A Mini Project Synopsis on

Fake News Detection

T.E. - I.T Engineering

Submitted By

Abdul Samad Ansari	**19104022**
Ekta Shantaram Gujar	**19104026**
Sindura Rajendra Dasi	**19104015**

Under The Guidance Of

Prof. Yaminee Patil

DEPARTMENT OF INFORMATION TECHNOLOGY

A.P.SHAH INSTITUTE OF TECHNOLOGY

G.B. Road, Kasarvadavali, Thane (W), Mumbai-400615

UNIVERSITY OF MUMBAI

Academic year : 2021-22

CERTIFICATE

This to certify that the Mini Project report on Fake News Detection has been submitted by Abdul Samad Ansari(19104022), Ekta Shantaram Gujar(19104026) and Sindura Rajendra Dasi (19104015) who are a Bonafede students of A. P. Shah Institute of Technology, Thane, Mumbai, as a partial fulfilment of the requirement for the degree in **Information Technology**, during the academic year **2021-2022** in the satisfactory manner as per the curriculum laid down by University of Mumbai.

Ms. Yaminee Patil
Guide

Prof. Kiran Deshpande
Head Department of Information Technology

Dr. Uttam D.Kolekar
Principal

External Examiner(s)
1.
2.

Place:A.P.Shah Institute of Technology, Thane
Date: 25/4/2022

TABLE OF CONTENTS

Chapter 1

Introduction

1.1. Purpose :

Fake news is one of the biggest discouragements in our digitally connected world. Fake news spreads at lightning-fastspeed impacting millions of people in the form of clickbait, trigrams everyday. Therefore, noticing fake news becomes a vital problem attracting huge research efforts. Detection of fake news from social media always creates a new challenge. It is written on social media to mislead readers. In the 2016 US presidential election, fake news propagated more on Facebook than authentic news. Fake news detection on social media has attracted politicians to researchers. The detection of fake change the mindset of people or society or country. So, it is very important for those readers who read news on daily basis on social media to know whether the news is real or fake. So, they always try to read news from authenticating sites or authors. In this report, we present a survey on the state of the art pertaining to the type of fake news and solutions that are being proposed. The research in this field has been going on for a long time and in the Indian context, the ill effects of spreading fake news are far from what anyone might think. The year 2020 was a full pandemic year due to coronavirus. So all the school examinations and university exams had been postponed for some months. As usual, the CBSE 10th and 12th ituation, it was extended to May and June. The CBSE board only released the timetable of the exam but on some social media sites, the exact date of each paper was also showing which was false. This fake news distracted the mind of students and they started to follow this fake timetable of the exam. After confirmation of CBSE regarding the timetable of the exam that it has not published yet, all the students get very disappointed. This one example of fake news shows how much fake news can poorly affect human beings' lives. One fake news can suppress the image of society and can change the thinking skill of human beings. Due to this, it is very necessary to detect fake news before spreading it.

1.2. Objectives :

- To implement TF-IDF Vectorizer for automatic classification of text into positive and negative vectors.
- To design the system using best suited algorithm in such a way that it can easily predict the false news as soon as the user enters the data.
- To process the system to obtain the better accuracy results.

1.3. Scope

- Can be used for reducing the time required to search for a place leading to quicker decision making with respect to places to visit.
- Automatic fake news detection may be done using the latest news and training the model regularly to get the best results. So this can be used as a filter to upload the news.

Chapter 2

Problem Definition

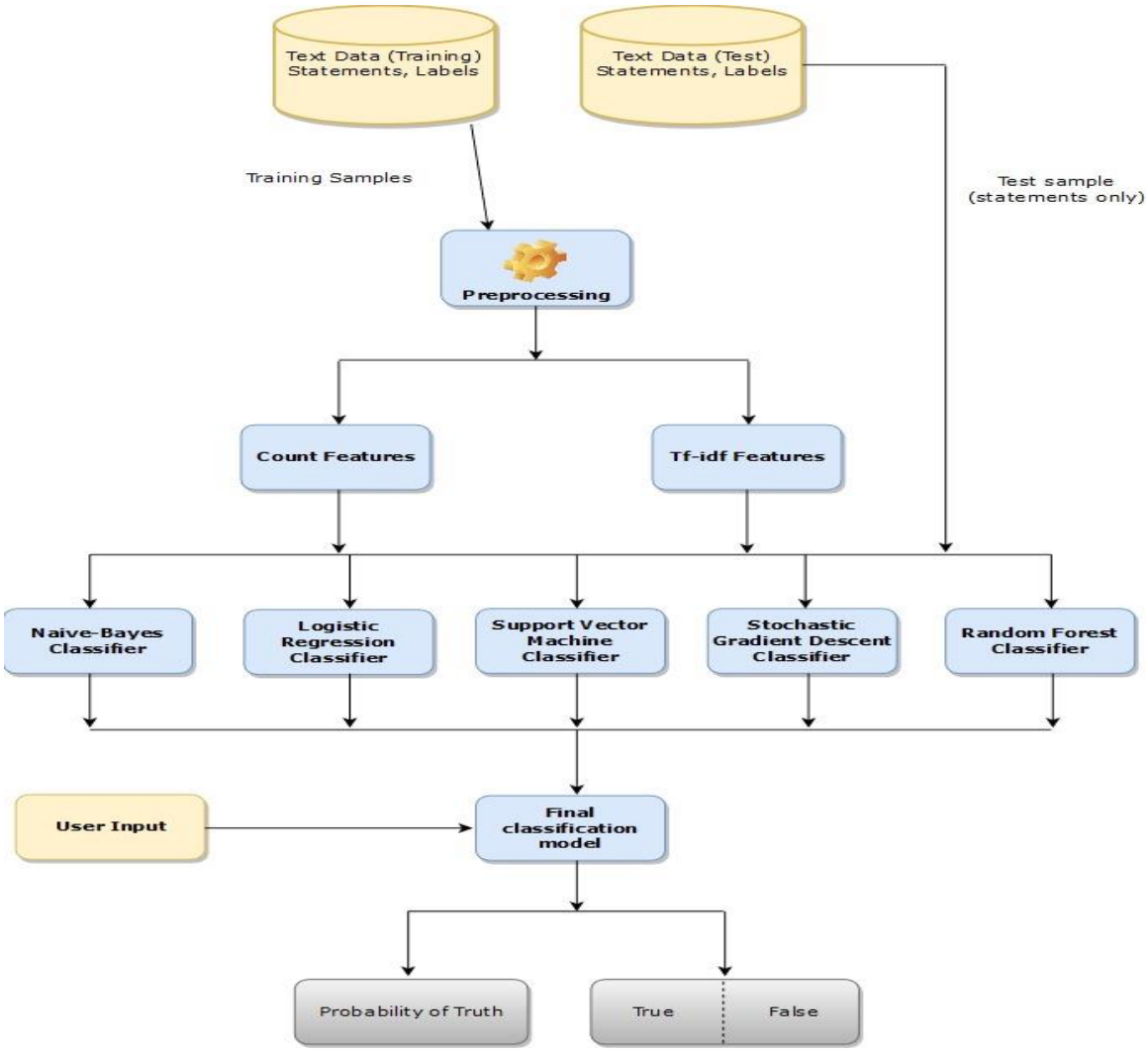
The widespread problem of fake news is very difficult to tackle in today's digital world where there are thousands of information sharing platforms through which fake news or misinformation may propagate. It has become a greater issue because of the advancements in AI which brings along artificial bots that may be used to create and spread fake news. The situation is dire because many people believe anything they read on the internet and the ones who are amateur or are new to the digital technology may be easily fooled. A similar problem is fraud that may happen due to spam or malicious emails and messages. So, it is compelling enough acknowledge this problem take on this challenge to control the rates of crime, political unrest, grief, and thwart the attempts of spreading fake news. Fake news, information bubbles, news manipulation and the lack of trust in the media are growing problems within our society. "Fake news" has been used in a multitude of ways in the last half a year and multiple definitions have been given. For instance, the New York Times defines it as "a made-up story with an intention to deceive". Measuring fake news or even defining it properly could very quickly become a subjective matter, rather than an objective metric. In its purest form, fake news is completely made up, manipulated to resemble credible journalism and attract maximum attention and, with it, advertising revenue. Despite all these shortcomings, several entities have tried to categorize fake news in different manners. However, in order to start addressing this problem, an indepth understanding of fake news and its origins is required. Only then one can look into the different techniques and fields of machine learning (ML), natural language processing (NLP) and artificial intelligence (AI) that could help us fight this situation."

Chapter 3

Proposed System

3.1. Features and Functionalities :

- Publicaly Available Dataset
- Fast checking of knowledge based news content
- Rumor Classification
- Detecting True facts from Multiple conflicting sources
- Helps to identify the True and False news.



Chapter 5

Software Requirements:

Front-end: HTML, CSS, JavaScript

The HyperText Markup Language or HTML is the standard markup language for documents designed to be displayed in a web browser. It can be assisted by technologies such as Cascading Style Sheets (CSS) and scripting languages such as JavaScript. Web browsers receive HTML documents from a web server or from local storage and render the documents into multimedia web pages. HTML describes the structure of a web page semantically and originally included cues for the appearance of the document. CSS is designed to enable the separation of presentation and content, including layout, colors, and fonts.[3] This separation can improve content accessibility; provide more flexibility and control in the specification of presentation characteristics; enable multiple web pages to share formatting by specifying the relevant CSS in a separate .css file, which reduces complexity and repetition in the structural content; and enable the .css file to be cached to improve the page load speed between the pages that share the file and its formatting. JavaScript is a high-level, often just-in-time compiled language that conforms to the ECMAScript standard. It has dynamic typing, prototype-based object-orientation, and first-class functions. It is multi-paradigm, supporting event-driven, functional, and imperative programming styles. It has application programming interfaces (APIs) for working with text, dates, regular expressions, standard data structures, and the Document Object Model (DOM).

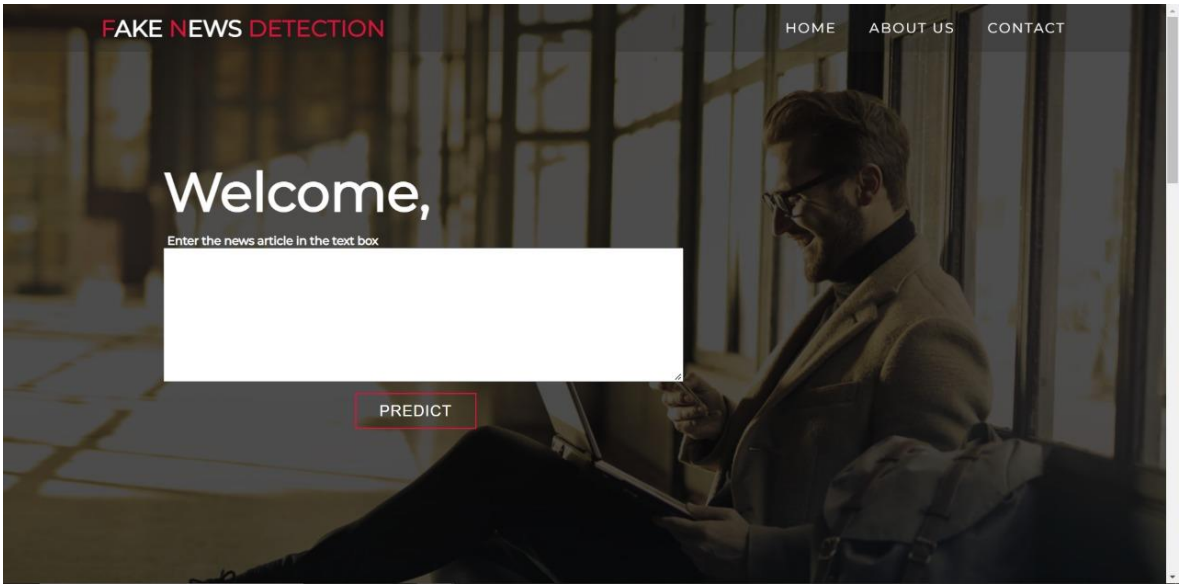
Back-end: Jupyter Notebook

The Jupyter Notebook is the original web application for creating and sharing computational documents. It offers a simple, streamlined, document-centric experience. program used to mix code, comments, and visualizations in an interactive document called notebook that can be shared, reused, and reworked in a web browser. Jupyter Notebook (formerly IPython Notebooks) is a web-based interactive computational environment for creating notebook documents. A Jupyter Notebook document is a browser-based REPL containing an ordered list of input/output cells which can contain code, text (using Markdown), mathematics, plots and rich media.

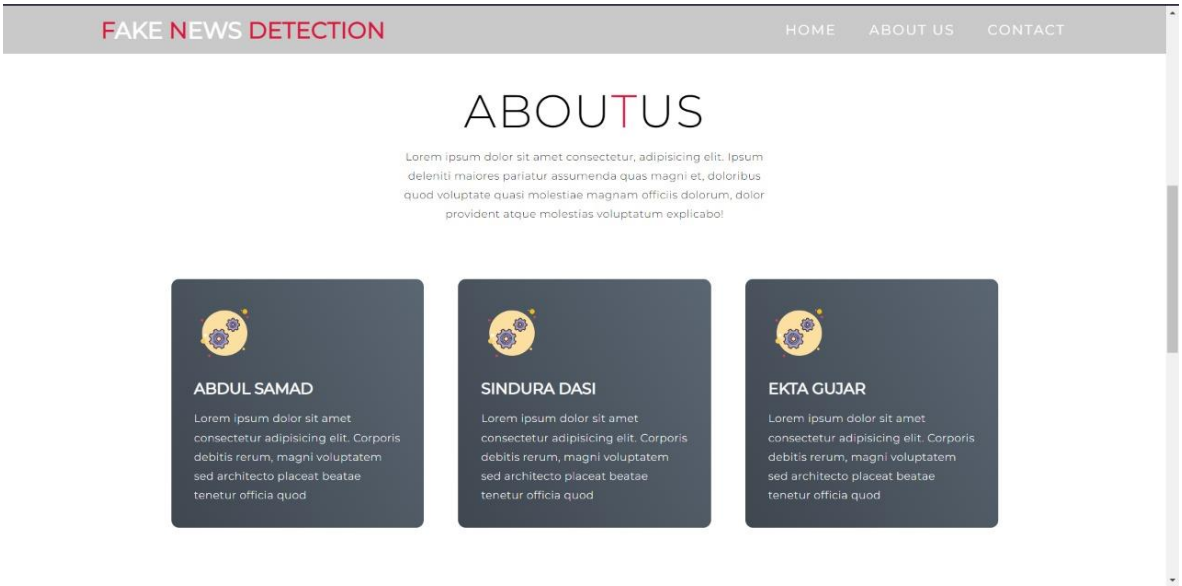
Chapter 6

Project Design

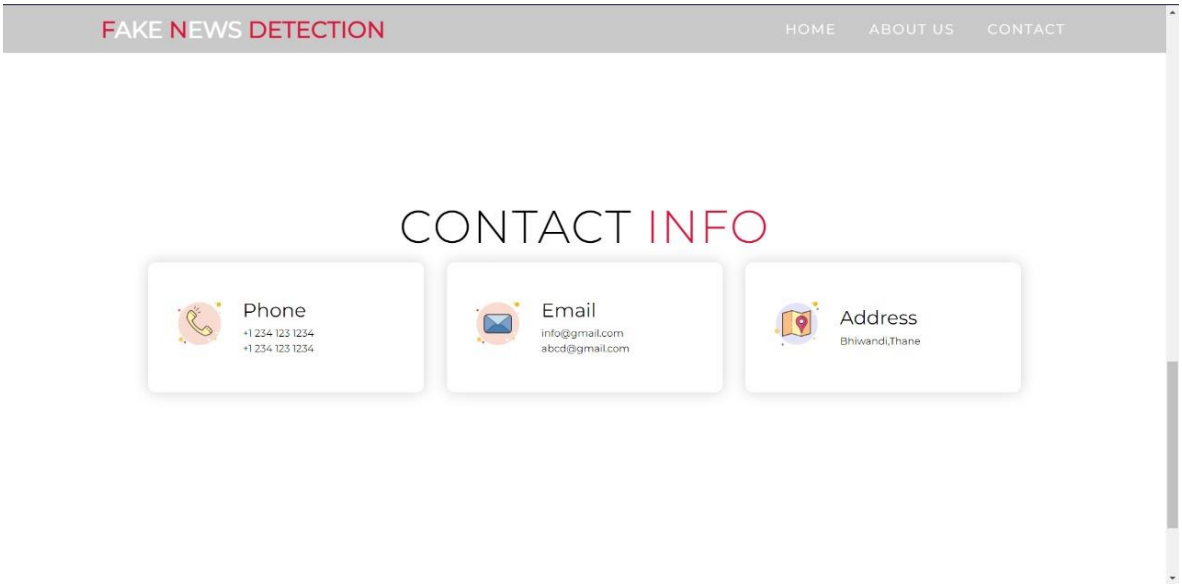
Home Page:



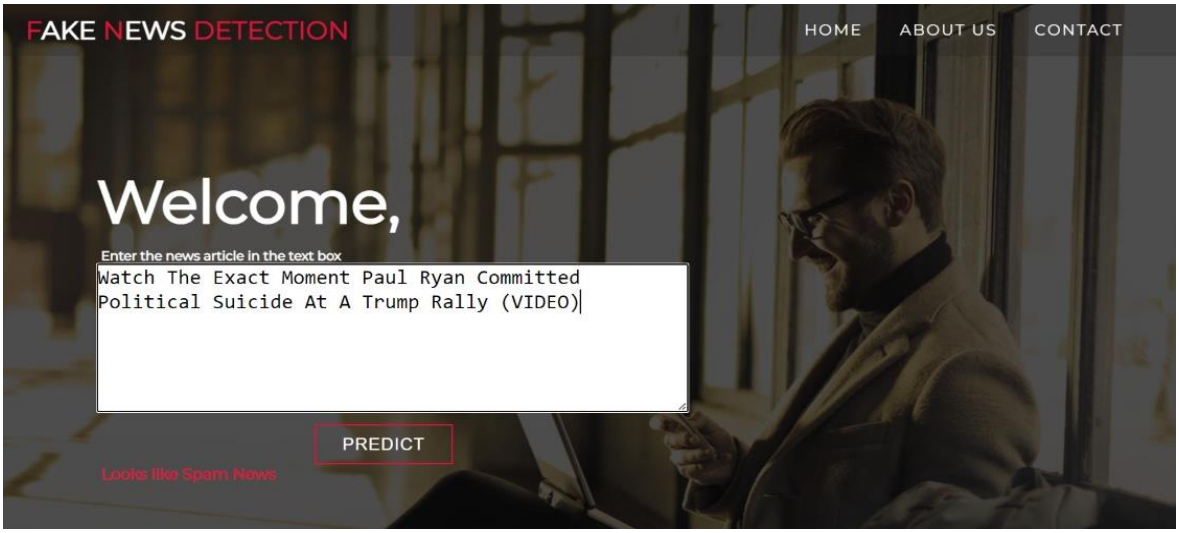
About Us Page:



Contact Info:



Prediction Page:



Jupyter Notebook :

```
import os
os.chdir("C:\Users\Abdul Samad\OneDrive\Desktop\Workspace\Fake news")
```

```
import pandas as pd
```

```
dataframe = pd.read_csv('news.csv')
dataframe.head()
```

```
x = dataframe['text']
y = dataframe['label']
```

```
x
```

```
0        Daniel Greenfield, a Shillman Journalism Fello...
1        Google Pinterest Digg Linkedin Reddit Stumbleu...
2        U.S. Secretary of State John F. Kerry said Mon...
3        — Kaydee King (@KaydeeKing) November 9, 2016 T...
4        It's primary day in New York and front-runners...
                               ...
6330     The State Department told the Republican Natio...
6331     The 'P' in PBS Should Stand for 'Plutocratic' ...
6332      Anti-Trump Protesters Are Tools of the Oligar...
6333     ADDIS ABABA, Ethiopia —President Obama convene...
6334     Jeb Bush Is Suddenly Attacking Trump. Here's W...
Name: text, Length: 6335, dtype: object
```

```
y
```

```
0        FAKE
1        FAKE
2        REAL
3        FAKE
4        REAL
         ...
6330     REAL
6331     FAKE
6332     FAKE
6333     REAL
6334     REAL
Name: label, Length: 6335, dtype: object
```

```
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import PassiveAggressiveClassifier
from sklearn.metrics import accuracy_score, confusion_matrix
```

```
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.2,random_state=0)
y_train
```

```
2402    REAL
1922    REAL
3475    FAKE
6197    REAL
4748    FAKE
       ...
4931    REAL
3264    REAL
1653    FAKE
2607    FAKE
2732    REAL
Name: label, Length: 5068, dtype: object
```

```
tfvect = TfidfVectorizer(stop_words='english',max_df=0.7)
tfid_x_train = tfvect.fit_transform(x_train)
tfid_x_test = tfvect.transform(x_test)
```

- max_df = 0.50 means "ignore terms that appear in more than 50% of the documents".
- max_df = 25 means "ignore terms that appear in more than 25 documents".

```
classifier = PassiveAggressiveClassifier(max_iter=50)
classifier.fit(tfid_x_train,y_train)
```

```
PassiveAggressiveClassifier(C=1.0, average=False, class_weight=None,
                            early_stopping=False, fit_intercept=True,
                            loss='hinge', max_iter=50, n_iter_no_change=5,
                            n_jobs=None, random_state=None, shuffle=True,
                            tol=0.001, validation_fraction=0.1, verbose=0,
                            warm_start=False)
```

```
y_pred = classifier.predict(tfid_x_test)
score = accuracy_score(y_test,y_pred)
print(f'Accuracy: {round(score*100,2)}%')
```

```
Accuracy: 93.69%
```

```
cf = confusion_matrix(y_test,y_pred, labels=['FAKE','REAL'])
print(cf)
```

```
[[575  40]
 [ 40 612]]
```

```
def fake_news_det(news):
    input_data = [news]
    vectorized_input_data = tfvect.transform(input_data)
    prediction = classifier.predict(vectorized_input_data)
    print(prediction)
```

```
fake_news_det1("""Go to Article
President Barack Obama has been campaigning hard for the woman who is supposedly going to extend his legacy four more years. The only problem with stumping for Hillary Clinton, however, is she€™s not exactly a candidate
```

Python

```
['FAKE']
```

```
fake_news_det1("""U.S. Secretary of State John F. Kerry said Monday that he will stop in Paris later this week, amid criticism that no top American officials attended Sunday€™s unity march against terrorism.""")
```

Python

```
['REAL']
```

```
fake_news_det('''U.S. Secretary of State John F. Kerry said Monday that he will stop in Paris later this week, amid criticism that no top American officials attended Sunday€™s unity march against terrorism.''')
```

Python

```
['REAL']
```

Chapter 7

Project Scheduling

Sr. No	Group Member	Time duration	Work to be done
1	Abdul Samad Ansari	1st February To 28th February	Reviewing few previously published research paper and starting with the pre-processing of dataset.
2	Sindura Dasi	1st March To 31st March	Implementing Front-end of the project
3	Ekta Gujar	1st April To 25th April	Performing comparison of algorithms and calculating accuracy.

Chapter 8

Conclusion

With the current scenario ongoing one can easily tell that in today's world nobody is anything less than a journalist who can just post what they intend to. There are no criteria on which their authenticity can be based. Also, there are some mischievous elements in this society that have successfully developed bots that send continuous spam and fake news to social media for their own benefit or just some fun. Social media has become a hub for fake news that needs to be corrected before this fake news render these social media useless. For such there are many researches going on all over the world emphasizing on various points in a news or information snippet to identify and inform of any fake news. We have hence summed up the different types of existing methods that can be used for fake news identification along with few of researches each. Then we also provided with few of what are the available datasets for fake news detection, some focusing on rumours, some on stance etc.

References

[1] Abror Abduvaliyev, Al-Sakib Khan Pathan, Jianying Zhou, Rodrigo Roman and WaiChoong Wong ,“On the vital Areas of Intrusion Detection Systems in Wireless Sensor networks”,IEEE Communications Surveys & Tutorials, Accepted For Publications, 2013-in press.

[2] H.H. Soliman, et al,“A comparative performance evaluation of intrusion detection techniques for hierarchical wireless sensor networks”, Egyptian Informatics Journal (2012) 13, 225238.

[3] Giannetsos Athanasios, “Intrusion Detection in Wireless Sensor Networks”, Master THESIS, Carnegie Mellon University, April 8, 2008.

[4] K.Fall and K.Varadhan,“The NS Manual”, http://www.isi.edu/nsnam/ns/doc/ns doc.pdf., 1 Feb 2014.

[5] Jae Chung and Mark Claypool,“NS by Example-Tutorial”, http://nile.wpi.edu/NS/overview.html , 1 Feb 2014.

[6] Network Simulator blog, http://Mohittahilani.blogspot.com , 1 Feb 2014. [7] AWK Script for NS2, http://mohit.ueuo.com/AWK-Scripts.html , 1 Feb 2014.

ACKNOWLEDGEMENT

This project would not have come to fruition without the invaluable help of our guide **Prof. Yaminee Patil**. Expressing gratitude towards our HoD, **Prof. Kiran Deshpande**, and the Department of Information Technology for providing us with the opportunity as well as the support required to pursue this project. We would also like to thank our teacher Prof. Nahid Shaikh who gave us her valuable suggestions and ideas when we were in need of them. We would also like to thank our peers for their helpful suggestions.