Mini Project report on

Fake News Detection

By

OTHNIEL ALEXANDER

(2018450035)

PRATHAMESH HIRE

(2018450019)

Under the guidance of

Internal Supervisor

Prof. PALLAVI THAKUR

Department of Master Of Computer Applications

Sardar Patel Institute of Technology

University of Mumbai

2019-20

CERTIFICATE OF APPROVAL

This is to certify that the following students

OTHNIEL ALEXANDER

(2018450035)

PRATHAMESH HIRE

(2018450019)

Have satisfactorily carried out work on the project entitled

"Fake News Detection"

Towards the fulfillment of Mini Project, as laid down by University of Mumbai during year 2019-20.

_____

Project Guide

PROJECT APPROVAL CERTIFICATE

This is to certify that the following students

OTHNIEL ALEXANDER

(2018450035)

PRATHAMESH HIRE

(2018450019)

Have successfully completed the Mini Project report on

"Fake News Detection"

which is found to be satisfactory and is approved

At

SARDAR PATEL INSTITUTE OF TECHNOLOGY,

ANDHERI (W), MUMBAI.

INTERNAL EXAMINER

EXTERNAL EXAMINER

Abstract

The explosive growth in fake news and its erosion to democracy, justice, and public trust has increased the demand for fake news analysis, detection and intervention. This project comprehensively and systematically reviews fake news research. The project identifies and specifies fundamental theories across various disciplines, e.g., psychology and social science, to facilitate and enhance the interdisciplinary research of fake news. Current fake news studies is reviewed, summarized and evaluated.

These studies focus on fake news from four perspective:

(1) the false knowledge it carries,

(2) its writing style,

(3) its propagation patterns, and

(4) the credibility of its creators and spreaders.

We characterize each perspective with various analyzable and utilizable information provided by news and its spreaders, various strategies and frameworks that are adaptable, and techniques that are applicable. By reviewing the characteristics of fake news and open issues in fake news studies, we highlight some potential research tasks at the end of this project report.

Objectives

Following are the objectives of the Web Application.

- It helps users to verify the news whether its fake or not.
- It helps them to understand the sources they must follow in order to refer to news from the right source.

List of Figures

List of Tables

1. Introduction

1.1. Problem Definition

We characterize each perspective with various analyzable and utilizable information provided by news and its spreaders, various strategies and frameworks that are adaptable, and techniques that are applicable. By reviewing the characteristics of fake news and open issues in fake news studies, we highlight some potential research tasks at the end of this project report.

1.2. Objective And Scope

Objectives of the system:

- The proposed system will be online system, so it will not be time-consuming.
- Users would easily be able to verify the news.
- Accuracy of the Truth value is generated.

Scope of the proposed system:

- The system has been developed to be simple and user-friendly. Therefore, even a non-technical user(non tech savvy) would be able to use the system effectively and without any issues.
- Since it is upgrade-able, further enhancements are possible.
- The system can maintain speed and accuracy.

1.3. Existing System

- Users are operating through Whatsapp messages (communication and discussion) and news apps providing information about the news integrity.
- There's no existing system which verifies the news whether they are fake or not/
- These Web Applications are less user friendly and do not present the results in form of graphs telling the user stories.

1.4. Proposed System

- User can verify news articles or create their own articles and graphs and verify them.
- To know the approx. truth value of the news that is generated.
- To have an aesthetic UI pleasing to the eyes of the user.

1.5. System Requirements:

Hardware Requirements:

Processor	Dual Core Processor or Above
RAM	512MB RAM
Storage	10GB Hard Disk Space for smooth run

Software Requirements:

Back-End:

- Django with pickle

Front-End:

- JetBrains Pycharm, VS-Code and Jupyter notebook

Others:

- JavaScript Frameworks like EJS, Passport, Route Handlers.
- Bootstrap v3.3.7
- Cascading Style Sheets (CSS)
- Font Awesome
- Hyper Text Mark-up Language (HTML)

2. Literature Survey

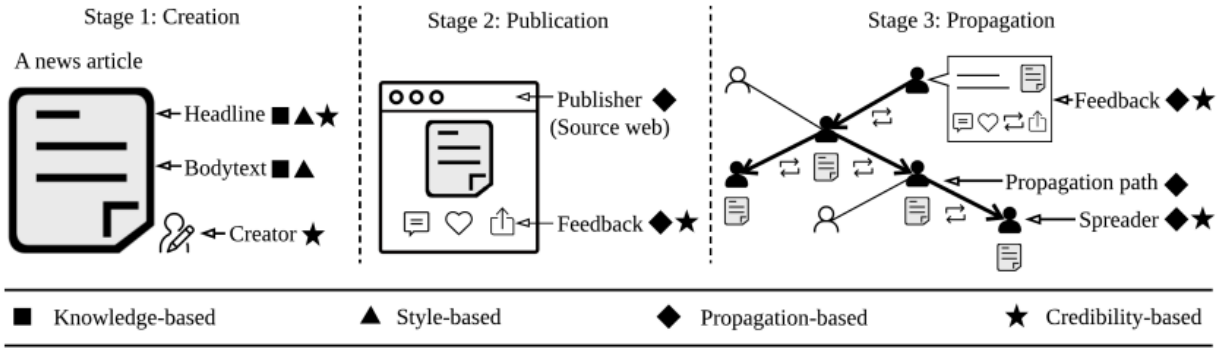
Fundamental human cognition and behavior theories developed across various discipline such as psychology, philosophy, social science, and economics provide invaluable insights for fake news analysis. Firstly, these theories introduce new opportunities for qualitative and quantitative studies of big fake news data, which to date, has been rarely available. Secondly, they facilitate building well-justified and explainable models for fake news detection and intervention, as well as introducing means to develop datasets that provide "ground truth" for fake news studies. We have conducted a comprehensive literature survey across various disciplines and have identified twenty well-known theories that can be potentially used to study fake news.

These theories are provided in Table 2 along with short descriptions. These theories can be used to study fake news from three different perspectives:
(I) style: how fake news is written, (II) propagation: how fake news spreads, and (III) users: how users engage with fake news and the role users play (or can play) in fake news creation, propagation, and intervention. In the following, we detail how each perspective and its corresponding theories facilitate fake news analysis.

I. **Style-based Fake News Analysis.** These fundamental theories address how fake news content and writing style can be different from true news. For instance, reality monitoring indicates that actual events can be expressed by higher levels of sensory-perceptual information.

II. **Propagation-based Fake News Analysis**. Epidemic models, which can mathematically model the progression of an infectious disease, can be used or extended to model fake news propagation. However, selecting or developing proper epidemic models relies on making reasonable assumptions. Some real-world phenomena can help simplify these assumptions and in turn, simply such epidemic models. Examples includes backfire effect, conservatism bias and Semmelweis reflex, which indicate that "fake news is incorrect but hard to correct" i.e., it propagates with minimum resistance.

III. **User-based Fake News Analysis**. These theories investigate fake news from a user's perspective, considering how users engage with fake news and what roles users play in fake news creation, propagation and intervention. In sum, users that participate in fake news activities can be grouped into (i) malicious users, who intentionally create and/or propagate fake news motivated by some benefits and (ii) normal users, some of whom spread fake news along with malicious users. These normal users are often called naïve users as their engagement is unintentional and driven by self-influence or social influence, e.g., naïve users can participate in fake news spreading



Fake News can be studied with respect to four perspectives:

(i) knowledge-based, focusing on the false knowledge in fake news;
(ii) style-based, concerned with how fake news is written;
(iii)propagation-based, focused on how fake news spreads, and;
(iv)credibility-based, investigating the credibility of its creators and spreaders.

Each perspective targets some of fake news characteristics (i.e., its authenticity, intentions, or whether it is news) by using some types of information (news-related or social-related) using different techniques.

JetBrains PyCharm IDE for writing, running and debugging code:

- PyCharm is an integrated development environment (IDE) used in computer programming, specifically for the Python language. It is developed by the Czech company JetBrains. It provides code analysis, a graphical debugger, an integrated unit tester, integration with version control systems (VCSes), and supports web development with Django.

- Coding assistance and analysis, with code completion, syntax and error highlighting, linter integration, and quick fixes
- Project and code navigation: specialized project views, file structure views and quick jumping between files, classes, methods and usages
- Python refactoring: including rename, extract method, introduce variable, introduce constant, pull up, push down and others.

3. Software Requirement Specification (SRS) & Design:

3.1 Introduction:

Product Perspective:

The software described in this SRS is the software for a complete Fake News Detection System. The system merges various hardware and software elements and further interfaces with external systems. Thus, while the software covers the majority of the system's functionality, it relies on a number of external interfaces for persistence and unhandled tasks, as well as physically interfacing with humans.

Since this is a machine learning and NLP project we need to train the model using datasets which will enable the model to perform various actions.

3.2 Overall Description:

3.2.1 Product Functions :

The product functions include:

1. Authentication:
Users/investors are required to Sign-up and Log-in. Users can view articles but cannot create or discuss without signing in. System redirects users to login page if he/she tries to review,create,edit,discuss the article without log-in into the system.

2. **Viewing News(Graphs):**
System facilitates users viewing of various articles created by other users/investors with or without authentication.

3. Creating News:
4. Investors can create articles(graphs) and provide full details of it-name, month,year ,commodity , etc.

5. Reviewing News:
Investors/users can provide feedback and can review the articles created by themselves or other investors. This facility demands log-in into the system, means authenticate himself/herself.

5. Users/investors will get get the news rate for the desired month.Also they will be provided with the average ETF (1 ETF = gram of news) .

6. Edit/Delete News:
Only the owner can edit/delete its content(articles, review, graphs). For security purpose apart from owner, admin can delete users review, articles and graphs.

Software Interfaces:

The interfaces setup an infrastructure by Incorporating required libraries Sklearn, Numpy, Pandas, Matplotlib, Pickle and Seaborn.

The interfaces then perform dataset preparation which incorporates data from different sources, processes the data into pandas dataframe, the data is normalized by using the min_max function.

The web application communicates with the server and plots the graph based on the user inputs.It does this, as the model is trained to perform such actions by using the past data.

User interfaces: If the user is not a first-time user, he/she should be able to see the articles page directly when the application is opened. Here the user chooses the type of search he/she wants to conduct. Every user should have a profile page where they can edit their e-mail address, phone number and password. Also, the user can set the Web application to his/her preferred language.

3.2.2 User characteristics:-

- Django Admin: Admin has the full access to the system which means he is able to manage any activity with regard to the system. He is the highest privileged user who can access to the system.

Key functions of admin are:

- Allocate resources
- Manage administrator charges
- Manage News Articles

User:

Interacts with the systems most often to supply service to customers.

- Checks for News details and verifies its truth

Constraints:

- System is wirelessly networked with an encryption.
- Only administrator can access the whole system.
- Each user should have initial id and password.
- Database is password protected.

Assumptions and dependency:

- Each user must have a valid user id and password.
- Server must be running for the system to function.
- Users must log in to the system to access any article.
- Only the administrator can delete the articlesalong with the owner(user who created the article).

3.3 Specific Requirements:

3.3.1 Functional Requirements:

• The system will allow user to view News whether they are True or False

3.3.2 Non-functional Requirements:

1. **Usability**: The system provides a good and easy to use GUI for users to interact with the system.
2. **Security**: The users are required to authenticate themselves before impacting system. Impacting system means creating any articles, discuss, review.
It is the facility provided by system to prevent unauthorized access.
3. **Performance**: The system must have high performance rate when executing user's input and should be able to provide response within short period of time. System is persistent. Data is directly stored onto the database and gives quick flash message to user.
4. **Availability**: The system must be available for access 24 hours, 7 days a week. Also in occurrence of any major malfunctioning, the system should online again as soon as possible. Also system will give the same performance even if many users log-on at the same time.
5. **Error Handling**: Errors must be considerably minimized and appropriate error message that guides the user to recover from error should be provided.

4. Project Analysis And Design

4.1 Detailed Life Cycle of the Project

Rapid Application Development (RAD) Model is used for designing this system. Unlike Waterfall methods, RAD emphasizes working software and user feedback over strict planning and requirements recording. In this model, we go through prototyping and collecting feedback iteratively until satisfactory product is developed. RAD requires system to following:

Phase I: Figure out the requirements
Phase II: Build Prototypes
Phase III: Get user feedback
Phase IV: Repeat Phase II and III (Sprints)
Phase V: Testing

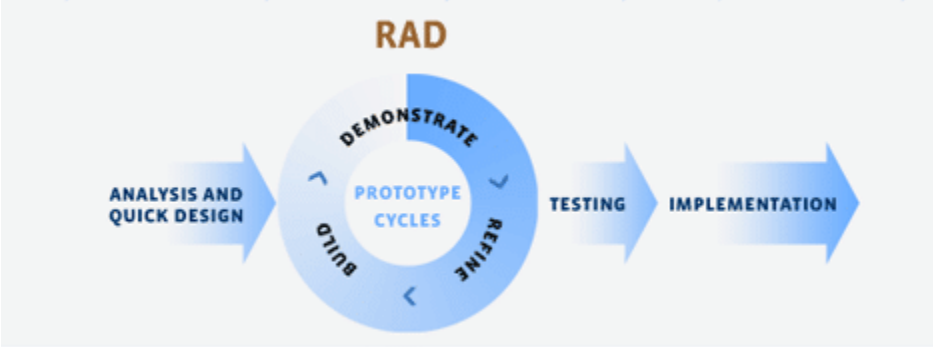


Fig 4.1 Diagrammatic Representation of RAD Model

The process undertaken this project are as follows :

- 1. Set Up Infrastructure
  - a) I Python Notebook
  - b)Incorporate required libraries Sklearn, Numpy, Pandas, Matplotlib and Seaborn
  - c)Git Project Organization
- 2. Dataset Preparation
- a)Incorporate data from different sources
- b)Process the data into pandas dataframe

 c)Normalize the data using Sklearn’s MinMax Function
 d)Timeseries Split with n_splits=10
 3. Develop Benchmark Model
a) Set up basic decision tree regressor with default parameters as benchmark model.
 4. Develop Solution Model & improve using GridsearchCV
 5. Ensemble top three performing models
 6. Evaluate all the solution models and benchmark model & Plot the result
 7. Refine the models using Feature selection
 8. Develop Benchmark, solution and ensemble models
 9. Evaluate and compare the results with Original feature models
 10. Plot, analyze and describe the results for report.

News Rates Prediction system has gone through following sprints:

10.1 Setting up infrastructure
10.2 Dataset preparation
10.4 Testing Front-end
10.4 Evaluate solution model and plot results
10.5 Testing the security and quality for robustness.

4.2 Use Case Diagram with Report

Use Case Description

Actors:
1. User
2. Organizer
3. Administrator

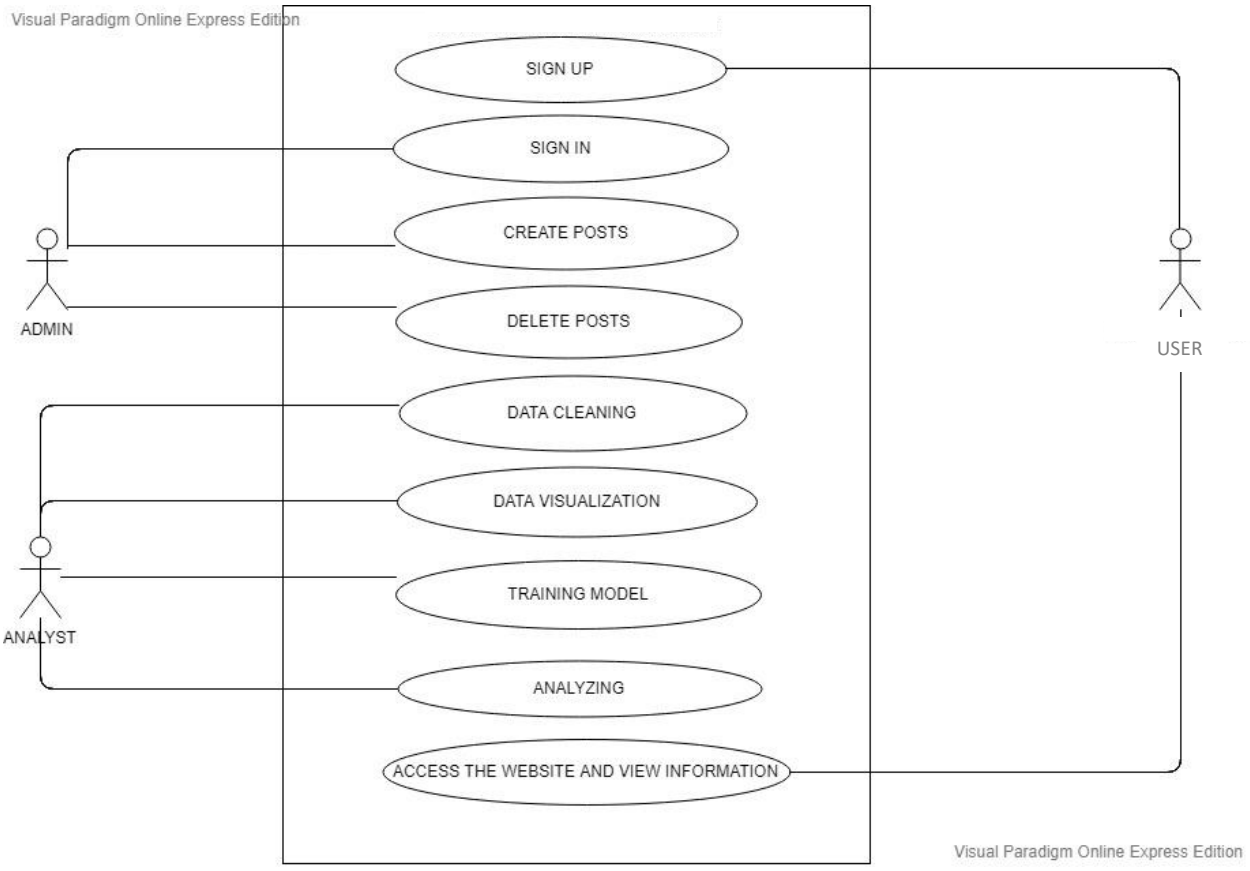


Fig 4.2 Use Case Diagram

Use Cases:

1. Sign-in
2. Create Posts
3. Delete Posts
4. Data Cleaning
5. Data Visualization
6. Training Model
7. Analyzing
8. Access the Website and View Information

Table 4.2: Use Case Tables

Use Case Id	1
Use Case Name	Sign-In
Actor	User, Admin
Pre-Condition	They must register themselves first.
Post-Condition	Users can view and review articles, Admin can control the activity.
Flow of events	Register, login, access system

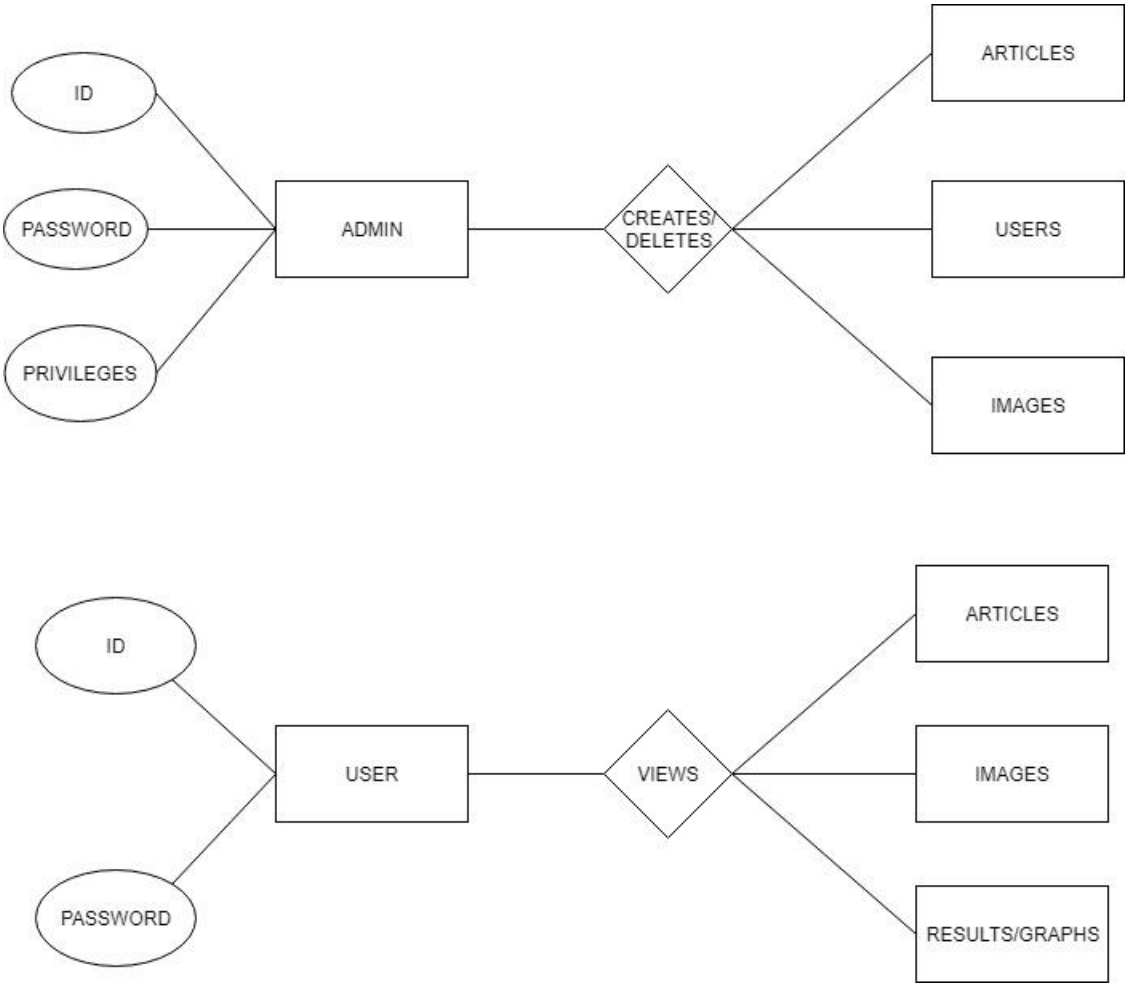
Use Case Id	2
Use Case Name	Create Posts
Actor	Admin
Pre-Condition	Sign-in is necessary.
Post-Condition	Admin can can create, review and can delete.
Flow of events	Login, Create Articles
Use Case Id	3
Use Case Name	Delete Posts
Actor	Admin
Pre-Condition	Login is must.
Post-Condition	Admin can delete their reviews. Other users, can view it.
Flow of events	Login, make review, and delete it.

Use Case Id	4
Use Case Name	Analyzing
Actor	Analyst
Pre-Condition	User must login
Post-Condition	Model Analyzed
Flow of events	Data cleaning,Data Visualization,Training Model,Analyzing

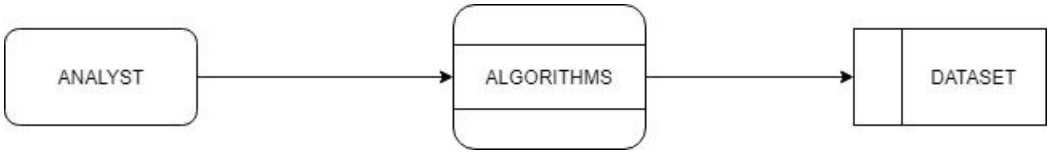
Use Case Id	5
Use Case Name	Access website and view information
Actor	User
Pre-Condition	Login is must.
Post-Condition	User as well as admin can view the information but only admin can delete it
Flow of events	Login, create, view, delete.

4.3 Database Schema/Design

4.3.1. E-R Diagram



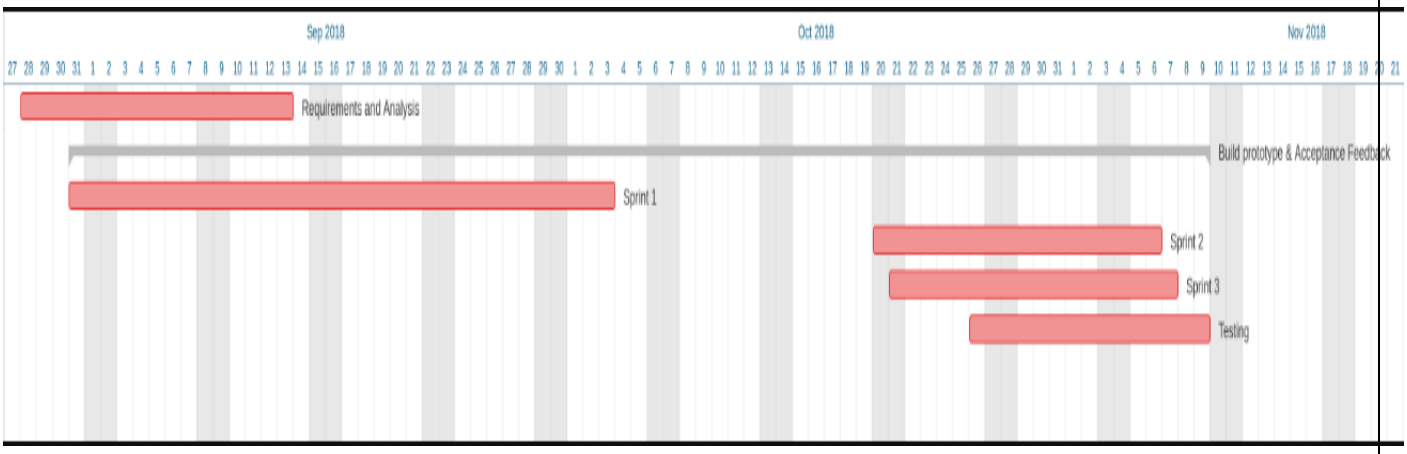
4.3.2 DFD Diagrams



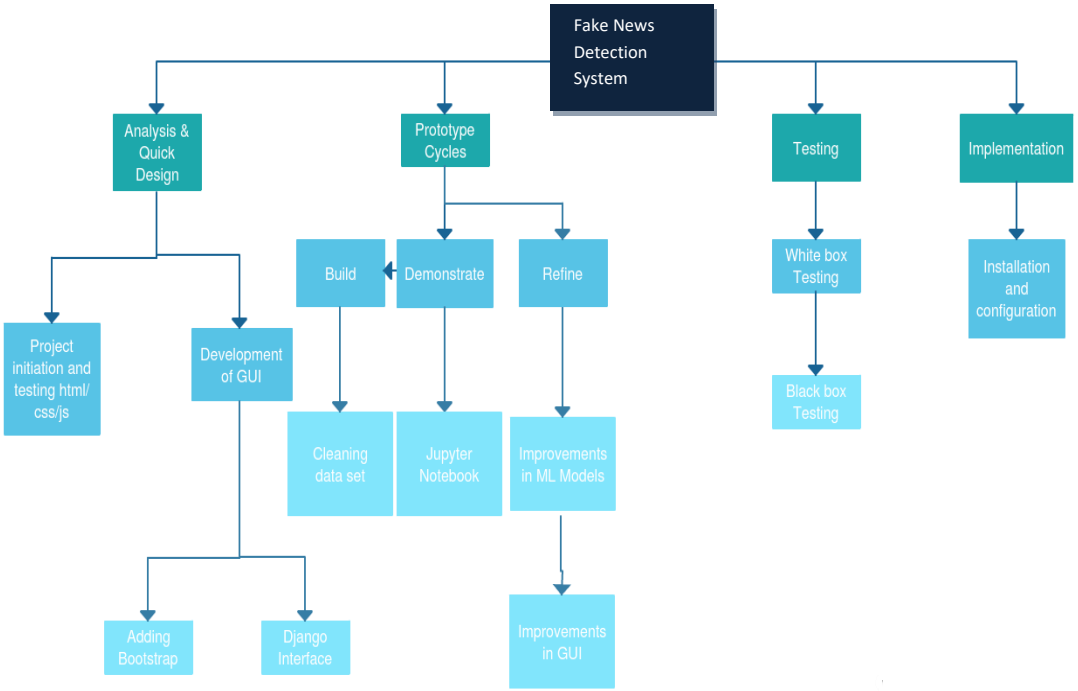
5. Project Implementation and Testing

5.1 Gantt Chart

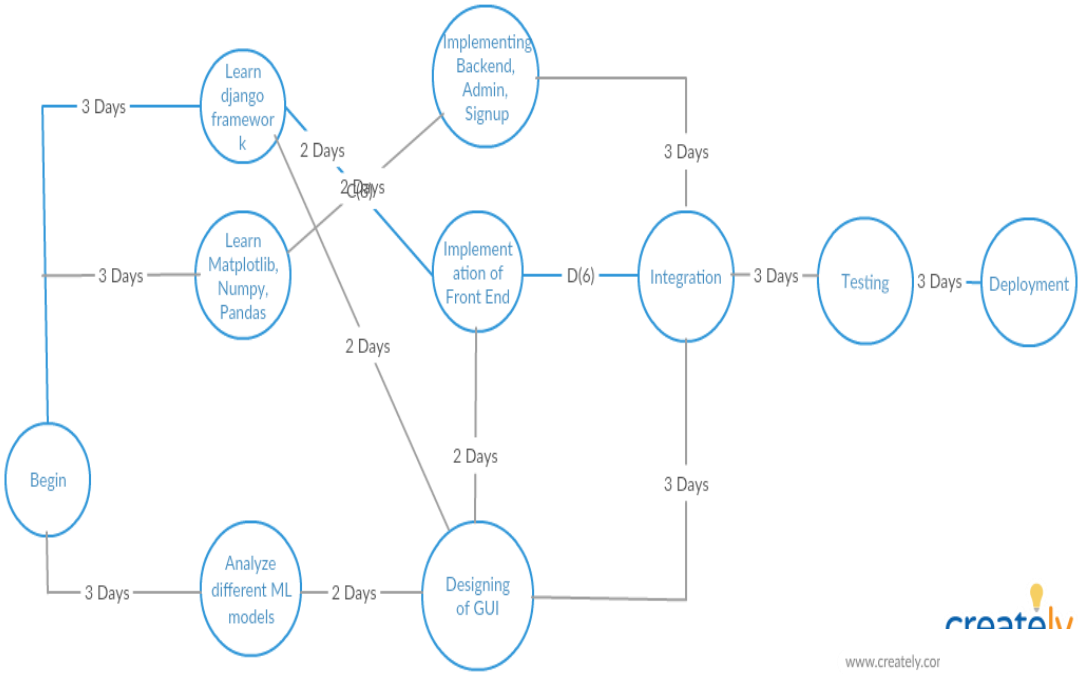
Fake News Detector	Sub Tasks	Assignee	Est. Hours	Start Date	Due Date	Task Prog.
0 Requirements and Analysis		Unassigned	-	28/Aug	13/Sep	0%
Build prototype & Acceptance Fee…		Unassigned	-	31/Aug	09/Nov	0%
2 Sprint 1		Unassigned	-	31/Aug	03/Oct	0%
3 Sprint 2		Unassigned	-	20/Oct	06/Nov	0%
4 Sprint 3		Unassigned	-	21/Oct	07/Nov	0%
5 Testing		Unassigned	-	26/Oct	09/Nov	0%



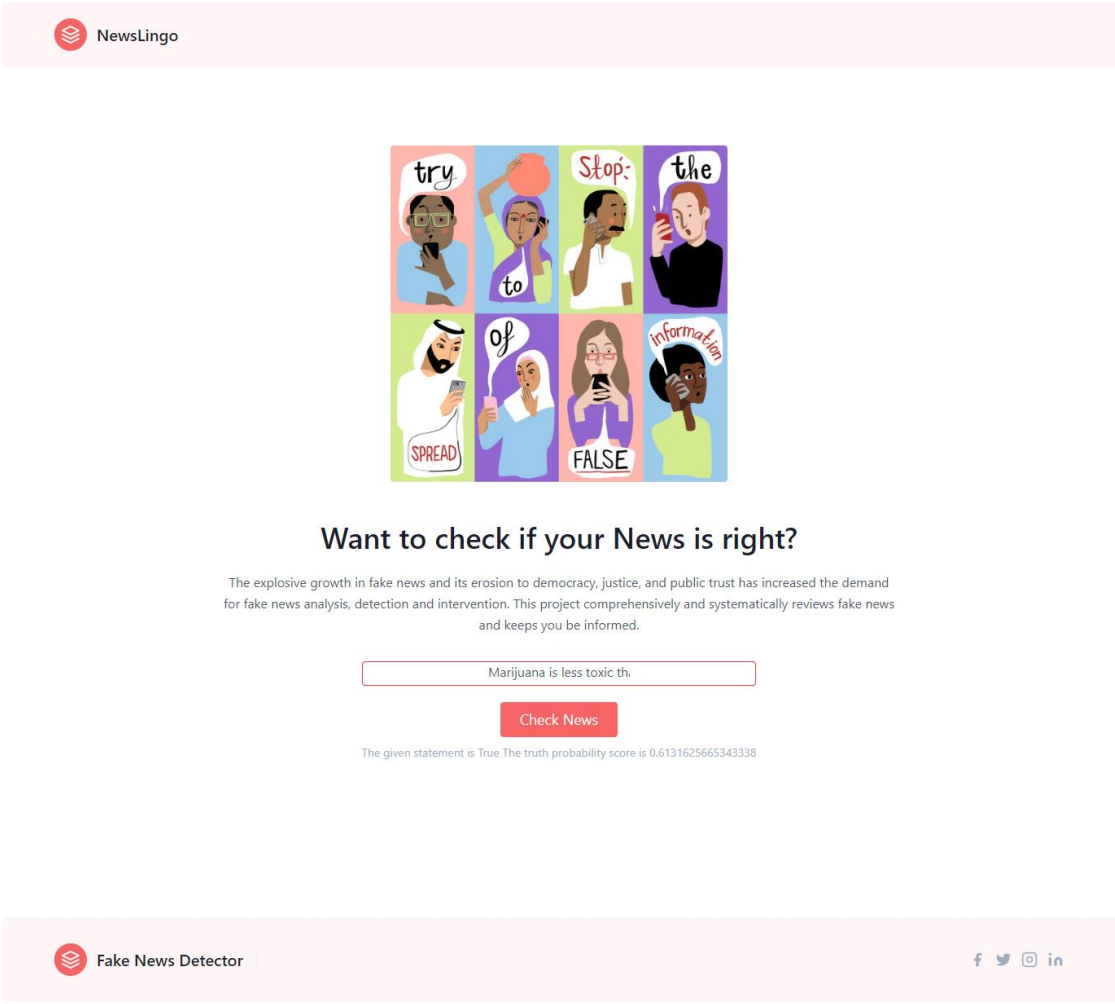
5.2 Work Breakdown Structure



5.3 PERT Chart



5.4 Snapshot:


NewsLingo
try
to
Stop:
the
SPREAD
of
FALSE
information
Want to check if your News is right?
The explosive growth in fake news and its erosion to democracy, justice, and public trust has increased the demand for fake news analysis, detection and intervention. This project comprehensively and systematically reviews fake news and keeps you be informed.
Marijuana is less toxic th
Check News
The given statement is True The truth probability score is 0.6131625665343338
Fake News Detector

5.4 Test Cases

Table 5.4.1: Test Cases for Log-in

Test Case Id	Test Case Name	Test Data	Expected Output	Actual Output	Result
1	User enters News Text	Trump is Black	False	True	Fail
2	User enters News Text	Obama is Black	True	True	Pass

6. Limitations

With the increasing popularity of social media, more and more people consume news from social media instead of traditional news media. However, social media has also been used to spread fake news, which has strong negative impacts on individual users and broader society. In this project, we explored the fake news problem by reviewing existing literature in two phases: characterization and detection.

In the characterization phase, we introduced the basic concepts and principles of fake news in both traditional media and social media. In the detection phase, we reviewed existing fake news detection approaches from a data mining perspective, including feature extraction and model construction. We also further discussed the datasets, evaluation metrics, and promising future directions in fake news detection research and expand the field to other applications.

7. Future Enhancements

- To implement the existing system to list all the fake news which were generated from the previous user queries.
- Highlighting links where the fake news exists.

8. References

Fan Yang, Yang Liu, Xiaohui Yu, and Min Yang. Automatic detection of rumor on sina weibo. In Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics, page 13. ACM, 2012.

Robert B Zajonc. Attitudinal effects of mere exposure. Journal of personality and social psychology,

Robert B Zajonc. Mere exposure: A gateway to the subliminal. Current directions in psychological science,

Arkaitz Zubiaga, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Procter. Detection and resolution of rumours in social media:

Data set:

William Yang Wang. ” liar, liar pants on fire”: A new benchmark dataset for fake news detection.