RUMOR DETECTION ON SOCIAL MEDIA: AN EXPERIMENTAL STUDY

BY

JOY SARKER
ID: 181-15-11053

FALGUNI MALAKAR
ID: 181-15-11162

AND
SAJOL KUMAR ROY
ID: 181-15-11145

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering.

Supervised By

TAPASY RABEYA
Lecturer
Department of CSE
Daffodil International University

Co Supervised By

Ms. Israt Jahan
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2022

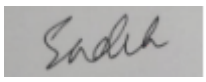
APPROVAL

This Project titled **"RUMOR DETECTION ON SOCIAL MEDIA : An Experimental Study**", submitted by Joy Sarker, Falguni Malakar and Sajol Kumar Roy to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering (BSc) and approved as to its style and contents. The presentation has been held on 05/12/21.

BOARD OF EXAMINERS

Chairman

Dr. Touhid Bhuiyan (DTB)
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Internal Examiner

Md. Sadekur Rahman (SR)
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Afsara Tasneem Misha (ATM)
Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Shah Md. Imran
Industry Promotion Expert
LICT Project, ICT Division, Bangladesh

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **TAPASY RABEYA, Lecturer, Department of CSE,** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

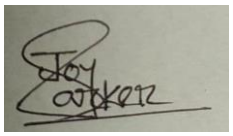
Supervised by:

TAPASY RABEYA
Lecturer
Department of CSE
Daffodil International University

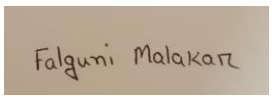
Co-Supervised by:

Ms. Israt Jahan
Lecture
Department of CSE
Daffodil International University

Submitted by:



Joy Sarker
ID: 181-15-11053
Department of CSE
Daffodil International University



Falguni Malakar
ID: 181-15-11162
Department of CSE
Daffodil International University

Sajol Kumar Roy
ID: 181-15-11145
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We are really grateful and wish our profound our indebtedness to **TAPASY RABEYA, Lecturer,** Department of CSE, Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Rumor Detection"* to carry out this project. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr.Touhid Bhuiyan**, Professor and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Social media stages have been utilized for information and news gathering, and they are very profitable in numerous applications. However, they lead to the spreading of rumors and fake news. Numerous endeavors have been taken to detect and expose rumors on social media by analyzing their substance and social context using machine learning procedures. This paper gives a diagram of the later studies in the rumor location field. It gives a comprehensive list of datasets utilized for rumor detection, and audits the imperative studies based on what sorts of data they exploit and the approaches they take. And more critically, we moreover show several new headings for future research.

TABLE OF CONTENTS

CONTENTS **PAGE**

LIST OF FIGURES

FIGURES **PAGE**

LIST OF TABLES

TABLES **PAGE**

CHAPTER 1

Introduction

1.1 Introduction

Bangladesh is a country of youth generation. These youth are trying to keep themselves connected by using social medias. They share all their feelings of cherish and despise, hostility and viciousness over these social media stages. These youths frequently discover themselves in a rough virtual space that has contrarily impacted their every day genuine life. We are attempting to get it youthful people's states of mind and behavior over social media stages and endeavors to clarify the social media as an uncontrolled and savage stage for the young people.

1.2 Motivation

Emotions are the best way to express what a person is feeling at that particular time. Now a day's people share their views, emotions on social networking sites such as, Facebook, twitter, instagram etc. Recently most of the people post status in Social media. Most of the time it becomes tough to understand the rumor reading those sentences. That's why we have decided to detect rumor from text. We started to read research papers and found out that lots of work has been done with detecting rumor in English. Then we start searching papers related to detecting rumor. Very few works has been done. So, we thought that we will work rumor detection from text and link. We will detect rumor from sentences and link by using dataset.

1.3 Rationale of the study

Detecting rumor from text and link is quite complex. As most of the rumor depend on the word expression. That's why we have decided to detect only two things from a text First of all it will check the text and get decision real or fake and the another is it will go to the link and check the result.

1.4 Research question

Research Question 1: Does every news have a distinct rumor such as real or fake?
Research Question 2: If a news has rumor, does it contain any proper reference text or link which will prove it is rumor?

1.5 Expected output

We have chosen to detect rumor from text or link, a very few works has been done till now detecting but text and link at a same project has not done yet. Our expected output is at first it will find out whether the news is fake or real. And then after the analysis it will detect the rumor of the news either true or false.

1.6 Report layout

The paper is organized into five sections. Following this introduction, Chapter 2 provides brief background details of Rumor detection field from an information systems perspective, a survey on text analysis those have been published in different information system journals, moreover the scope of the issue and its challenges. Chapter 3 provides a details description about our research techniques including data collecting procedure, statistical analysis. Chapter 4 presents the experimental result of applied methodology, brief description of the analysis. And finally Chapter 5 describes the summary of the empirical research, important limitations of the approach, implication for Further Study.

CHAPTER 2

BACKGROUND

2.1 Introduction

Do you accept all the news you tune in from social media? All news are not honest to goodness, right? So how will you recognize the fake news? We'll be utilizing Multinomial Credulous Bayes strategy to classify the news article rejected from the news article connect and content, as fake or genuine.

In our research, we have mainly worked on how we can compute rumor and thus sort out the rumor from a given text which has been collected from social media such as Facebook status and also from newspaper, noble, online news portal etc

Detecting rumor as well as emotion from a Text is a very tough and complicated task to carry out. In this research, we have tried to identify the positivity, negativity or neutral expression on a given text or link with respect to some trained data. We have used a trained data set to sort out the possibility of a given query in a text and extract an unparalleled feature to give an output whether the text or link is real or fake.

2.2 Related works

Many works have done yet. We want to do something new. In this work we want to add not only text detect but also detect the original news link. There are so many works has done but they have not done it by using multiple function like link and text at a same time in the same project.

Name of the paper	Writers	Publishing year	Main theme
Analysis of Techniques for	Ajeet Ram Pathaka		There's an expanded inquire about slant

Rumor Detection in Social Media	, Aditee Mahajana , Keshav Singha & others	2020	towards discovery of rumors due to overpowering utilize of social media Platform for dispersal of data and news. Considerable thinks about have centered on recognizing the source of The rumors and recognizing the rumors.
A Tweet level propagation context based deep neural networks for early rumor detection in Social Media	Jie Gao, Sooji Han, Xingyi Song, Fabio Ciravegna	2020	In this paper, we addressed the task of message-level ERD in early development stages of social media rumors where limited information is available.
Rumor Detection on Social Media with Bi-Directional Graph Convolutional Networks	Tian Bian, Xi Xiao, Tingyang Xu,Peilin Zhao, Wenbing Huang, Yu Rong,Junzhou Huang	2019	Early detection aims to detect rumor at the early stage of Propagation. Which is another important metric to evaluate the quality of the method.
Leveraging the Implicit Structure within Social Media for Emergent Rumor Detection	Justin Sampson, Fred Morstatter, Liang Wu, Huan Liu	2016	In this paper, we analyzed how the verifiable structure within social media discussions can be utilized to al- low for the location of emanant rumors. We discover that implicit linkage can altogether move forward the capacity to clas- sify rumors inside their developmental stages by leveraging in- formation inside related conversations.

Automatic Rumor Detection on Microblogs: A Survey	Juan Cao, Junbo Guo, Xirong Li & others	2018	It works for decreasing the rumor on social media. It works on manual efforts, either by human experts or by common user, to combat the ever-increasing rumors.
Rumor Detection Using Machine Learning Techniques on Social Media	Akshi Kumar and Saurabh Raj Sangwan	2018	This paper displayed the essential concepts of rumor discovery. As much as social media has ended up an priceless source for sharing real-time and significant information, it is additionally a breeding stage for rumors.
Automatic Detection of Rumor on Social Network	Qiao Zhang1,2, Shuiyuan Zhang1,2, Jian Dong3, Jinhua Xiong2(B) , and Xueqi Cheng2	2015	In this paper we center on recognizing rumor on social organize. To distinguish rumors from ordinary messages, we propose a rumor location strategy based on implicit highlights of substance and users.
Multiple Time-Series Data Analysis for Rumor Detection on Social Media	Lijun Qian, Chandra Mouli Madhav Kotteti	2018	It works on social media for existing rumor or fake news detection

2.3 Research Summary

Research is an organized way to find solutions of existing problems or problems that nobody has worked on before. It can be used for solving a new problem or it can be the expansion of past work on any particular field. Our Research is on detecting rumor from text & link that is associated with Machine Learning (Flask Web App) is challenging the human being to exceed human beings performance. There's been lots of work that has already done to detect rumor. We have studied lots of paper related to detecting rumor from text & link. They used different methods and among them we have chosen machine learning based method for rumor detection. For that reason we have developed a strong corpus that will match the words and news link according to its origin. At first we have decided to analyze rumor from a sentence then we will find the fake or real news according to that Dataset. We will divide the news whether it is fake or real. If it's true we will categorized the news then according to its dataset that has been allocated for positive and not rumor. If the sentence is false then it will be categorized rumor according to its allocated link

2.4 Scope of the Problem

Feeling affirmation from substance is incipiently an acknowledgement based classification which dilate the concept from Common dialect preparing counting Machine Learning as well. Rumor examination is a curious field of consider. These days, it has been including values to the industry as since rumor investigation bases its comes about on figure that are so inalienably compassionate, it is bound to ended up one of the basic drivers of numerous commerce choices in future. Moved forward precision and consistency in content mining methods can makes a difference to overcome the current issues. Now a days as the following wave of information revelation, content examination is accomplishing high commercial values. In this investigate we are going analyze content from social media news for finding related opinion of each sentence like fake or real. After recognizing the estimation of every sentence we'll at that point attempt to discover unmistakable feeling of each sentence like rumor or not.

2.5 Challenges

The huge dataset and same word having different meaning are another challenge. As we are working with generated expressions from text & link, so it is possible to have huge number of expressions. And same expression with different meaning is also happen. As a result it can’t be properly identify the rumor. For this sometimes it will be challenging to detect rumor.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

Fake news discovery may be a basic however challenging issue in Machine Learning. The quick rise of social organizing stages has not as it were yielded a tremendous increment in data openness but has moreover quickened the spread of fake news.

In this way, the impact of fake news has been developing, now and then amplifying to the offline world and undermining open security. Given the enormous sum of Web substance, programmed fake news discovery may be a down to earth machine learning issue valuable to all online substance suppliers, in arrange to diminish the human time and exertion to distinguish and avoid the spread of fake news. In this paper, we portray the challenges included in fake news discovery conjointly portray related errands.

We systematically overview and compare the task subtle elements, datasets courses of action that have been made for this task, also conversation almost the conceivable outcomes and limitations of them. Based on our bits of knowledge, we chart promising ask approximately heading, checking more fine-grained, point by point, sensible, and commonsense disclosure models. We moreover highlight the differentiate between fake news area and other related errands, and the importance of machine learning courses of action for fake news location.

3.2 Research Subject

The main goal of this research is to detect rumor from a text and link. In order to come up with an emotion associated with it by using a technique.

In case of finding the rumor of a news analysis can make it very specific. A news with

no fake news it may not have any detection. Our research subject is to find the rumor of a text or link to see how people generally express the rumor while writing.

3.3 Data collection procedure

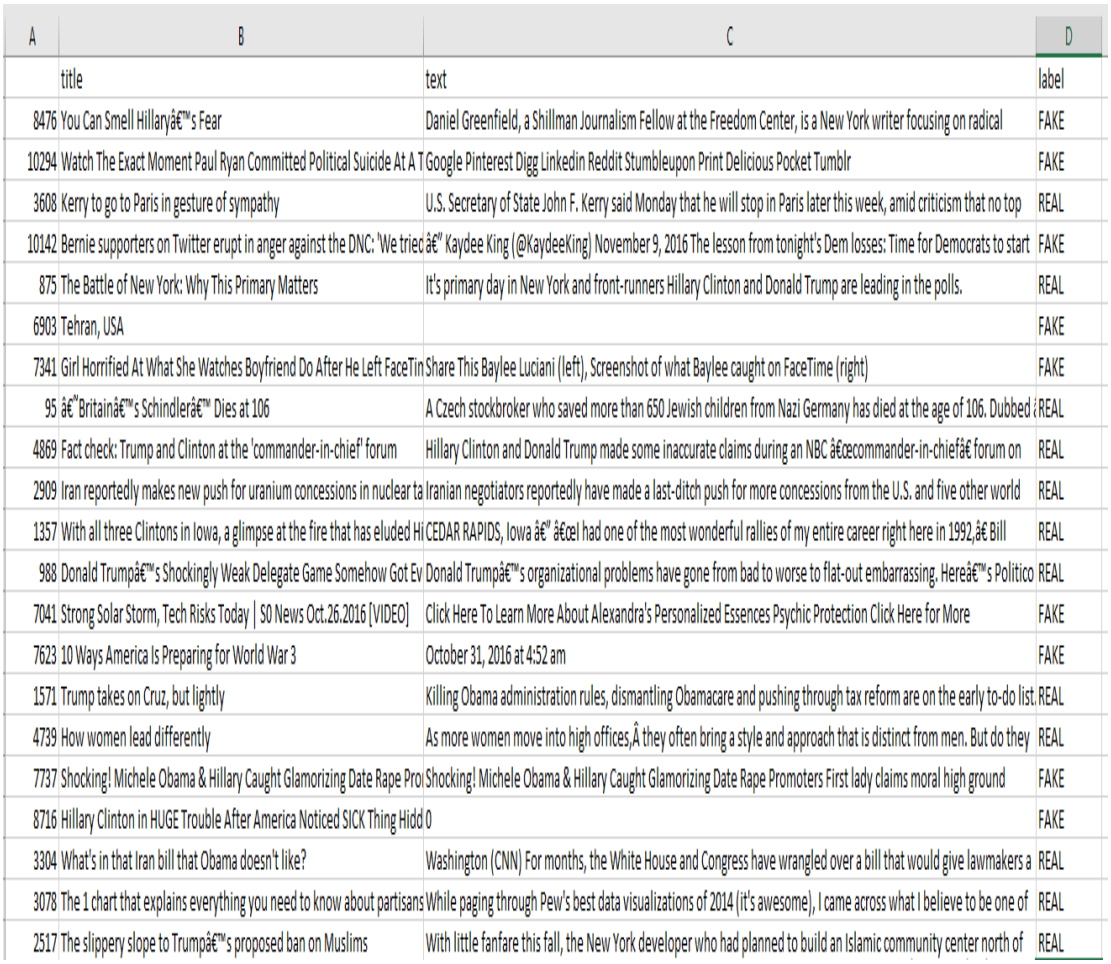
In order to come up with precise and objective results, effective research depended heavily on both primary and secondary data.

Primary data is unprocessed information that is primarily used for the original purpose and is obtained through interviews and surveys. Secondary data, on the other hand, is gathered for purposes other than those for which it was originally obtained.

Online news can be obtained from a variety of sources, including social media sites, search engines, news agency homepages, and fact-checking websites. These datasets have been widely utilized to determine the validity of news in a variety of research studies. The sources of the dataset used in this study are mentioned briefly in the following sections.

The study was conducted with the use of secondary data. We'll call the dataset we'll use for this Python project news.csv. The shape of this dataset is 77964. The first column identifies the news, the second and third columns contain the title and text, and the fourth column identifies the author.

Figure 3.1 shows the collection of our raw data which we have collected from different sites like Facebook, newspaper, website etc.



A	B	C	D
	title	text	label
8476	You Can Smell Hillaryâ€™s Fear	Daniel Greenfield, a Shillman Journalism Fellow at the Freedom Center, is a New York writer focusing on radical	FAKE
10294	Watch The Exact Moment Paul Ryan Committed Political Suicide At A T	Google Pinterest Digg Linkedin Reddit Stumbleupon Print Delicious Pocket Tumblr	FAKE
3608	Kerry to go to Paris in gesture of sympathy	U.S. Secretary of State John F. Kerry said Monday that he will stop in Paris later this week, amid criticism that no top	REAL
10142	Bernie supporters on Twitter erupt in anger against the DNC: 'We tried â€" Kaydee King (@KaydeeKing) November 9, 2016 The lesson from tonight's Dem losses: Time for Democrats to start	FAKE
875	The Battle of New York: Why This Primary Matters	It's primary day in New York and front-runners Hillary Clinton and Donald Trump are leading in the polls.	REAL
6903	Tehran, USA		FAKE
7341	Girl Horrified At What She Watches Boyfriend Do After He Left FaceTin	Share This Baylee Luciani (left), Screenshot of what Baylee caught on FaceTime (right)	FAKE
95	â€"Britainâ€™s Schindlerâ€™ Dies at 106	A Czech stockbroker who saved more than 650 Jewish children from Nazi Germany has died at the age of 106. Dubbed a	REAL
4869	Fact check: Trump and Clinton at the 'commander-in-chief' forum	Hillary Clinton and Donald Trump made some inaccurate claims during an NBC â€œcommander-in-chiefâ€ forum on	REAL
2909	Iran reportedly makes new push for uranium concessions in nuclear ta	Iranian negotiators reportedly have made a last-ditch push for more concessions from the U.S. and five other world	REAL
1357	With all three Clintons in Iowa, a glimpse at the fire that has eluded Hi	CEDAR RAPIDS, Iowa â€" â€œI had one of the most wonderful rallies of my entire career right here in 1992,â€ Bill	REAL
988	Donald Trumpâ€™s Shockingly Weak Delegate Game Somehow Got Ev	Donald Trumpâ€™s organizational problems have gone from bad to worse to flat-out embarrassing. Hereâ€™s Politico	REAL
7041	Strong Solar Storm, Tech Risks Today | S0 News Oct.26.2016 [VIDEO]	Click Here To Learn More About Alexandra's Personalized Essences Psychic Protection Click Here for More	FAKE
7623	10 Ways America Is Preparing for World War 3	October 31, 2016 at 4:52 am	FAKE
1571	Trump takes on Cruz, but lightly	Killing Obama administration rules, dismantling Obamacare and pushing through tax reform are on the early to-do list.	REAL
4739	How women lead differently	As more women move into high offices,Â they often bring a style and approach that is distinct from men. But do they	REAL
7737	Shocking! Michele Obama & Hillary Caught Glamorizing Date Rape Pro	Shocking! Michele Obama & Hillary Caught Glamorizing Date Rape Promoters First lady claims moral high ground	FAKE
8716	Hillary Clinton in HUGE Trouble After America Noticed SICK Thing Hidd	0	FAKE
3304	What's in that Iran bill that Obama doesn't like?	Washington (CNN) For months, the White House and Congress have wrangled over a bill that would give lawmakers a	REAL
3078	The 1 chart that explains everything you need to know about partisans	While paging through Pew's best data visualizations of 2014 (it's awesome), I came across what I believe to be one of	REAL
2517	The slippery slope to Trumpâ€™s proposed ban on Muslims	With little fanfare this fall, the New York developer who had planned to build an Islamic community center north of	REAL

Figure 3.1 Raw Data

3.4 : Methodology and Data Analysis

The system is developed in three stages, as explained in this paper. The first section is static and relies on a machine learning classifier to function. We investigated and trained dataset before settling on the optimal one for final precaution. After that the portion is the news text and link searches the internet for the likelihood of the news being true. The third and final section verifies that the URL entered by the user is valid.

Python and the dataset were utilized in this paper. Python comes with a big number of libraries and extensions that may be utilized in Machine Learning with ease. Closely all forms of machine learning algorithms are available for Python allowing for easy and speedy assessment of Machine Learning (Flask Web App) methods.

System Design

Figure 3.4.1: System Design of text detect

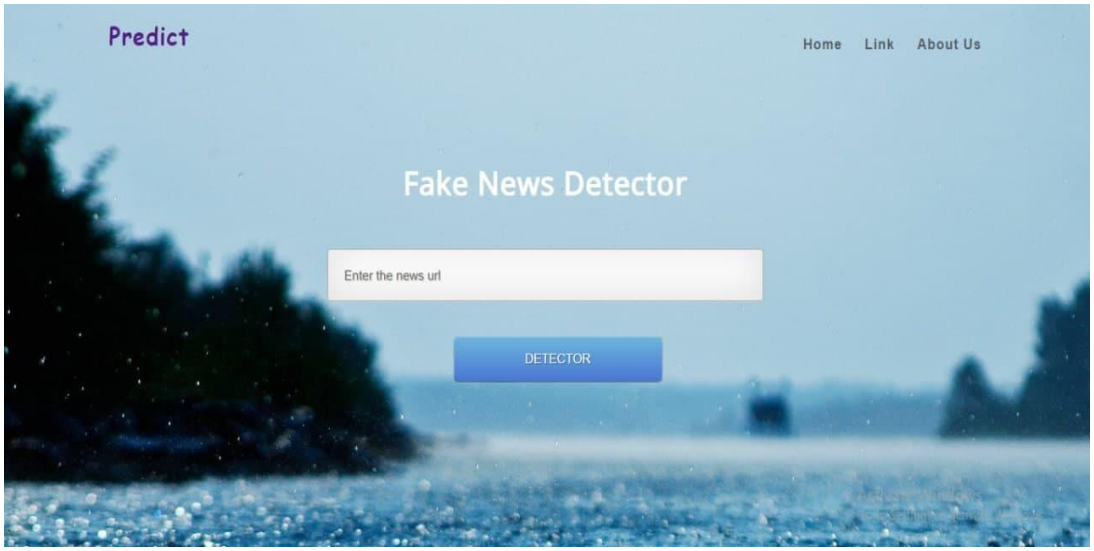


Figure 3.4.2: System Design of link detect

3.4.1 STATIC SEARCH FOR RUMOR DETECTION

The static element of the false news detection system's design is pretty straightforward, and it's built around the simple machine learning process. This system is designed by self-descriptive.

3.4.2 DYNAMIC SEARCH FOR RUMOR DETECTION

The site's second search form requests particular words to be searched on the internet, and it returns an appropriate % surety of that term being included in a paper or a comparable piece with those word references.

Searching for URLs

The site's third look field acknowledges a particular site space title, and the usage looks our genuine destinations or boycotted locales databases for the location. The true sites database contains domain names that deliver accurate and reliable news on a regular

basis. If the site isn't in either database, the execution isn't going to work.

The news aggregator does not exist, according to the domain classification.

3.4.3 PRE-PROCESSING DATA

Social media data is significantly unstructured bigger portion of them are casual communication with sorts, slangs and bad-grammar etc. Travel for extended execution and unflinching quality has made it essential to make strategies for utilization of resources to make taught choices. To achieve more better encounters, it is crucial to clean the data a few time as of late it can be utilized for prescient illustrating. For this reason, fundamental pre- planning was done on the News planning data. This step was comprised of- Information Cleaning: Whereas scrutinizing data, we get information inside the organized or unstructured orchestrate. A organized organize highlights a well- characterized plan whereas unstructured data has no appropriate structure. In between the 2 structures, we have a semi-structured organize which can be a comparably more better organized than unstructured organize.

Cleaning up the substance data is imperative to highlight qualities that were pointing to require our machine learning system to select up on. Cleaning (or pre- dealing with) the data customarily comprises of a number of steps:

1. Remove punctuation

Accentuation may give linguistic setting to a text which underpins our understanding. But for our vectorizer which tallies the number of words and not the setting, it does not include esteem, so we evacuate all extraordinary characters. eg: what are you doing? -> What are you doing

2. Tokenization

   Tokenizing isolates content into units such as sentences or words. It gives structure to already unstructured content. eg: Plata o Plomo -> Plata,o,Plomo.

3. Remove stopwords

   Stopwords are common words that will likely show up in any content. They don't tell us much around our information so we evacuate them. eg: silver or lead is fine for me -> silver, lead, fine.

4. Stemming

   It makes a difference decrease a word to its stem frame. It regularly makes sense to treat related words within the same way. It evacuates suffices, like ing, ly, s, etc. by a straightforward rule-based approach. It decreases the corpus of words but frequently the real words get ignored. eg: Entitling, Entitled -> Entitle.

   Note: A few look motors treat words with the same stem as equivalent words

3.4.4 ALGORITHMS USED FOR CLASSIFICATION

This area bargains with preparing the classifier. Distinctive classifiers were explored to anticipate the lesson of the content. We investigated particularly four diverse machine-learning calculations Multinomial NaÃ¯ve Bayes Detached Forceful Classifier and Calculated regression. The usage of these classifiers were done utilizing Python library Sci-Kit Learn.

Brief presentation to the algorithms-

1. NaÃ¯ve Bayes Classifier:

   This method is based on Bayes hypothesis, which accept that the nearness of a specific highlight in a course is free of the nearness of any other include. It gives way for calculating the back probability.

   Importance of a term within the report and whole corpus.

   TF stands for Term Recurrence: It calculates how habitually a term shows up in a archive. Since, each report measure changes, a term may show up more in a long measured record that a brief one. In this way, the length of the archive frequently separates Term frequency.

2. Random Forest:

   It could be a trademark term for a gathering of choice trees. In Arbitrary Timberland, weve collection of choice trees (so known as Woodland). To classify a modern protest based on qualities, each tree gives a classification that means it define the class. The woodland chooses the classification having the foremost votes (over all the trees within the timberland). The arbitrary timberland may be a classification calculation comprising of numerous choices trees. It employments sacking and include arbitrariness when building each person tree to undertake to make an uncorrelated woodland of trees whose forecast by committee is more accurate than that of any person tree. Irregular woodland, like its title implies, consists of a huge number of person choice trees that work as an outfit. Each person tree within the irregular woodland spits out a course expectation and the course with the foremost votes gets to be our models expectation.

   The reason that the irregular woodland show works so well may: Be a expansive number of moderately uncorrelated models (trees) working as a committee will beat any of the person constituent models.

3. Logistic Regression:

   It may be a classification not a relapse calculation. It is utilized to assess discrete values (Parallel values like 0/1, yes/no, real/fake) based on given set of free variable(s). In straightforward words, it predicts the likelihood of event of an occasion by fitting data to a calculated work. Thus, it is additionally known as calculated

relapse. Mathematically, the log chances of the result are demonstrated as a straight combination of the indicator factors.

4. Passive Aggressive Classifier:

   The Passive Aggressive Algorithm is used for online algorithm; perfect for classifying gigantic streams of information (e.g. twitter). It is simple to execute and exceptionally quick. For an example, learning from it and after that tossing it absent. Such a calculation remains passive for a rectify classification result, and turns forceful within the occasion of a miscount, updating and altering. Not at all like most other calculations, it does not focalize. Its purpose is to form upgrades that rectify the misfortune, causing exceptionally small alter within the standard of the weight vector.

3.5 IMPLEMENTATION STEPS

Step 1: In to begin with step, we have extricated highlights from the as of now pre-processed dataset

Step 2: Here, we have built all the classifiers for foreseeing the fake news discovery. The extricated highlights are encouraged into diverse classifiers. From this project we have used Naive-bayes, Logistic Regression, and Random forest classifiers from sklearn. Each of the extricated highlights was utilized in all of the classifiers.

Step 3: After that we have compared & checked the confusion matrix.

Step 4: After testing all the classifiers, 2 best performing models were chosen as candidate models for fake news classification.

Step 5: We have performed parameter tuning by executing GridSearchCV strategies on these candidate models and chosen best performing parameters for these classifier.

Step 6: At last our final model was selected for fake news detection with the probability of truth.

Step 7: Our final selection the best performing classifier was Logistic Regression which was then saved on disk. It will be used to classify the fake news.

It takes a news article as input from client at that point show is utilized for last classification yield that's shown to user at the side likelihood of truth.

CHAPTER 4
EXPERIMENTAL RESULTS AND DISCUSSION

4.1Introduction

In afterward times fake news area has finished up one of the prevalent subjects among the examiners all through the world. Diverse considers nitty gritty fake news revelation strategies. Fake news not because it were remains a word like a few time as of late it gets to be one of the greatest evenhanded perils these days. Anyone or any source can be the medium of fake news maker. It may be out of individual thought process or envy, to require correct revenge, it can be for political reasons, to harmed commerce competitors or to advance things or to accomplish reputation. Fake news imitates faster than veritable news. Where truth once in a whereas diffused among thousands of people, the fake news diffused between 1000-100,000 individuals interior a number of time. Development of the utilize of the internet and the utilize of social media increases the advancement of off-base news. The effect of fake news into the American society is preventing. Concurring to a report as it were 41% of grown-ups accept the neighborhood news as to be genuine and as it were 27% of the grown-ups believe the national news as being true. In 2016 the US decision spread purposefulness fake news did a significant harm. The commercialization and globalization of fake news could be a genuine issue now. Internet and social media is supposed to assist individuals to put through and understand, however it has presently too ended up a way of making chaos. Fake news is most predominant within the Facebook moreover found in YouTube, Twitter, blogs and surveyors.

4.1 Experimental Results

The sentences that we collected from different sources for analysis purpose have been stored as four kind of lexicon in our corpus. Recalling research question part, we have started our research with four questions. For getting the answers of our research questions that we applied to find rumor associated with text or link. After analyzing a

sentence, systems work is to say whether the input sentence contain any rumor it shows detected. Recalling research question part, we have started our research with four questions getting the answers of our research questions that we applied a back tracking technique to find rumor associated with sentence or link. After analyzing a sentence, systems work is to say whether the input sentence contain any fake news called rumor or not. We have performed three test with three different data set of data. Test results after performing the link tests has been posturized in the Figure

.

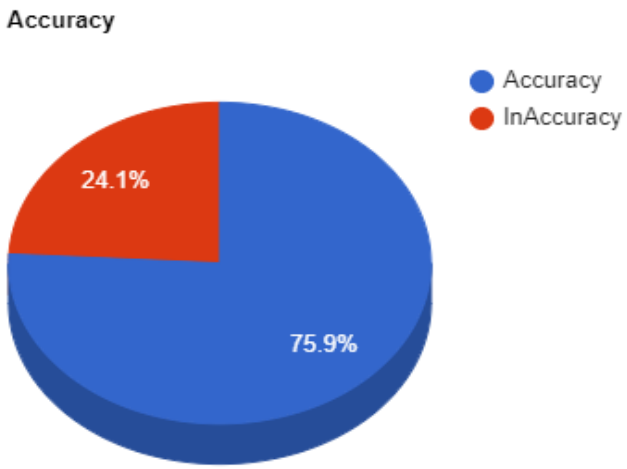


Fig4.2.1: Pic-chart for link accuracy

From the test we got 75.9% accuracy for detecting rumor from a link.

4.3 Comparison

4.3.1 Accuracy Table of rumor detection from text

Total link	Right Prediction	Wrong Prediction	Error Prediction
210	173	19	18

True Positive	True Negative	False Positive	False Negative
95	37	55	5

Fig: Accuracy Table of rumor detection from text

Recall precession & F-measurement are used for evaluating the performance of our algorithm which are defined as follows:

Precision = TruePositives / (TruePositives + FalsePositives) … … … (1)

Recall = TruePositives / (TruePositives + FalseNegatives) … … … (2)

F-Measure = (2 * Precision * Recall) / (Precision + Recall) … … … (3)
Accuracy

By using this formula we get the accuracy of the text from rumor detector is 71.8%

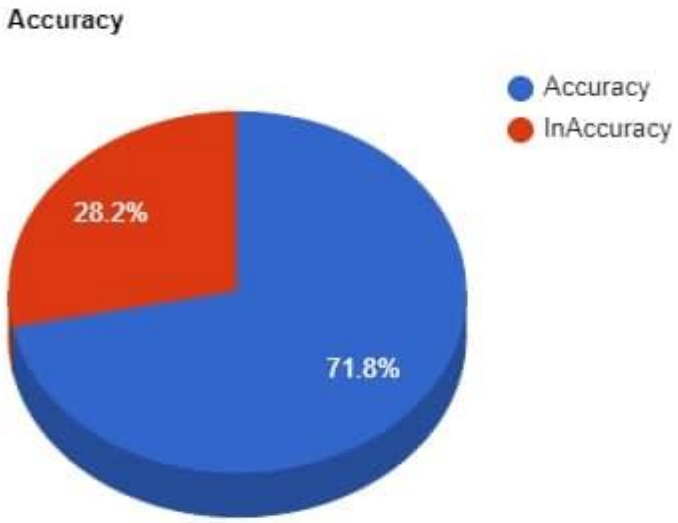


Fig for text accuracy

4.3.2 Accuracy Table of rumor detection from link

Total link	Right Prediction	Wrong Prediction	Error Prediction
210	173	19	23

True Positive	True Negative	False Positive	False Negative
87	34	59	7

Fig: Accuracy Table of rumor detection from link

Again from the same process Recall precession & F-measurement are used for evaluating the performance of our algorithm which are defined as follows:

Precision = TruePositives / (TruePositives + FalsePositives) … … … (1)

Recall = TruePositives / (TruePositives + FalseNegatives) … … … (2)

F-Measure = (2 * Precision * Recall) / (Precision + Recall) … … … (3)
Accuracy

By using this formula we get the accuracy of the text from rumor detector is 75.9%

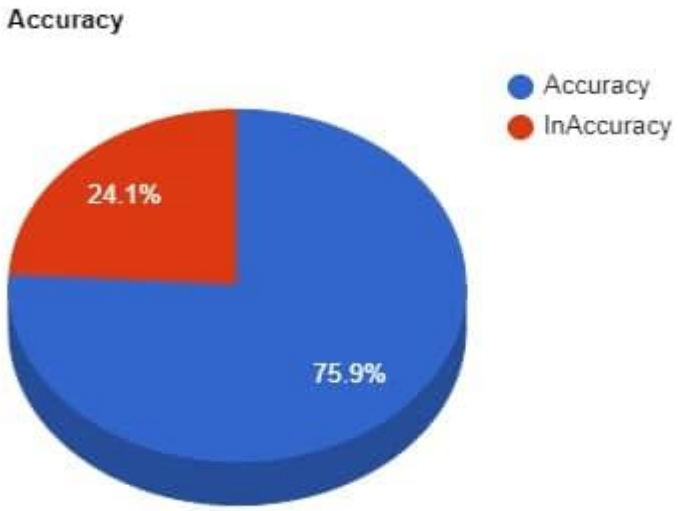


Fig for link accuracy

We are inspired to increase our accuracy from others and we show the comparison is clear that our accuracy is more than others.

4.2 Summary

This chapter shows the clear result of our four research questions. After experimenting we have found that, a text or link may have rumor, or it may be neutral. If a text or link has rumor then we can find the rumor (either fake or real) associated with it. If a sentence has no rumor it doesn't show any null. We will do it in future and after the experimental result we have come up with the answer of our fourth research question with 77.16% accuracy. Therefore the entire accuracy refers that it will contribute a big for our social life

CHAPTER 5
CONCLUSION

5.1 Summary of the Study

In this research we tried to detect rumor from text using rumor detection. Classifying the expressions into either true or false using rumor detection to detect rumor that was our main goal. It's a lexicon based approach where real and fake word is stored in database as a unique hash value. For each text & link an expression will be generated whether it's true or false. Then rumor will be detected. Our trained data and our test data both showed our expected results. After this experimental result finally we have come up with a decision that the text or link is fake or real.

5.2 Impact on society

All the social media users in Bangladesh are not concern about rumor. They believe any news without justifing anything. As a result their movement or action will not right. By following this trends social violence is increasing day by day. As a result our project will able to detect the news which is real or fake. For this it has a huge number of social values to protect violance. This project can make awareness of social people. It can Make a great impact on social media and in our day to day life.

5.3 Conclusion

Detecting rumor from a text or link was not that easy as people disagrees on identifying exact rumor of same sentence. Our text & linked based trained data helps us to detect exact rumor that majority people think about. Among different approaches we have used dataset to extract semantic information from a text or link for giving exact rumor. We obtained 77.16% accuracy.

5.4 Recommendation

As at present, our corpus doesn't have sufficient lexicons and links. So generating right expression from given input text &link will be failed if it does not contain specific lexicons. So before going for test add necessary key words to the database. While giving input keep focus on the spelling of the words and text. In case of spelling mistake, program will fail to generate right expressions. So user may get wrong answer.

5.5 Implications for Further Research

Now days, the demand for data mining analyst is highly appreciated. This is due to presence of abundant amount of data in our surroundings. To be more precise, it is the right time to work with these sorts of complex data, so that a new pattern can be introduced to resolve different critical problems. Rumor analysis is one of the fundamental branches of data mining. The experimental study which we have carried out on rumor detection with a satisfactory outcome is leaving a strong footprint behind our work. It has been observed that works on rumor analysis in English have lots of valuable impacts in everyday life. Research can be done on how our project can be utilized to build something for disables those who are especially blind.

5.6 Future Work

- We will enrich our dataset with more words and links
- We have done our preliminary experiment on very small amount of data. We will work with large amount of data so that the result will be more accurate.
- We have work on just two emotions either true or false. In future we will work on three basic words like; fake, real & null .
- We want to integrate our system with text to speech converter in order to build something helpful for blind people.

REFERENCES

1. Pathak, Ajeet Ram, Aditee Mahajan, Keshav Singh, Aishwarya Patil, and Anusha Nair. "Analysis of techniques for rumor detection in social media." *Procedia Computer Science* 167 (2020): 2286-2296.

2. Gao, Jie, Sooji Han, Xingyi Song, and Fabio Ciravegna. "RP-DNN: A tweet level propagation context based deep neural networks for early rumor detection in social media." *arXiv preprint arXiv:2002.12683* (2020).

3. Bian, Tian, Xi Xiao, Tingyang Xu, Peilin Zhao, Wenbing Huang, Yu Rong, and Junzhou Huang. "Rumor detection on social media with bi-directional graph convolutional networks." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, pp. 549-556. 2020.

4. Sampson, Justin, Fred Morstatter, Liang Wu, and Huan Liu. "Leveraging the implicit structure within social media for emergent rumor detection." In *Proceedings of the 25th ACM international on conference on information and knowledge management*, pp. 2377-2382. 2016.
5. Cao, J., J. Guo, X. Li, Z. Jin, H. Guo, and J. Li. "Automatic rumor detection on microblogs: a survey (2018)." *arXiv preprint arXiv:1807.03505*.

6. Kumar, Akshi, and Saurabh Raj Sangwan. "Rumor detection using machine learning techniques on social media." In *International conference on innovative computing and communications*, pp. 213-221. Springer, Singapore, 2019.

7. Zhang, Qiao, Shuiyuan Zhang, Jian Dong, Jinhua Xiong, and Xueqi Cheng. "Automatic detection of rumor on social network." In *Natural Language Processing and Chinese Computing*, pp. 113-122. Springer, Cham, 2015.

8. Kotteti, Chandra Mouli Madhav, Xishuang Dong, and Lijun Qian. "Multiple time-series data analysis for rumor detection on social media." In *2018 IEEE International Conference on Big Data (Big Data)*, pp. 4413-4419. IEEE, 2018.

rumor defense check

ORIGINALITY REPORT

24%	18%	10%	15%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ijert.org Internet Source	6%
2	Submitted to October University for Modern Sciences and Arts (MSA) Student Paper	6%
3	Farzana Islam, Mohammad Minhazul Alam, S. M. Shahadat Hossain, Abdul Motaleb, Sabrina Yeasmin, Mehedi Hasan, Rashedur M. Rahman. "Bengali Fake News Detection", 2020 IEEE 10th International Conference on Intelligent Systems (IS), 2020 Publication	3%
4	Submitted to Daffodil International University Student Paper	2%
5	arxiv.org Internet Source	1%
6	www.public.asu.edu Internet Source	1%
7	Akshi Kumar, Saurabh Raj Sangwan. "Chapter 23 Rumor Detection Using Machine Learning	1%