platforms for any news. A news need not be fully human made for it to be called fake news. Even by changing e single date of a real news also makes the news equally false or just by changing a name of the place an incident took place. Rumors are also defined as fake news which is for an unknown reason very favourite of people now a days. Every single day a new rumor circulates in the internet for gaining some public hype and attention.

1.3 Impacts of Fake News

Not even a decade ago people used to wake up in the morning and first thing they did was to check weather today's news paper is delivered or not. Spotting people in road side tea stalls sipping their morning tea with a news paper in hand was a regular picture. This reminds us how significant news papers are in our daily life to cope up with the world.

But with time this scenario is rare now to even imagine. Now people do first thing in morning is to open their mobile phones and scroll through the news in internet. If we are talking about the impact of news then the impact of fake news also comes into light. Rumors and distorted news have a huge ability to change peoples view and behaviour towards a particular incident.

There are many examples in Bangladesh of different tragic incidents due to rumors. In July 2019, five people were beaten to death and ten injured by mobs as a result of widespread news about the expected human sacrifice in the construction of the Padma Bridge [7]. At the time of 2016 US election, 25 percent of the Americans browsed a fake news website in the period of election which has been hypothesized as one of the issues that influenced the final results [5]. Like this there are vast amount of incidents which shows us the impact of falsified news so detection of fake news is unavoidable in every aspect.

1.4 Motivation

Information and knowledge are just a click away in the world of internet. Even for tiny amount of query people tend to search it up on web. And there are thousands of answers of same question. But we need to remember not all of them are true or acceptable. Many people take advantage of internet and peoples innocence to distract them in a way that is a total waste of time for them. Some have pre decided intention for doing this and some just find it fun. Indeed what a ridiculous way to have

fun! But what can one do, this is the way world is revolving now. People use internet to keep track with the world news. But if you are keeping track of fake news then what is the motive of this total concept of keeping track with world. Most of the researches in this field revolves around English language mostly and to our knowledge there is not so many researches done to detect and tackle fake news spread in Bangla. So to outline and analyse a way to deal with fake news issue, this work is presented.

2 Literature Review

As stated previously there have many attempts to detect fake news from internet by many researchers. And most of the attempts were made regarding content based features. The news content has many parts like image,text etc but as a huge part is consisted by text so main focus always remains on text based fake news detection. But the limitation in this way is not negligible because even though previously the pattern of text in fake news and real news could be separated, now it is not the case. So text based detection in loosing its credibility with time.

To avoid loopholes of text based detection many studies focus on context based features which includes the idea of linguistic features of the news. In this approach user based features and network based features are focused on [12]. Which means the properties of news source and news link are studied to find out if the news is true or not. Just like different features every studies uses different methods for their corresponding work. Most of the previous studies used Support Vector Machine(SVM) [11]. It was intended to show that absurdity,punctuation marks and grammar are best for identifying poignant news by collecting a number of news articles from some news websites. Among all the studies, the research done by Vishwakarma et al. is a noteworthy one. His study proposed URL search based result fake news detection using a set of reliable links. This research shows 85 percent accuracy and which shows the effectiveness of using web search links in detecting fake news. But the idea of the this whole study was not new as it used the concept of white list based phishing detecting [8]. Even after considering the effectiveness of this study there are still some drawbacks of this research which is it needs human participation to collect the set of links.

Another research by Perez Rosa et al. which includes previously mentioned SVM model used only linguistic features of text. They also collected a datadset of 240 legitimate news from different news websites in US and made another dataset consisting the fake version of news from the previous dataset. They used crowd sourcing to generate this fake version of news using Amazon Mechanical Turk(AMT) [9].Moreover as we can see here also human intervention is needed to collect the relevent dataset and also to generate false version of the news collected in the dataset. Moreover to extract more secretive characteristics of false news neural network based model are undoubtedly more reliable [13].

There are also some researches of clickbait detection which also uses linguistic features and neural network with dataset that normally contains some clickbaits from various news websites [3]. Not only English but there are some studies done on a couple of low resource languages like Indonesian using Naive Bayes classifier [10].A noteworthy Research was done in Bangla language using SVM and MNB classifier that performs better than Nave Bayes [7]. Another mentionable work in Bangla fake

news detection was the creation of a first publicly accessible news dataset with the aim of collecting 50k data [6].

A new approach to detect fake news was initiated using Link2vec mechanism, which is a self supervised learning model in order to differentiate between fake and real news gained from web search result using Word2vec [12]. The result of this study was very appreciable as they showed that this way is far better than whitelisting method and doesn't need human intervention. Another attempt was made for automatic detection of fake news combining content as well as social signals [4]. This study result shows even though social based method gives good performance they gradually degrade for news item. To detect breaking news rumors of emerging topics in social media a research was done based on twitter only using both Word2vec and LSTM-RNN model, which was capable to detect breaking news rumors [1].

All these studies and many others which were not mentioned here somehow plays a role in detecting fake news in internet using different techniques and methods and models, which have some advantage and disadvantage,success and drawbacks but each and every study is important for making a further better model or finding a better way to detect fake news than the already existing ones. This paper also aims to create a new dimension with a low resource language like Bangla to make the task of detecting fake and falsified news in the internet easy.

3 Problem Statement

When reading a news from an article or a link not many people cares about the authenticity of the news. But the news they read does keeps a trace in their mind for some time or throughout the day. If someone reads a news about the weather he or she will take an umbrella in case the news shows that there is a chance of rain and will carry a water bottle in case the news shows it will be a very sunny day. Here the authenticity of the news is not too necessary but if the news turns out to be false and someone is carrying an umbrella throughout the day it will just be an extra baggage and nothing else. Though this is just an example but still because of a simple false news someone is facing a problem. So if there is a way that one can detect if the news is true or not even if it is a small and not harmful news one can relieve themselves from some extra hard work like carrying an umbrella.

When a human try to find if a news is fake or not by his or her own intelligence the first question will be why do they think it is fake? The answers can be:

- The news has no trustworthy source.

- The content seems unrealistic.

- They had some knowledge about the content from before.

- They found some mismatch with some other news.

. But who will think this much when reading a news? So they just read it as it is may it be fake or fraud. Which later effects their judgement towards any situation. So to avoid any kind of misjudgement, the problem of spreading of fake news is considered as an important field for researches.

There have been many attempt made to detect fake news from internet like using social media platforms as domains but if we are going to use the social media as a domain then we can only get the information an person or a news site has posted in their profile nothing more than that. If we are going to detect fake news by making a whitelist of the news found in different web portals then we will be needing human intervention for making that whitelist which is too much of a labor. There also have been many studies that uses linguistics features to identify which news is fake and which is true but as fake news are becoming more similar to the pattern of real news this process is not so useful now. Other than that the most thought about problem in this study is the language barrier. All this studies we are talking about were mostly done in English language or a high resource language like English but what about the low resource languages like Bangla? So keeping this in mind this study will focus on Bnagla language to detect fake and falsified news from web.

The idea of detecting fake news needs some collection of data from which we can classify that this
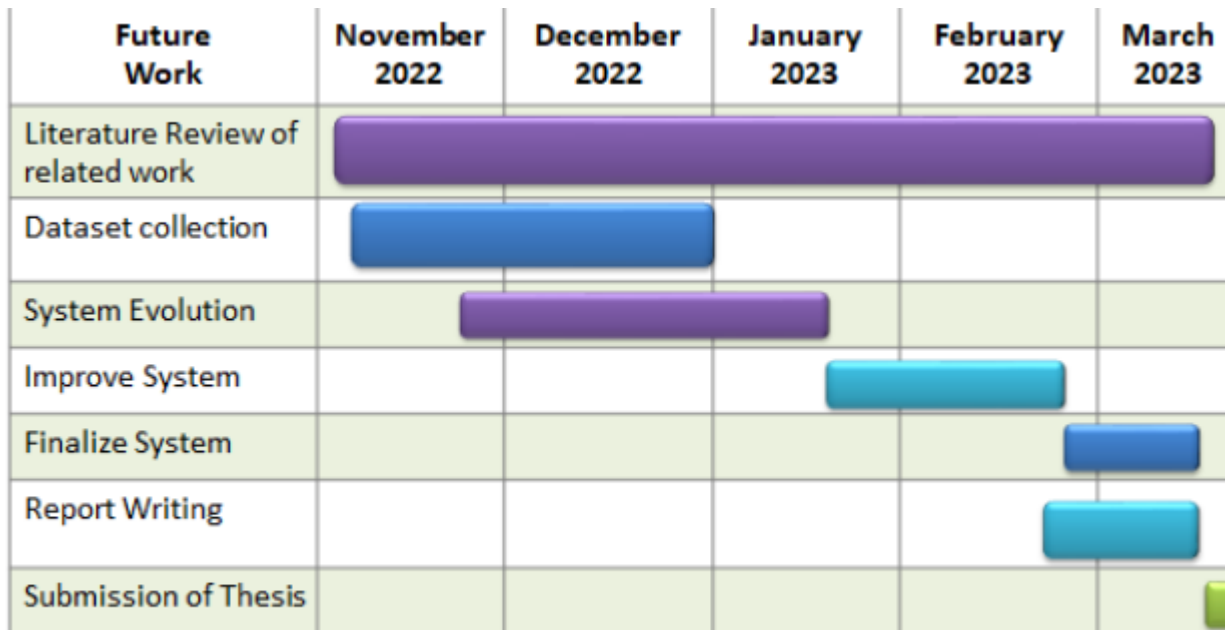
news is fake and this news is authentic. And we can decide whether our proposed model is working or not. So keeping that in mind we have collected a dataset from a platform consisting of different type of bangla news. And in our study we want to use Support Vector Model(SVM) which has also been previously used in many other fake news detection researches and successfully implemented and showed an impressive result.There may further other models which will be used but this study also focuses on web search based bangla fake news detection so for that purpose a link2vec based model is also considered to be used.

4 Preliminary Result

With the mindset to detect fake news in bangla and to get desired result from this study a step by step approach is needed which will be thoroughly prepared so that the work goes smoothly. The first step should be deciding the language which we have already decided as Bangla language and next the model we will be using which for now is decided as SVM classifier model. Also we will be needing a dataset of bangla news from different news websites which will help us to get a whitelist of news. Till now we have collected a dataset from 'kaggle.com' where there are many datasets available for fake news in different language. As for implementation we are deciding to choose link2vec model that uses the mechanism of word2vec.

5   Conclusion and Future Work

In this paper we discussed the basic idea of our selected topic. In order to complete our study further we need to implement all the models selected till now and use some other models also if needed. This paper focuses on SVM classifier considering its success in many other researches and also will use a word2vec mechanism based link2vec model to detect fake news from web search. Lin2vec model can find out the web search result links by itself so no human intervention is needed here like many other models where human intervention was needed. It also may have drawbacks which we will find out in our process of implementing the total research. At last we hope that this study will be a good contribution in automatic fake news detection research field.

Future Work	November 2022	December 2022	January 2023	February 2023	March 2023
Literature Review of related work					
Dataset collection					
System Evolution					
Improve System					
Finalize System					
Report Writing					
Submission of Thesis					

!!   Figure: Time management of Future work

6 References

[1] Sarah A Alkhodair et al. "Detecting breaking news rumors of emerging topics in social media". In: *Information Processing & Management* 57.2 (2020), p. 102018.

[2] Alessandro Bondielli and Francesco Marcelloni. "A survey on fake news and rumour detection techniques". In: *Information Sciences* 497 (2019), pp. 38–55.

[3] Abhijnan Chakraborty et al. "Stop clickbait: Detecting and preventing clickbaits in online news media". In: *2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*. IEEE. 2016, pp. 9–16.

[4] Marco L Della Vedova et al. "Automatic online fake news detection combining content and social signals". In: *2018 22nd conference of open innovations association (FRUCT)*. IEEE. 2018, pp. 272–279.

[5] Edouard Grave et al. "Learning word vectors for 157 languages". In: *arXiv preprint arXiv:1802.06893* (2018).

[6] Md Zobaer Hossain et al. "Banfakenews: A dataset for detecting fake news in bangla". In: *arXiv preprint arXiv:2004.08789* (2020).

[7] Md Gulzar Hussain et al. "Detection of bangla fake news using mnb and svm classifier". In: *2020 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*. IEEE. 2020, pp. 81–85.

[8] Linfeng Li et al. "Towards a contingency approach with whitelist-and blacklist-based anti-phishing applications: what do usability tests indicate?" In: *Behaviour & Information Technology* 33.11 (2014), pp. 1136–1147.

[9] V Pérez-Rosas et al. "Automatic fake news detection". In: *Proceedings of the 27th International Conference on Computational Linguistics. Association for Computational Linguistics, Santa Fe*. 2018.

[10] Inggrid Yanuar Risca Pratiwi, Rosa Andrie Asmara, and Faisal Rahutomo. "Study of hoax news detection using naıve bayes classifier in Indonesian language". In: *2017 11th International Conference on Information & Communication Technology and System (ICTS)*. IEEE. 2017, pp. 73–78.

[11] Victoria L Rubin 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, pp. 7–17.

[12]   Jae-Seung Shim, Yunju Lee, and Hyunchul Ahn. "A link2vec-based fake news detection model using web search results". In: *Expert Systems with Applications* 184 (2021), p. 115491.

[13]   Kai Shu et al. "Fake news detection on social media: A data mining perspective". In: *ACM SIGKDD explorations newsletter* 19.1 (2017), pp. 22–36.