

 2<sup>nd</sup> National Conference on Engineering Applications of Emerging Technology in association with International Journal of Scientific Research in Science,
Engineering and Technology | Print ISSN: 2395-1990 | Online ISSN: 2394-4099
[ https://ijsrset.com | doi : https://doi.org/10.32628/IJSRSET ]

# **Fake News Detection and Sentiment Analysis**

Divya Bhat N1, Poojashri P S1, Sahana P1, Sharan S 1, Reshma S2

<sup>1</sup>Student, <sup>2</sup>Assistant Professor

Department of Computer Science and Engineering, Global Academy of Technology, Bangalore, Karnataka,

India

# ABSTRACT

Social media used for societal knowledge is a double-edged sword. On the one side it is the rapid propagation, low cost, easy availability of information at just the click of a button leads people to avail news from the social platform. The other side proves the wide spread of "fake news", that are middling news with consciously incorporated false information. The immense propagation of fake news can potentially prove to deliver extremely negative impacts on individual people and the society. Hence, Detection of fake news on the social platform has currently become upcoming research alluring a large amount of attention. False news detection on digital media portrays attributes and challenges that are unique and that which make existing detection algorithms futile or not applicable. The Dataset used for training and testing the Machine Learning algorithm is downloaded from Kaggle.com which is used to perform classification of the message or text. And for Sentiment Analysis and developing various other surveys to provide a completely full-fledged report/analysis the aim is to retrieve tweets of the past 30 days. The procedure towards developing Sentiment Analysis and Survey reports involve the need for Natural Language Processing. As a Summary, the project is spread mainly across 2 domains involving Data Science to scrape for Tweets and Machine Learning in providing the necessary working prototype of the Project.

### I. INTRODUCTION

As an increasing amount of people's lives is spent interacting through various social media platforms. People tend to hunt out more and gain news from social network instead of traditional news organizations. The explanations for this, is the nature of these social media platforms:

 (i) it is often timelier and less extravagant to absorb news on social media in comparison to conventional journalism, like television or newspapers; and (ii) it is effortless to further share and debate the news with friends and other readers on social media.

Not to our surprise, it is also found that social media now surpasses television; the vital news source. Though having benefits from social media, the standard of information on social media is less when compared to conventional news organizations. However, as it is inexpensive to share news online and works at a much faster rate and since it is easy to propagate via social network, huge volumes of false news are generated online for

**Copyright:** © the author(s), publisher and licensee Technoscience Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited



purposes like financial and political gain. The increasing spread of false news has impacted negatively on individuals and the society. Firstly, fake news can destroy the counterpoise of the news ecosystem. For instance, it is transparent that the most popular fake news was outstretched on Facebook than the most obtained original mainstream news while the U.S. presidential election (2016). Second, fake news deliberately convinces consumers to simply adhere to false or biased beliefs. Third, fake news manipulates the way people interpret information and their answers along with their perspective. For instance, some fake news was just generated to trigger people's suspicion and make them bewildered; approaching their capacities to differentiate what's true from what's not. To assist alleviate the negative effects created by false news, it is critical that we develop methods to automatically discover fake news broadcast on social media.

# **II. LITERATURE REVIEW**

What is fake news? It is the conscious outspread of disinformation via social media or via traditional media. False information news spreads exceptionally fast. This is illustrated by the fact that, when one fake information website is demounted, another will swiftly replace it. Additionally, fake news can become identical to accurate journalism as it extends very quickly. We can download and upload articles on websites, share the data and information and re-share it and at the end the fake information will be gone far from its initial website that it might become indifferentiable from real information or news.

There are several techniques to differentiate fake news from real news like *n*-grams, bag-of-words, TFIDF, GloVe, and contextual embeddings. Techniques such as bag-of-words will not comprise conditions and depend on frequencies of words, albeit, researchers have also used semantic analyses to discover accuracy in a subject. We can also see a well deep syntax approach using probability context free grammar (PCFG) parsing trees. This approach uses revised uses of sentences to study dissimilarity in syntax structures in real and fake news. Another linguistic appeal is to contemplate the topic of the article and examine its relevancy with the contents of the article. This is done by using linguistic attributes such as the length of the advertisements, headlines, author attributes, text patterns etc. Several machine-learning methods have been used as well for fake news detection: knearest neighborhood (KNN), SVM, LR, Naïve Bayes (NB), Random Forest (RF) and Decision Trees (DT). These methods have exhibited strength in classifying disinformation using many features. Since feature engineering is time- consuming, several neural network propositions such as word count (LIWC) features and LSTM with linguistic inquiry, RNN established models for users involvement including convolutional neural network (CNN) established model with confined attributes were implemented to detect fake news.

A Naive Bayes classifier is a machine learning algorithm, the variables that are used to create the model are absolute. It is determined that Naïve Bayes classifier produces fairly good results. The Naïve Bayes classification is directed by acquiring the maximum posterior, that is the maximal P(Ci

X). Naive Bayes is a well-known algorithm implemented to discover the correctness of the information whether it is fake or real using multinomial Naïve Bayes. There exists numerous algorithms that center on the common principle, hence it is not the only algorithm used for training such classifiers. Naïve Bayes can be used to examine the news if it is fake or real. Although false information detection in social network getting attention fairly lately, there has been irregularity of publications and research on the matter. The dataset issue should be considered before talking about fake news detection using machine learning. For news outlet stance detection to enable fake information detection problems on some techniques such as Natural Language Processing (NLP) have been implemented. FNC-1 have made the datasets accessible publicly and we are not far from possessing standard benchmarks to juxtapose all of the recently proposed methods.

# **III. LITERATURE SURVEY**

Table 1: Summary of Literature Survey

Title of	Methodolo	Advantages	Limitations
Paperand	gy		
Year			
"Fake	count	Naive Bayes	Simple
News	vectorizer	is:	context of n-
Detection	or a matrix	Bayes	grams and
usingNaïve	(frequency	theorem	POStagging
Bayes	of terms	with strong	have
Classifier"	and inverse	(naive)	permitted to
(2021)	document	independen	be
	frequency).	ce	inefficient
	Naive	assumptions	for the
	Bayes	. Simple	classification
	classifier.	technique	•
		for	
		constructin	
		g classifiers.	

Effect of	Collecting	Classificatio	Naive Bayes
Corpora on	data,	n accuracy	classifier has
Classificati	Preparing	about87%	some
on of Fake	NLP	and 92% by	drawback as
News	environme	naive Bayes	textual
usingNaive	nt, Data	Algorithm	processing.
Bayes	Processing,	with	
Classifier	Feature	enriched	
(2020)	Extraction,	corpora.	
	Confusion		
	matrix.		
Fake News	Classifiers,	The model	Relies on
Detection	Dynamicity	is trained	identifying
using	for Truth,	using a	highly
Machine	Probability,	suitable	retweeted
learning	Authenticit	dataset.	threads of
Algorithm	y, Web		conversation
(2020)	Scrapping.		
A Smart	SVM,	Using SVM	The
System for	Naïve	(Support	analytical
Fake News	Bayes	Vector	algorithms
Detection	classifier,	Machine),	and
Using	Semantic	Reduces the	classifiers
Machine	investigatio	computatio	are needed
Learning	n.	nal cost.	to work
(2019)		New	together for
		component	the
		which gives	documentati
		new	on of news
		suggestions	articles.
		if the news	
		is fake.	
Fake News	Compositio	Using SVM	SVM
Detection	n of	(Support	approach
in social	Support	Vector	that has
media	Vector	Machine)	difficulty
(2018)	Machines,	which	with large
	Naïve	tends to be	datasets.
	Bayes	very	Naïve Bayes

	classifier	accurate	classifier
	and	and	deems all
	Semantic	performs	features of a
	analysis.	extremely	document
		well on	being used,
		small and	to be
		concise	independent
		datasets	
		Adding	
		semantic	
		analysis	
		helps fix	
		one of the	
		biggest	
		weaknesses	
		of Naïve	
		Bayes	
		classifier.	
Machine	Sentence	Pre-	The model is
Learning	and	processing	relatively
or	Document	of small	unfazed by
Detection	Level	dataset, a	the rejection
of Fake	Baseline	CNN is	of some
News	(Logistic	required to	"giveaway"
2018)	Regression)	pick up on a	topic words
		various set	of the
		of subtle	training
		patterns of	dataset.
		language.	
News	A Rapid	Heuristic	Multiple
Credibility	Application	evaluation	calls to the
Checking	Developme	of the	API.
System for	nt (RAD).	extension.	
ſwitter			
2017)			

### IV. EXISTING SYSTEM

The currently available system is well-designed for fake news detection. Nevertheless, it has some following limitations: "Chen, Conroy, Ruben and others outline several advancements that prove to be promising towards accurately classifying the deceptive articles. They point that basic and simple context associated n-grams and lower POS have demonstrated to be imminent for the classification of the Dataset. Other contributing factor are:

- 1. Lack of an awareness of this system.
- 2. Implementation is difficult and complex.
- 3. Some security related issues may be created.
- 4. Cost Effectiveness

# V. PROPOSED SYSTEM

Retrieve all Tweets using the Twitter API for testing and to provide sentiment analysis using NLP and generate reports using Python libraries generally such as "matplotlib" and many more such libraries. The building of the prototype is considered to be brought about implementing Naïve Bayes Classifier.



Fig 1: Architecture of Fake News Detection



Fig 2: Architecture of Sentiment Analysis

# VI. METHODOLOGY

Prototyping a Fake News Detector was indeed a challenging accomplishment. For ensuring that we

complete the prototype efficiently without facing issues, which would have caused vital remodelling of the architecture in time and cost constrained project environment. We began with basic development of SRS (Software Requirement Specifications) and a completely detailed design of the prototype. After that we started our research on the dataset which we downloaded from Kaggle and further extending our research to Machine Learning Algorithms for classification using Naïve Bayes Classifier and what kind of NLP to use. We used Multinomial NB as our machine learning algorithm, which gave us accuracy of 97.3%.



Posterior Probability

 $P(c \mid \mathbf{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$ 

#### VII. CONCLUSION

Social Media has assumed quite high proportions with a constant increase in quantity of adolescents afflicting to having experienced as either a casualty or as an eye-witness. Anonymity and insufficiency of supervision in the social media two main contributing factors that have aggravated this social jeopardy. Fake news is a current happening which is has an exponential effect on the societal life. Fake news detection is a rising research domain that has gained popularity and people's interest but involved certain challenges due to the constrained amount of resources that are available. We propose in this project, a fake news detection prototype that is bound to use machine learning techniques, predominantly Naïve Bayes.

#### VIII. REFERENCES

- [1]. I.M.V.Krishna, Dr.S.Sai Kumar, "Fake News Detection using Naïve Bayes Classifier" Department of Information Technology, PVP Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh, India (2021).
- [2]. Farzana Islam Adiba, Tahmina Islam, M Shamim Kaiser, Mufti Mahmud, Muhammad Arifur Rahman, "Effect of Corpora on Classification of using Naive Bayes Fake News Classifier", Jahangirnagar University, Dhaka, Bangladesh, Nottingham Trent University, Nottingham, UK (2020).
- [3]. Uma Sharma, Shankar N Patil, Sidarth Saran, "Fake News Detection using Machine Learning Algorithm", International Journal of Creative Research Thoughts (IJCRTS) (2020).
- [4]. Anjali Jain, Harsh Khatter, Avinash Shakya, "A Smart System for Fake News Detection Using Machine Learning", Dr. APJ Abdul Kalam University, Lucknow, India (2019).
- [5]. R.J. Poovaraghan, M.V. Keerti Priya, P.V. Sai Surya Vamsi, Mansi Mewara, Sowmya Loganathan, "Fake News Accuracy using Naive Bayes Classifier", International Journal of Recent Technology and Engineering (IJRTE) (2019).
- [6]. Nicole O'Brien, "Machine Learning for Detection of Fake News", Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (2018).
- [7]. Kelly Stahl, "Fake News Detection in Social Media", Department of Mathematics and Department of Computer Sciences, California State University Stanislaus, 1 University Circle, Turlock (2018).
- Kristin Kinmont, "News Credibility Checking [8]. System for Twitter", Computer Science Honours, University of Cape Town (2017)