DANYAL HAIDER
01-134191-011
OSAMA NISAR
01-134191-100

News Mapping System

Bachelor of Science in Computer Science

Supervisor: Ms. Umarah Qaseem

Department of Computer Science
Bahria University, Islamabad

January 2023

© Osama Nisar And Danyal Haider, 2022

Certificate

We accept the work contained in the report titled "Newsverse", written by Mr. Danyal Haider AND Mr. Osama Nisar as a confirmation to the required standard for the partial fulfillment of the degree of Bachelor of Science in Computer Science.

Approved by . . . :

Supervisor: Ms. Umarah Qaseem

_____

Internal Examiner: Name of the Internal Examiner (Title)

_____

External Examiner: Name of the External Examiner (Title)

_____

Project Coordinator: Ms. Maryam Khalid Multani

_____

Head of the Department: Dr. Arif ur Rahman (Head of Department / Sr. Associate Professor)

_____

December 30^{th}, 2022

Abstract

The Internet has turned our world on its head. It has transformed communications to the point where it is now our best medium of communicating. We use the Internet in almost everything we do.People can get a lot of benefits from it, if it is used correctly.The purpose of this project in today's advance technological period is to aware people about situation and help them to get awareness because now days its very challenging for public to get authentic news in short period of time. Also, our project will help common public to get aware of news that are not showed on mainstream media by using different resources.Moreover,it will help law enforcement agencies to get aware about situation in mean time and also it is first GIS news Web application which shows news of Pakistan as markers on map which help people to get awareness in very less time.News analyst can also get benefits by plotting news on graphs to see trends about what is happening in the country.Authenticated Users will also be able report news in the application to help their community.Moreover by using machine learning techniques it will also be able to classify news into different categories so that it helps users to filter news so that they only see news of their interest.In general this project is based upon live news mapping system by using GIS and machine learning techniques so that it accurately perform its functionalities.It will help people of every field who are more concerned about situation of country and will help them to get awareness.

Acknowledgments

We are enormously indebted to Ms. Umarah Qaseem, Sr. Lecturer, Department of Computer Science, for providing us the opportunity to do this project under her constant supervision as well as for providing necessary information regarding the project and also for their support in completing the project.

DANYAL HAIDER
OSAMA NISAR
Islamabad, Pakistan

October 2022

"We think someone else, someone smarter than us,
someone more capable, someone with more resources will solve that problem.
But there isn't anyone else."

Regina Dugan

Contents

Abstract i

1 Introduction 1
- 1.1 Problem Statement . . . 1
- 1.2 Project Objectives . . . 1
- 1.3 Newsverse is providing . . . 3
- 1.4 Project scope . . . 3
- 1.5 Summary . . . 4

2 Literature Review 5
- 2.1 Sentiment Map . . . 5
- 2.2 NewsStand . . . 6
- 2.3 PhotoStand . . . 7
- 2.4 NewsStand 2 . . . 8
- 2.5 Implementation of Website News Map Function Based on Map World . . . 9
- 2.6 The Local News Map . . . 10
- 2.7 Comparison . . . 11

3 Requirement Specifications 13
- 3.1 Existing System . . . 13
- 3.2 Proposed System . . . 14
- 3.3 Feasibility Study . . . 15
- 3.4 Tools and Technologies . . . 15
- 3.5 Methodology . . . 16
  - 3.5.1 Collecting the data and prepossessing of data . . . 16
  - 3.5.2 Removing the duplication . . . 16
  - 3.5.3 Classifying the data . . . 16
  - 3.5.4 Mapping of the news . . . 16
- 3.6 Requirement Specifications . . . 16
  - 3.6.1 Resource Requirements . . . 16
  - 3.6.2 Functional Requirements . . . 17
  - 3.6.3 Non-Functional requirements . . . 17
- 3.7 Use Case . . . 18
- 3.8 Use Case Description . . . 24
  - 3.8.1 Login . . . 24
  - 3.8.2 Registration . . . 25
  - 3.8.3 View news map . . . 26

3.8.4 Report news . . . 27
3.8.5 View news graph . . . 28

4 Design **29**
4.1 System Architecture . . . 29
4.2 Design Methodology . . . 29
4.2.1 News Data retrieval . . . 29
4.2.2 Removal of duplicate news data . . . 30
4.2.3 Preprocessing of data . . . 30
4.2.4 Classification . . . 30
4.2.5 Mark news on map . . . 30
4.3 Flowchart Diagram . . . 30
4.4 Entity Relationship Diagram . . . 32
4.5 Sequence Diagram . . . 33
4.5.1 View map sequence diagram . . . 33
4.5.2 Advance search sequence Diagram . . . 34
4.5.3 View chart sequence Diagram . . . 35
4.6 Activity diagrams . . . 36

5 System Implementation **41**
5.1 System implementation . . . 41
5.2 System Architecture . . . 41
5.2.1 System architecture diagram . . . 42
5.3 System Internal Component . . . 43
5.3.1 Register . . . 43
5.3.2 Login . . . 43
5.4 Navbar . . . 44
5.5 Map . . . 44
5.6 Select Category . . . 45
5.7 News feed . . . 46
5.8 Homepage . . . 47

6 System Testing and Evaluation **49**
6.1 Introduction . . . 49
6.2 Graphical User Interface Testing: . . . 49
6.2.1 Usability testing: . . . 51
6.2.2 Performance testing: . . . 52
6.2.3 Security testing: . . . 53
6.3 Test Cases: . . . 54
6.3.1 Logging in application: . . . 54
6.3.2 Signing up in application: . . . 54
6.3.3 View news map: . . . 55
6.3.4 Report news: . . . 55
6.3.5 View news trends: . . . 56
6.4 Our Collected Dataset: . . . 56
6.4.1 Dataset collection: . . . 56
6.5 Visualization of data count in each category . . . 57

6.6 Preprocessing of data: . . . 58
6.7 Evaluation matrix: . . . 58
6.8 Confusion matrix: . . . 59

7 Conclusions **61**
7.1 Conclusion . . . 61
7.2 Future Enhancements . . . 61

A User Manual **63**
A.1 Sign in screen: . . . 63
A.2 Signup screen: . . . 64
A.3 Main screen: . . . 64
A.4 Select category option: . . . 65
A.5 Newsfeed: . . . 65
A.6 Report news Screen: . . . 66
A.7 News analysis: . . . 67

References **69**

List of Figures

1.1 Map of our application . . . 2

2.1 Sentiment Map . . . 6
2.2 NewsStand . . . 7
2.3 PhotoStand . . . 8
2.4 NewsStand 2 . . . 9
2.5 Map World . . . 9
2.6 Local News Map . . . 10

3.1 Google World Map . . . 14
3.2 Newsverse . . . 15
3.3 Use Case Login . . . 19
3.4 Use Case Registration . . . 20
3.5 Use Case View News Map . . . 21
3.6 Use Case Report News . . . 22
3.7 Use Case View News Graph . . . 23

4.1 Flowchart Diagram . . . 31
4.2 Entity Relationship Diagram . . . 32
4.3 View Map sequence diagram . . . 33
4.4 Advance search sequence diagram . . . 34
4.5 View chart sequence diagram . . . 35
4.6 Activity diagram for viewing news map . . . 37
4.7 Activity diagram for advance search . . . 38
4.8 Activity diagram for view news report . . . 39

5.1 System Architecture . . . 42
5.2 User Register . . . 43
5.3 User sign in . . . 43
5.4 Navbar . . . 44
5.5 Map . . . 44
5.6 Select region . . . 45
5.7 News feed . . . 46
5.8 Home Page . . . 47

6.1 Category of data . . . 57

A.1 User Manual: Sign in screen . . . 63

A.2 User Manual: Sign up screen 64
A.3 User Manual: main screen 64
A.4 User Manual: Select category option 65
A.5 User Manual: Newsfeed screen 65
A.6 User Manual: Report news Screen 66
A.7 User Manual: Report news Screen 67

List of Tables

2.1 Comparison of Literature Review 11

3.1 Tools and technologies 15
3.2 Login use case 24
3.3 Registration use case 25
3.4 View news map use case 26
3.5 Report news use case 27
3.6 View news graph use case 28

6.1 Test case GUI 50
6.2 Usability testing 51
6.3 Performance testing 52
6.4 Security testing 53
6.5 Test case: Logging in Application 54
6.6 Test case: Signing up in application 54
6.7 Test case: View news map 55
6.8 Test case: Report news 55
6.9 Test case: View news trends 56
6.10 Data count 57
6.11 Evaluation matrix 58
6.12 Confusion matrix 59

Acronyms and Abbreviations

GIS	Geo Information System
OOP	Object Oriented Programming
ML	Machine Learning
UML	Unified Modeling Language
UC	Use Case
HTML	Hypertext Markup Language
CSS	Cascading Style Sheets

Chapter 1

Introduction

Technology is automating the world in every domain and every aspect of life, easing the burden on humans as a result. We can flourish our lives if we use technology wisely. Proposed project is about scraping the news data from the different news platforms and then mark the News data on the map with respect to the location of occurrence. A web-based GIS application will be obtained to show the news and analyze data in graphical manner. Which will allow users to check news without going to different newspapers, news websites and search for news. It will cover the news about current situation going in the country or any specific area of country also user will be able to see trends of technological advancements, terror activities, crimes, or other related activities of specific region of the country also it will have different news categories and it will only show related news once the category is selected. Users will be also able to report any type of news or crimes of specific area and all reports will be in different section. Overall, in general it's a map-based news system which will save user's time and it will display the extracted news on map in structured manner.

1.1 Problem Statement

As we know it takes time to get knowledge about situation of whole country and sometimes people are left unaware of some important news. checking newspaper or watching news headlines takes a lot of time. This system will allow users to get aware of situation of whole in very less time. Sometimes when something like calamity, disaster or any suspicious activity occurs we don't know about the exact location so it will help to check the location of without going to other map apps.

1.2 Project Objectives

Newsverse is a web-based news mapping system, which provides nationwide news about:

- Business
- Technology
- Sports
- Natural Disasters
- Terrorism
- Crimes

Also this proposed application comes with following objectives:

- Provide a unified platform which should cater all the latest news as well as previous news over a map.
- Scrap news from different news websites which are authentic.
- Store all of the collected data in a database.
- Display news over a map with specific marker.
- Marker shape are related to type of news.
- Provide Crime rate of any area over a graph.
- Show stats of terror activities and crimes of every province which will help to compare different provinces and areas in terms of safety

Figure 1.1: Map of our application

1.3 Newsverse is providing

- Easy to learn and use.
- Efficient algorithm used to make this application efficient.
- Provide a unified platform which should cater all the latest news as well as previous news over a map.
- This system geographical information system to deliver good quality of news.
- Easy news mapping approach using GIS system for visual inspection.
- This platform is beneficial in providing awareness to common public.
- Date range and news type filter to analyze data in different perspectives.
- Proposed system will scrape data from news website only.
- System would have only news data of past as well as future.
- In initial stage system will only target Pakistan.
- Users will be able to report any suspicious activity and it will have different section.

1.4 Project scope

This application is provide a unified platform where users will be aware about situation nearby his location/home. This application will help people to get information about situation of there country. Overall, in general, it is an opportunity for people to raise awareness about what is going in society and care about their safety. It can also help news analysts to see patterns of news of their interest. The theme of this web-based application is to facilitate

General Public:

- We will keep those who are concerned about the ever-changing situation in their area informed through news alerts.

Third Party: This application also provide facility to those who are impacted directly or indirectly by the continuously changing situation in their area like

- Local agencies
- News analyst
- Civil protection institutions

1.5 Summary

The technology revolution refers to the advancement of technology from radio, Television to mobile and computers. Today majority of android mobile users and computer users want to there tasks done on their mobile and computers like buying seats, payments, communication and watching news of there own interest. Right now there is no web-based application available which is providing news over a map. That's why we develop a web-based application which provide a news visually over a map. It will also help users to see the crime ratio in there city and Provence.

Chapter 2

Literature Review

The literature review provides a clear perception of the standards being used in the past and the present. It gives us an insight of the business models that currently exist and changes with the passage of time. Due to the evolution of the Web Applications the many developers have faced many discontinuities which may include features which application offered with respect to demand of the time but cannot satisfy today's demands. Newsverse is time saving and very easy to use. Some of previous web-based news projects are listed below

2.1 Sentiment Map

The demand for websites that offer higher-quality material is rising along with the volume of content available on the web. Today, high-quality data is gathered and integrated by online sites from a variety of websites. Users no longer need to visit multiple websites in order to access different types of information. These websites include news portals like Google News. This system is a news visualizations tool that displays a news site's Internet reliability trend. If users are aware of the news site's trend, they can evaluate the legitimacy of each news story. This method analyses each news article's sentiments to identify the website's trend. The average feelings of news articles posted about a topic on each website are used to determine the trend of news websites. Using the sentiment dictionary produced by the previously suggested method, the sentiment values of news articles are determined. Our suggested system extracts news articles having the keywords from each specified news source when a user enters one or more topic keywords. The sentiment scores of news stories from each news website are likewise calculated by this algorithm, along with their average score for each sentiment. Of this research, we refer to the map as a sentiment map. The algorithm then constructs a bar graph from the four average values in each news site

and arranges all of the bar graphs on a global map using Google Map API. Users may quickly and easily comprehend trends across numerous websites on a particular topic by using the sentiment map. In this paper, they show how to create a sentiment map. [1]

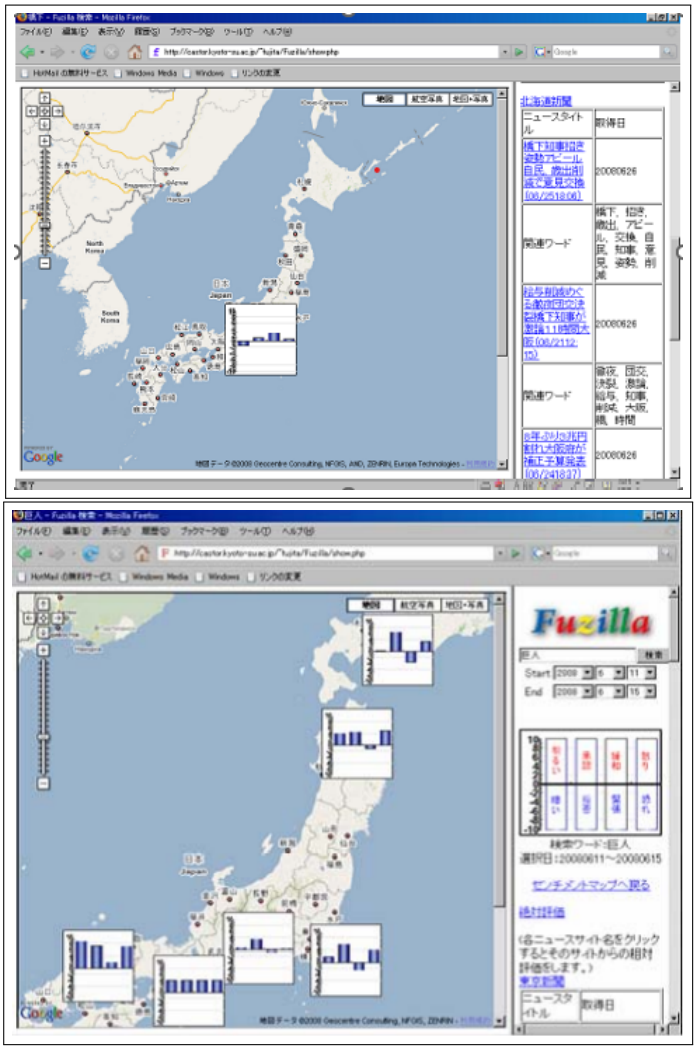

Figure 2.1: Sentiment Map

2.2 NewsStand

Every day, NewsStand collects the most recent news from thousands of different news sources and processes several of hundreds of new articles. Some few regions of NewsStand could use some advancement. NewsStand will generally display a geographic bias close to the areas about which news stories are typically composed, requiring more uniform news inclusion. Furthermore, the framework currently only cycles English-language articles, which could be improved by including articles and news sources in other languages.More semantic clues from the record, such as landmarks and waterways, may be utilized by

NewsStand's geotagged to aid in correct geotagging. Later, they will use geology, in addition to textual terms, to work on grouping news articles. We will consider how to use grouping to determine the geographic scope of the news provider (for example, the geographic area of the newspaper), as well as how to use it to work on geotagging and neighbourhood news inclusion. Finally, we intend to include media such as agent photos, recordings, and sound cuts on the guide itself. As a result, we're looking into methods for determining the best delegate image for a collection of news articles. [6]

Figure 2.2: NewsStand

2.3 PhotoStand

The demand for websites that offer higher-quality material is rising along with the volume of content available on the web. Today, high-quality data is gathered and integrated by online sites from a variety of websites. Users no longer need to visit multiple websites in order to access different types of information. These websites include news portals like Google News. This system is a news visualizations tool that displays a news site's Internet reliability trend. If users are aware of the news site's trend, they can evaluate the legitimacy of each news story. This method analyses each news article's sentiments to identify the website's trend. The average feelings of news articles posted about a topic on each website are used to determine the trend of news websites. Using the sentiment dictionary produced by the previously suggested method, the sentiment values of news articles are determined. Our suggested system extracts news articles having the keywords from each specified news source when a user enters one or more topic keywords. The sentiment scores of news stories from each news website are likewise calculated by this algorithm, along with their average score for each sentiment. Of this research, we refer to the map as a sentiment map.

The algorithm then constructs a bar graph from the four average values in each news site and arranges all of the bar graphs on a global map using Google Map API. Users may quickly and easily comprehend trends across numerous websites on a particular topic by using the sentiment map. In this paper, they show how to create a sentiment map. [3]

Figure 2.3: PhotoStand

2.4 NewsStand 2

They used the strength of spatial synonyms to analyze the design objectives and performance of the News Stand system for reading the headlines on the Web while viewing a map. News Stand provides an example of how removing geographic information from news items reveals a previously hidden layer of data that can help readers comprehend news events through time and geography. The letters NEWS stand for north, east, west, and south. Systems like News stand will be able to find appealing applications in other knowledge fields as a result of the increasing availability of geotagged content on the Internet. Sentiment analysis, for instance, as well as hot-spot analysis based on news, tweets, or other data stream sources, can reveal how a specific news story is interpreted by people in different nations or languages. Additionally, Newsstand makes a contribution to the developing discipline of computational journalism.[4]

Figure 2.4: NewsStand 2

2.5 Implementation of Website News Map Function Based on Map World

This research paper is in chines language so first we converter it in English language in order to get useful information for our project. In this project they Event, time and space which are the three elements of news. Traditional news puts more emphasis on reporting the process and time of event, but a little description on its position. Aiming at the problem, this paper applies Map World to news page through using HTML and JavaScript language on the basis on MyEclipse software platform, by which way it makes up the shortcoming of fuzzy spatial information successfully. The paper introduces the code part of news map page in detail. At the end of document, considering the simplicity and ease of the function, it implements one-click generate code of website news map and news information's query function.[5]

Figure 2.5: Map World

2.6 The Local News Map

This study addresses the literature on the geography of news, as well as the efficacy of mapping and other types of data visualisation as tools for supporting trans-disciplinary research. The authors show how news geography maps may be integrated with other data to illustrate their value for multidisciplinary research that improves journalism studies. The project's growing database of coded local news items is used to construct maps that reflect the spatial patterns of local news coverage supplied by various mainstream and ethnic news providers. Users may use these maps to analyse the diversity of stories, the volume of news, and the spatial patterns of coverage created by local news organisations, and then integrate the results with other data.[2]

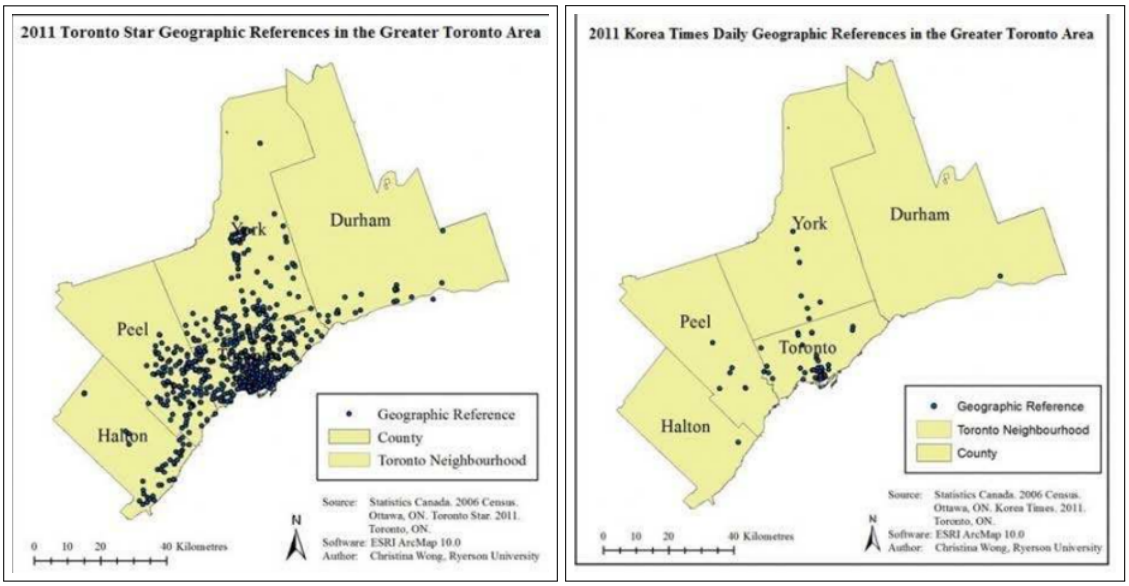

Figure 2.6: Local News Map

2.7 Comparison

Table 2.1: Comparison of Literature Review

NAME	**Title**	**Year Of Publish**	**Features**	**Country**
Yukiko Kawai	Sentiment Map	2012	A sentiment map is a cutting-edge method for visualizing news website statistics that helps readers identify differences in feelings among many websites for the same topic keyword they have looked for.	Japan
1.Benjmin Teitler 2.Michael Lieberman 3.Danele Panzzo 4.Hannan Samet	NewsStand	2008	Where did news happen?	USA
1)Hannan Samet 2)Marco Adelfio 3)Brendan Fruin 4)Michael Lieberman 5)Jaigan Sankaranarayanan	PhotoStand	2013	Using a Natural Language Processing (NLP) technique makes it easier to identify things (people, places, and organizations) in news articles.	Maryland

1)HANAN SAMET 2)JAGAN SANKARA-NARAYANAN 3)MICHAEL D.LIEBERMAN 4)MARCO D. ADELFIO 5)BRENDAN C.FRUIN 6)JACK M.LOTKOWSKI 7)DANIELE PANOZZO 8)JON SPER-LING 9)BEN-JAMIN E.TEITLER	Reading News with Maps by Ex-ploiting Spatial Synonyms	2014	The New Stand map query interface keeps track of the output from more than 10,000 RSS news sources and connects articles with the places they refer to within moments of pub-lishing.	USA
1)Lei Tang 2)Yi Zhu Kunwang Tao 3)Yong Feng	Implementation of Website News Map Function Based on Map World	2013		Beijing
1)April Lindgren, 2)Christina Wong	The Local News Map	2012		Columbia

Chapter 3

Requirement Specifications

In the following chapter, we have described the existing systems which show news on maps but with minimal functionalities in it. In our proposed system section we have clearly described what our system is about and how it is different from the existing systems. This chapter also contains the functional and non-functional requirements of our system and use cases as well.

3.1 Existing System

Technology has made users very smart and demanding. The user today prefers short cuts to save their time. Therefore, many applications are introduced for the ease in news to facilitate them. The existing system is made by a google called google world news map in which:

- System shows world news on map and it only get news updates from BBC through API.
- It only focuses on international political and terrorism news with no past records.
- It's features the conflicts in different regions which are already defined by them.
- User cannot see news of their interest
- No functionality of showing news headlines.
- This system doesn't show data in graphical manner like representation in graphs, charts, or tables.

Figure 3.1: Google World Map

3.2 Proposed System

In the proposed system we are developing a news web application in which:

- Users are able to see news over a map.
- News is placed over a cities where it belongs to.
- It is a web-based news mapping system, which provides nationwide news about business, technology, sports, natural disasters, terrorism, and crimes on a map.
- It gets news and alerts for most areas in Pakistan.
- This news data will be scrapped from news articles from different local news websites.
- This system also has its own news tip service, allowing users to report any suspicious activities in their area only if they are authenticated.
- It will also show stats of terror activities and crimes of every province which will help to compare different provinces and areas in terms of safety.

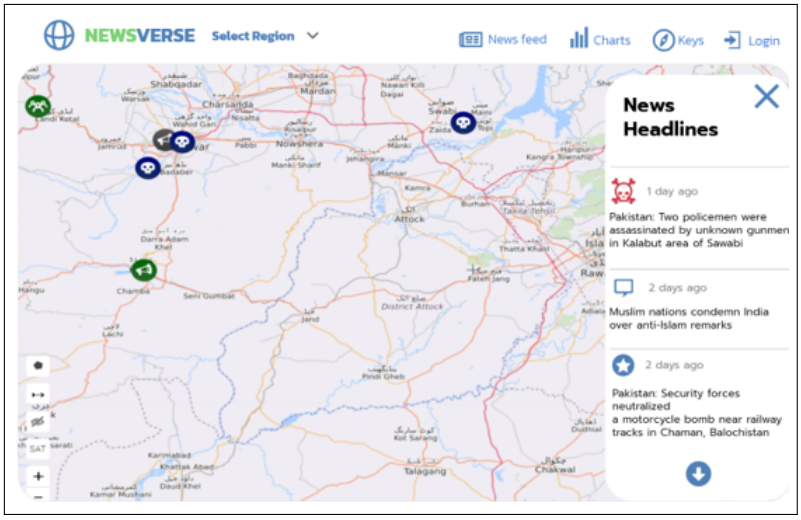

Figure 3.2: Newsverse

3.3 Feasibility Study

The idea of this project is to scrap news data from different platforms, perform classification of the news data in structured manner using NLP and show it on map. It will classify the data accurately as possible but in field of Machine learning getting 100 percent accurate results is nearly to impossible. However it will alert the people about situation.

3.4 Tools and Technologies

Table 3.1: Tools and technologies

Technology	Version	Rationale
python	3.8	Programming
MongoDB	5.1	Database
HTML	5	Web Development
CSS	3	Stylesheet
Javascript	ES5 or ES6	Scripting Language
Django	3.0	Backend
VS code	1.65	IDE height

3.5 Methodology

This project will employ the agile technique for development of web application. Project will be divided into different modules so that series of development cycles will be used to break down a complex tasks.
Modules are given below:

3.5.1 Collecting the data and prepossessing of data

News data will be scraped from different news channel and after that we will be cleaning that data using different techniques and then we will do the prepossessing to get data ready for the NLP classifiers.

3.5.2 Removing the duplication

After classifying the news data removal of duplication of news will be start because as mentioned above data will be scraped from multiple news websites or other authentic resources so it has chances that we will be getting duplicate news data so for that again we have to use ML algorithms to remove duplication.

3.5.3 Classifying the data

After the prepossessing of the data, we will be using different NLP algorithms to classify the data into different categories like political, crime, sports, business, technology, terrorism etc.

3.5.4 Mapping of the news

After the removal of duplication, the data will be mapped on map using location coordinates (longitude and latitude).

3.6 Requirement Specifications

The software requirements specification outlines the functions and performance standards for our project. The features that the product must provide is also described.

3.6.1 Resource Requirements

In order to run this system, it requires internet with at least 1Mb/s of speed. Users should have 4 Gb ram in their devices. It will be responsive web application so it will run on mobile phone, Tablets, and PC. Also, users should use latest browsers like chrome, Firefox, or MS edge.

3.6.2 Functional Requirements

The system should:

- Crawl through and extract data from news feeds.
- Classify news into different categories and save it in the database.
- Extract information from news data.
- Display extracted information on the map.
- Allow a user to view map.
- Allow user to use view news of every defined category.
- Allow user to make an account.
- Allow a user to request the report.

3.6.3 Non-Functional requirements

The interface should be simple to operate and explore. The system should only require occasional maintenance.

3.6.3.1 Usability

- The system should allow user to view map on a single click.
- The system should allow user to view news of his interest without getting confused or any difficulty.

3.6.3.2 Performance

- The system should generate results with high accuracy.
- The system should schedule scraper twice a day to keep to information up to date.

3.6.3.3 Security

- The system will be secured. Users will authorized after their authentication so that not every body have privilege to access or manipulate the important information.
- Access management and role management will be the basic goal.
- Security misconfigurations will be avoided.
- The system should Establish encrypted link between the server and browser.

3.6.3.4 Compatibility

- The system will be compatible with all browsers by using cross-browser compatible libraries and frameworks

3.6.3.5 Scalability

- The system should be able to handle multiple users at the same time.
- The system should be able to show news from multiple platforms to multiple users at same time.

3.7 Use Case

A use case is a group of actions and instructions that describe how the system, and the actors will interact to produce a certain outcome. Use cases and actors illustrate what the system performs and how the actors engage with it in use-case diagrams, but they do not describe how the system functions internally.

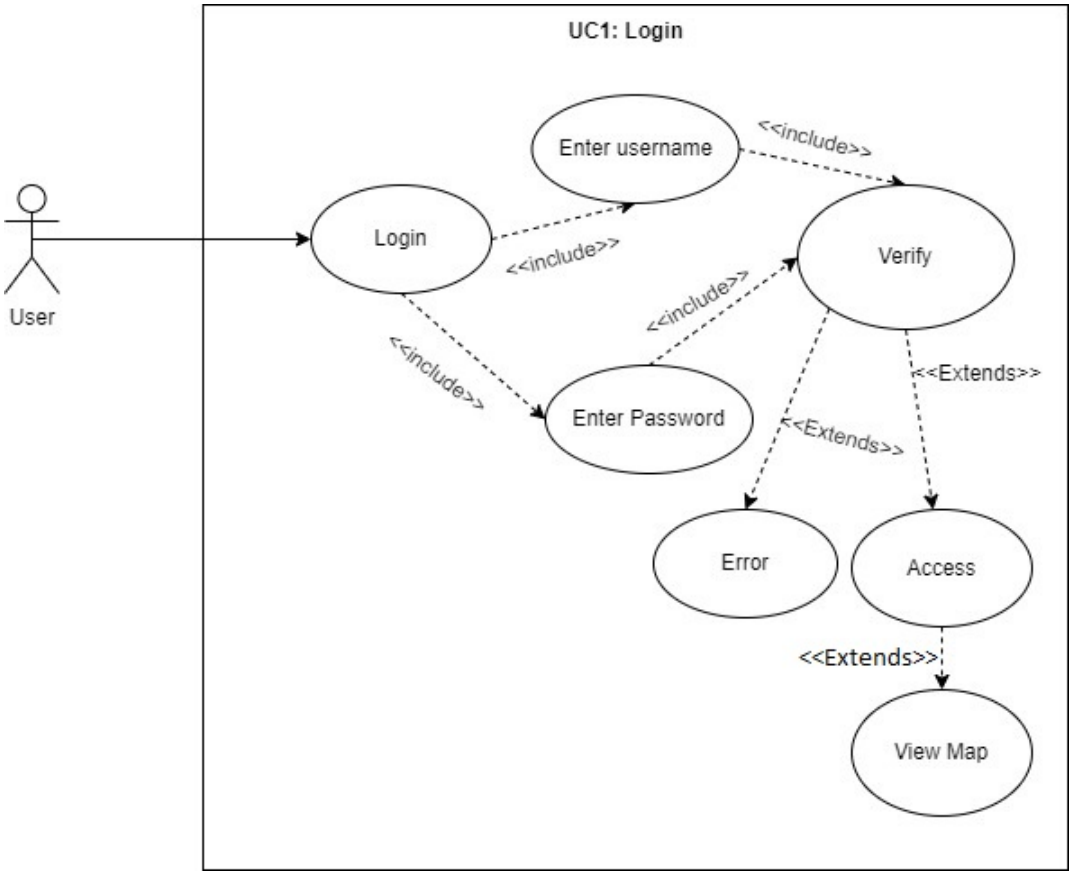

Figure 3.3: Use Case Login

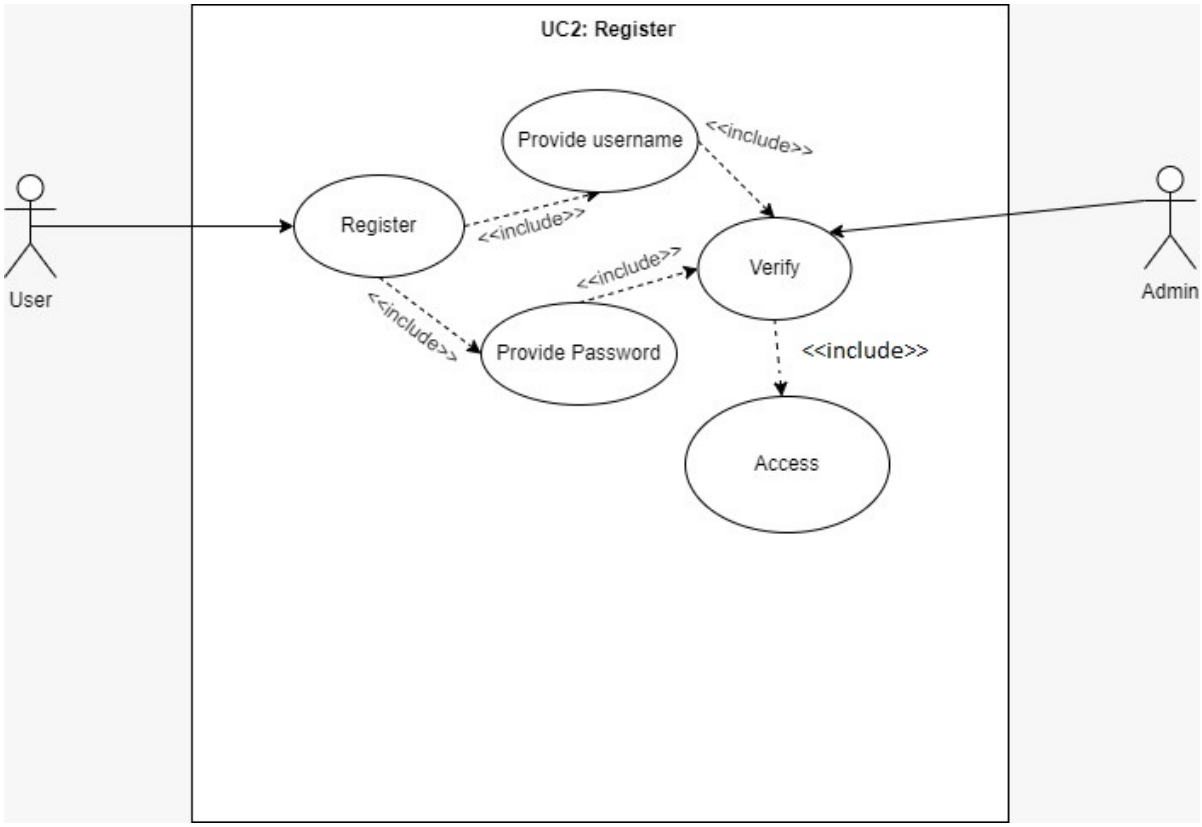

Figure 3.4: Use Case Registration

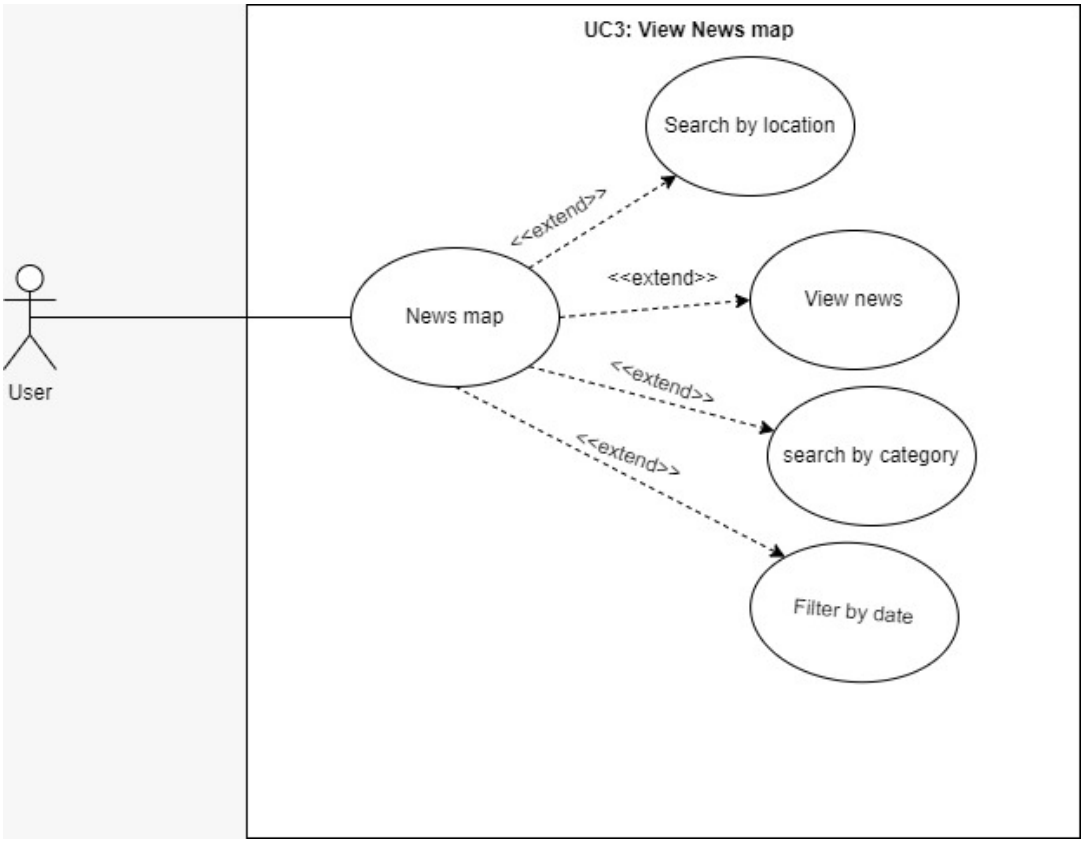

Figure 3.5: Use Case View News Map

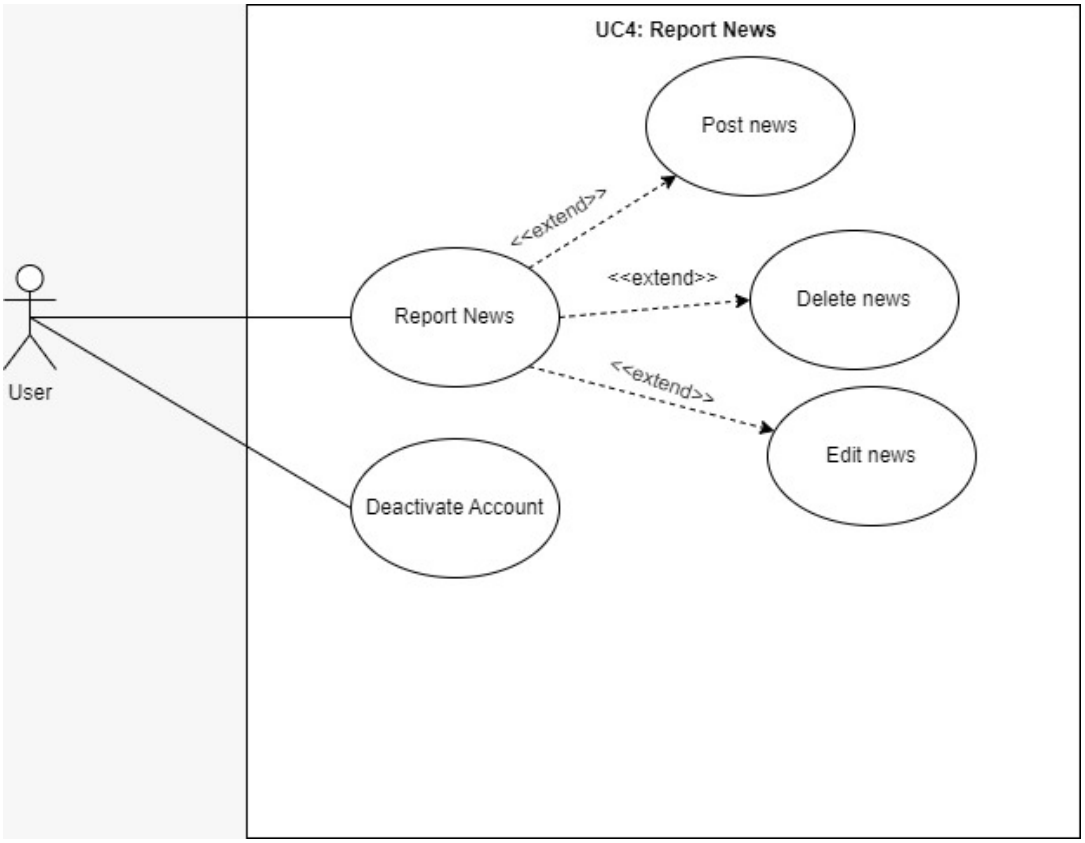

Figure 3.6: Use Case Report News

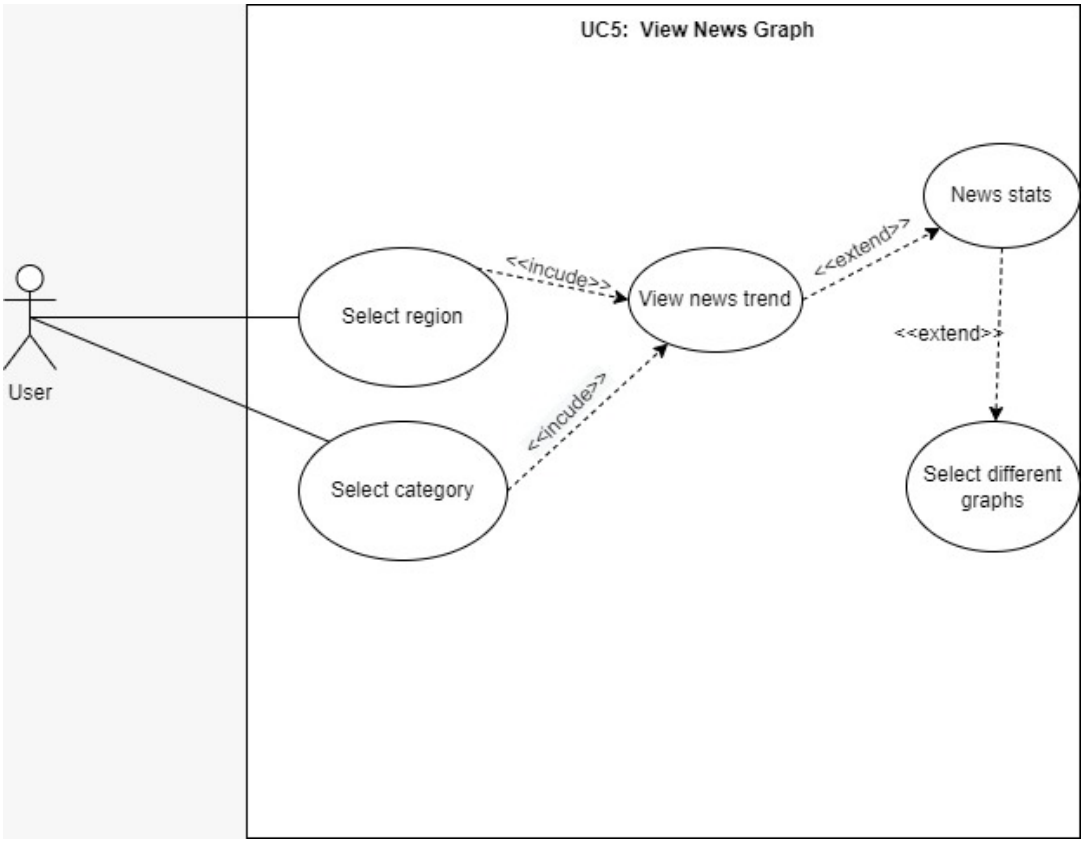

Figure 3.7: Use Case View News Graph

3.8 Use Case Description

3.8.1 Login

Table 3.2: Login use case

Login	
Use Case ID	UC1
Scope	Newsverse
Primary Actor	User
Stakeholders & Interests	User wants to login to post new
Pre-Conditions	User has registered email and password.
Post-Conditions	The user now able to post news
Main Success Scenario	
Actor	**System**
• Users enter credentials in app and go to homepage	• The system validates credentials and display homepage
Alternate Scenario	If user not registered, it would ask again for credentials.

3.8.2 Registration

Table 3.3: Registration use case

Register	
Use Case ID	UC2
Scope	Newsverse
Primary Actor	User, Admin
Stakeholders & Interests	User wants to register to access the feature post news
Pre-Conditions	User must have valid email account or google account
Post-Conditions	After providing details user should be provided secure access to the post news feature with same login credentials.
Main Success Scenario	
Actor	**System**
• Users enter credentials in app and their account will be registered and post news feature displayed.	• The system verifies that the user is entering correct data. After successful registration it maintain an open account. The system provide access to the user through a secure channel.
Alternate Scenario	The system may notify the user in case of incorrect information is entered and registration was not successful.

3.8.3 View news map

Table 3.4: View news map use case

View News Map	
Use Case ID	UC3
Scope	Newsverse
Primary Actor	User
Stakeholders & Interests	User wants to view news on map
Pre-Conditions	User have active internet connection
Post-Conditions	After run application, user able to view news map, user can also be able to view news by selecting category.
Main Success Scenario	
Actor	**System**
• The user selects the news by using news feeds section. • The user can post news. • The use can view news on map.	• System provides latest news on map using various news sources.
Alternate Scenario	If no latest news is available, then system will be shown past news.

3.8.4 Report news

Table 3.5: Report news use case

Report News	
Use Case ID	UC4
Scope	Newsverse
Primary Actor	User
Stakeholders & Interests	User wants to report news.
Pre-Conditions	User must login to his account by providing credentials.
Post-Conditions	Reported news by user is successfully saved and displayed on map.
Main Success Scenario	
Actor	**System**
• The users enter credentials in app and post news option is shown in header. • The user can post news by entering details of news.	• The system verifies that the user is entering correct information and then save it into database and display it on map.
Alternate Scenario	The system may notify the user in case of incorrect information is entered.

3.8.5 View news graph

Table 3.6: View news graph use case

Report News	
Use Case ID	UC5
Scope	Newsverse
Primary Actor	User
Stakeholders & Interests	User wants to view graph
Pre-Conditions	User must have active internet connection.
Post-Conditions	Graph is created successfully and showed to the user
Main Success Scenario	
Actor	**System**
• The user can select graph option from the header. • The user can select the type of graph	• The system asks the type of graph. • System generate graph by using news data.
Alternate Scenario	-

Chapter 4

Design

4.1 System Architecture

Users of the Newsverse App will be greeted by a splash screen when they first launch it, and depending on their current state, the app will then ask them to sign up or sign in. The user will be welcomed on the main page after signing in. where they can view news map of their choice. Users will also have choices, to report any update where they will provide specific information about the news and location of occurrence. Also, they will be able to analyze news of different categories.System architecture represents the application as it's a high-level logical representation of the system. It shows the components and their relationship. In figure the system architecture of our application is shown. System architecture includes:

- Presentation Layer
- Business Layer
- Database Layer

4.2 Design Methodology

The Newsverse App will be build using agile methodology. In agile, for development of Web application complex will be divided into modules given in following

4.2.1 News Data retrieval

News will be retrieved from different platforms like news websites and other authentic resources.

4.2.2 Removal of duplicate news data

After retrieving news data, it will check for duplication and remove duplicate data as mentioned above news data will be retrieved from different resources.

4.2.3 Preprocessing of data

when removal of duplicate news is done, preprocessing will be the next step to get the most out of data and make it ready for classification algorithms.

4.2.4 Classification

Classification algorithms will be applied to predict the category of news data one by one.

4.2.5 Mark news on map

After the prediction of category news will be marked on the map using legends.

4.3 Flowchart Diagram

A flow diagram depicts a series of actions, movements within a system, and/or decision points. They are a detailed explanation of each step in a process, regardless of its complexity.

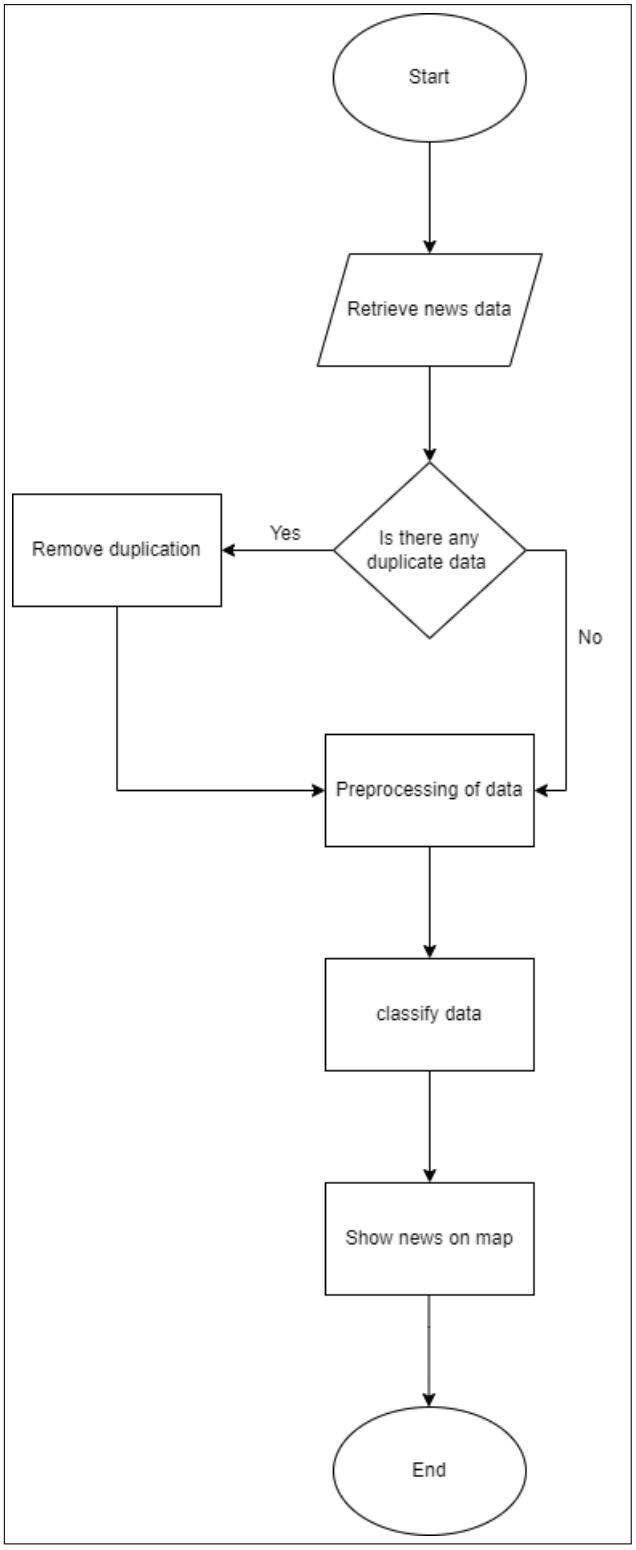

Figure 4.1: Flowchart Diagram

4.4 Entity Relationship Diagram

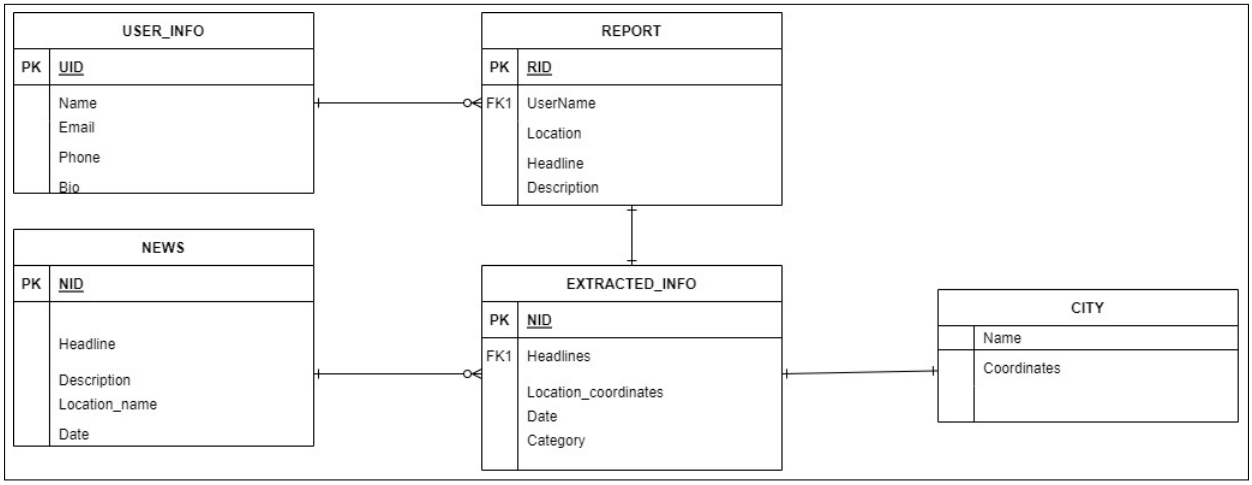

Figure 4.2: Entity Relationship Diagram

4.5 Sequence Diagram

Sequence diagram shows the interaction between the objects of the system application.

4.5.1 View map sequence diagram

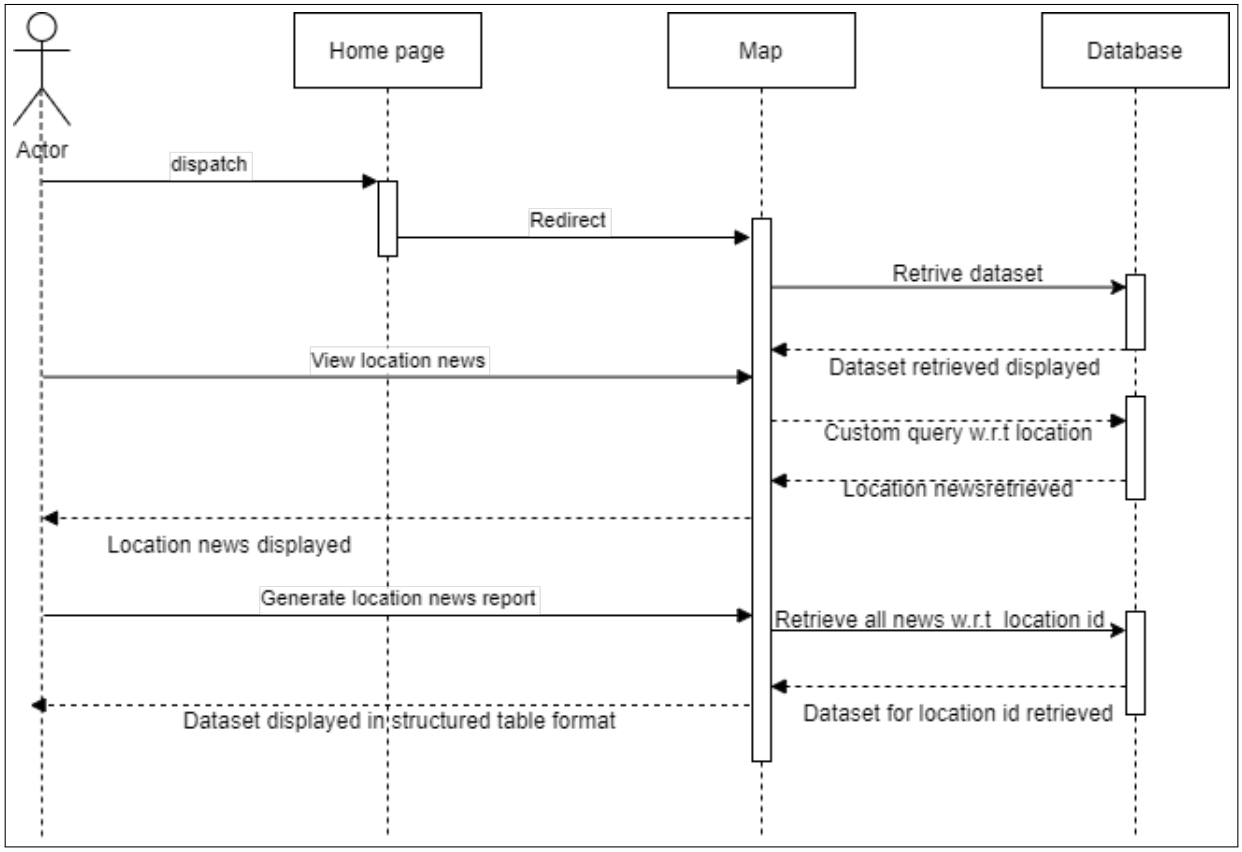

Figure 4.3: View Map sequence diagram

User able to see map with the news mark on it and newsfeed containing news in a structured form.

4.5.2 Advance search sequence Diagram

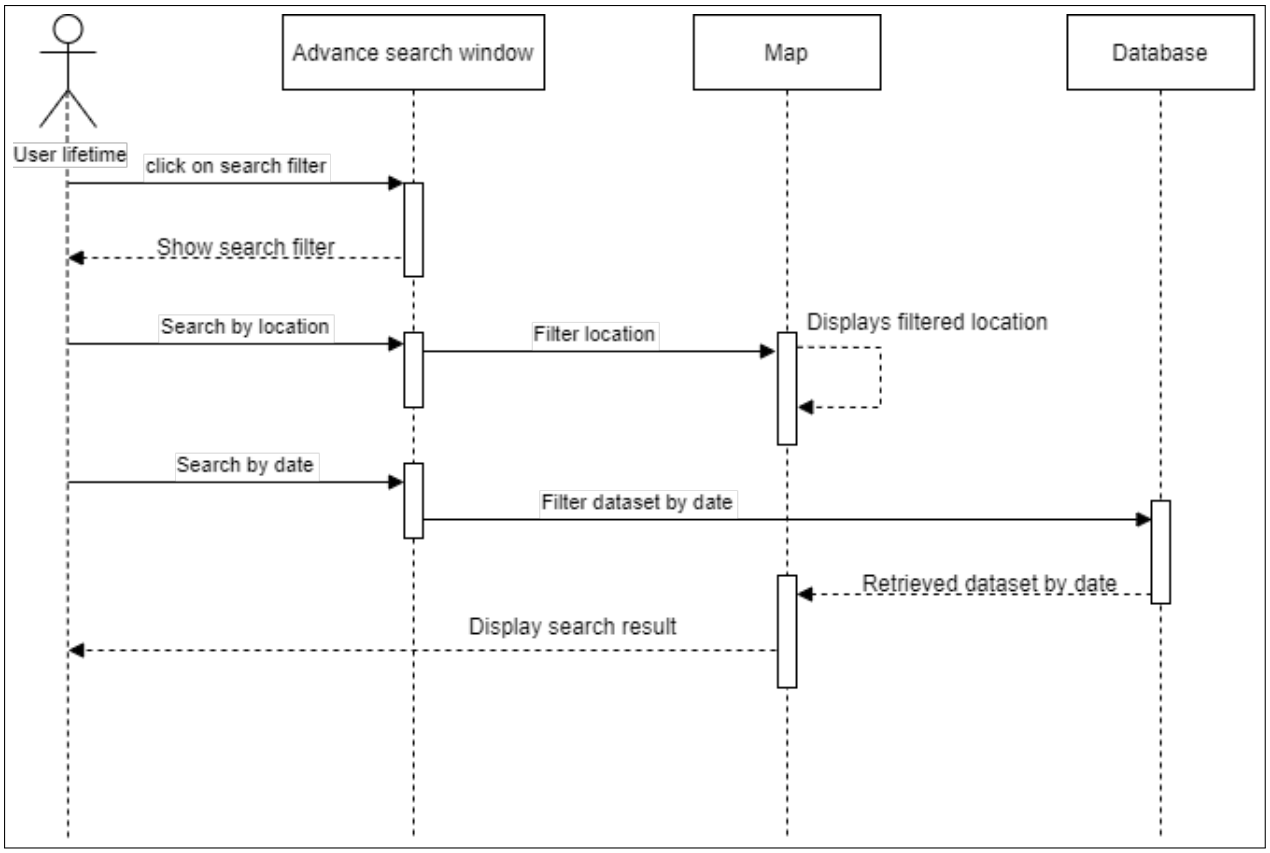

Figure 4.4: Advance search sequence diagram

User can also view news by using advance searches of specific area and also user can view past news by selecting past date.

4.5.3 View chart sequence Diagram

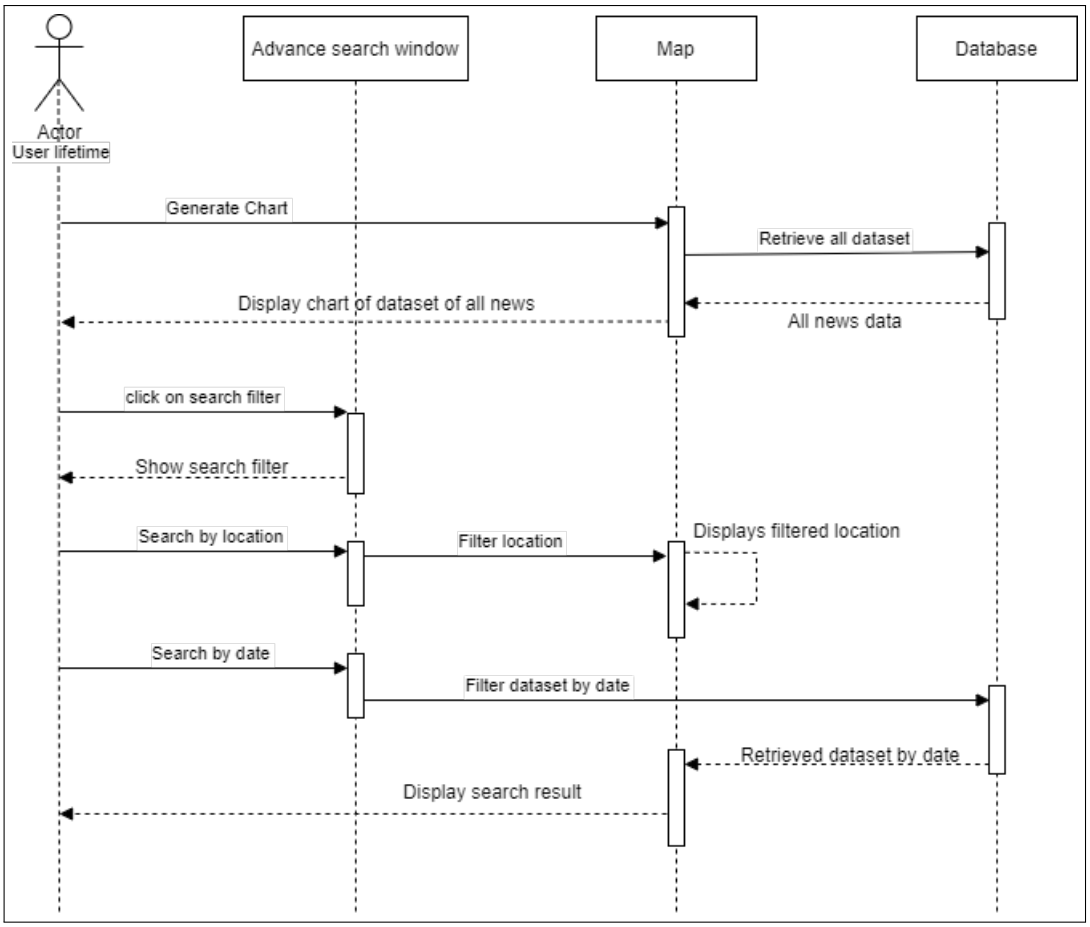

Figure 4.5: View chart sequence diagram

System will be able to show data in a graphical manner in a form of charts so it will help users to analyze situations

4.6 Activity diagrams

The activity diagram, which is used to illustrate dynamic features of the system, is another significant behavioural diagram in UML diagrams. An activity diagram is a more sophisticated version of a flow chart that displays the movement of information from one activity to the next.

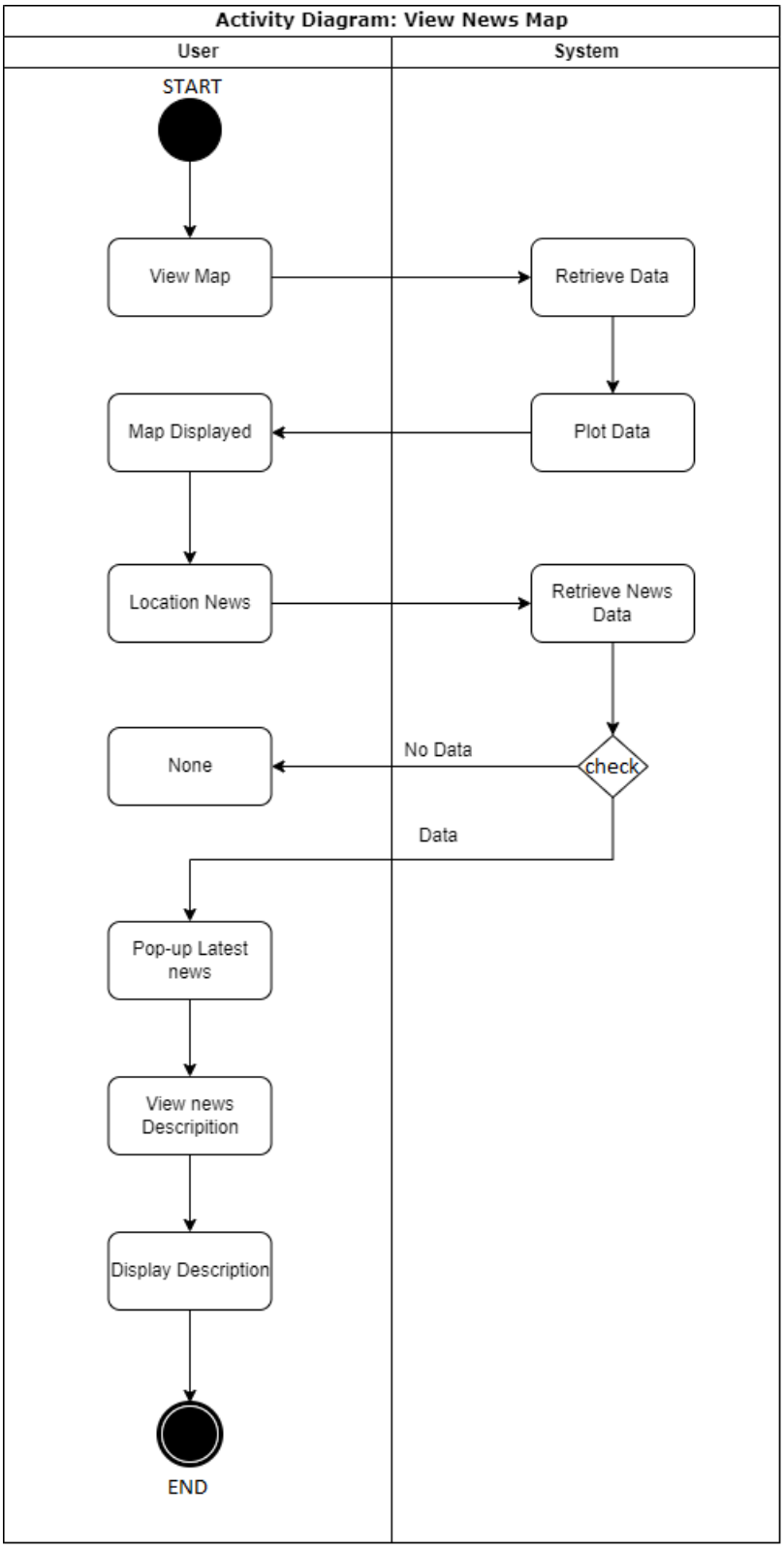

Figure 4.6: Activity diagram for viewing news map

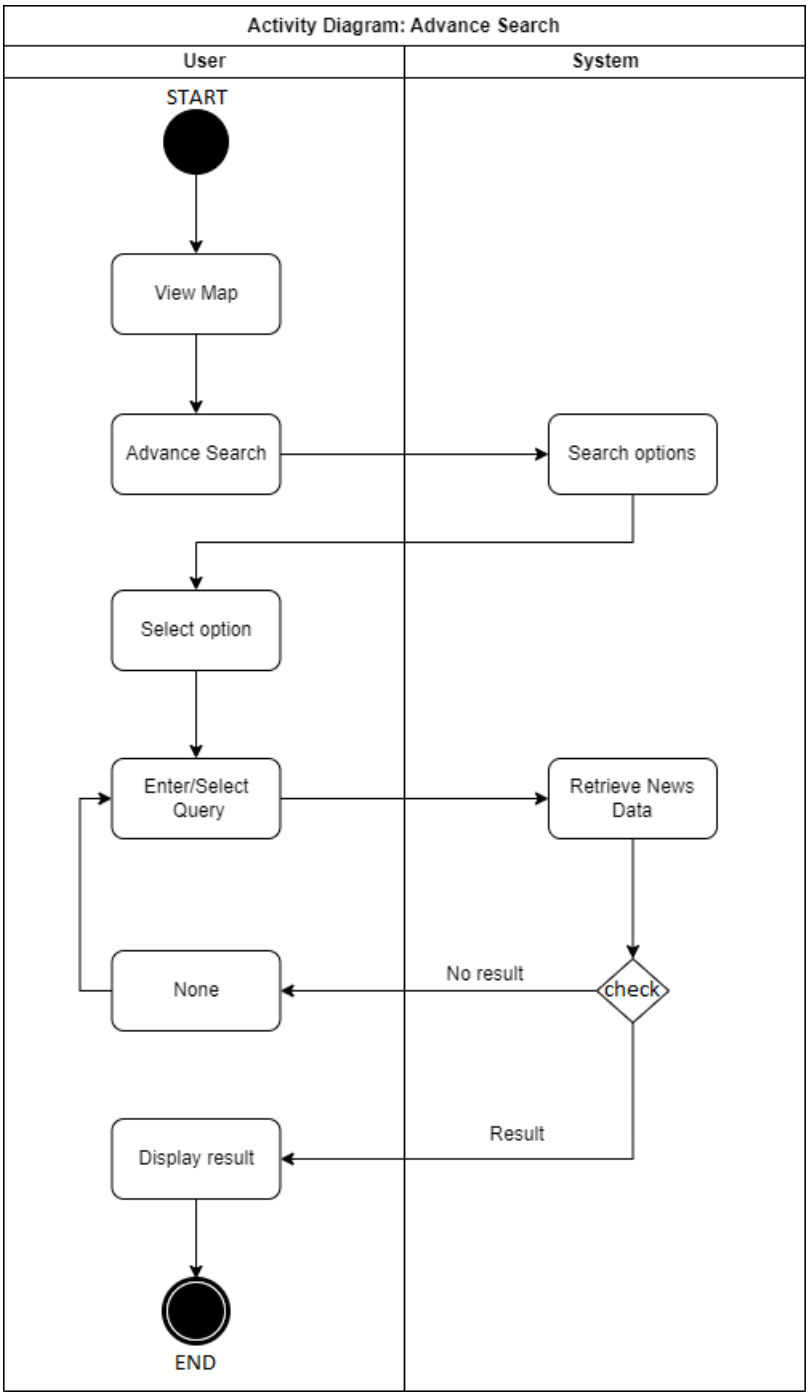

Figure 4.7: Activity diagram for advance search

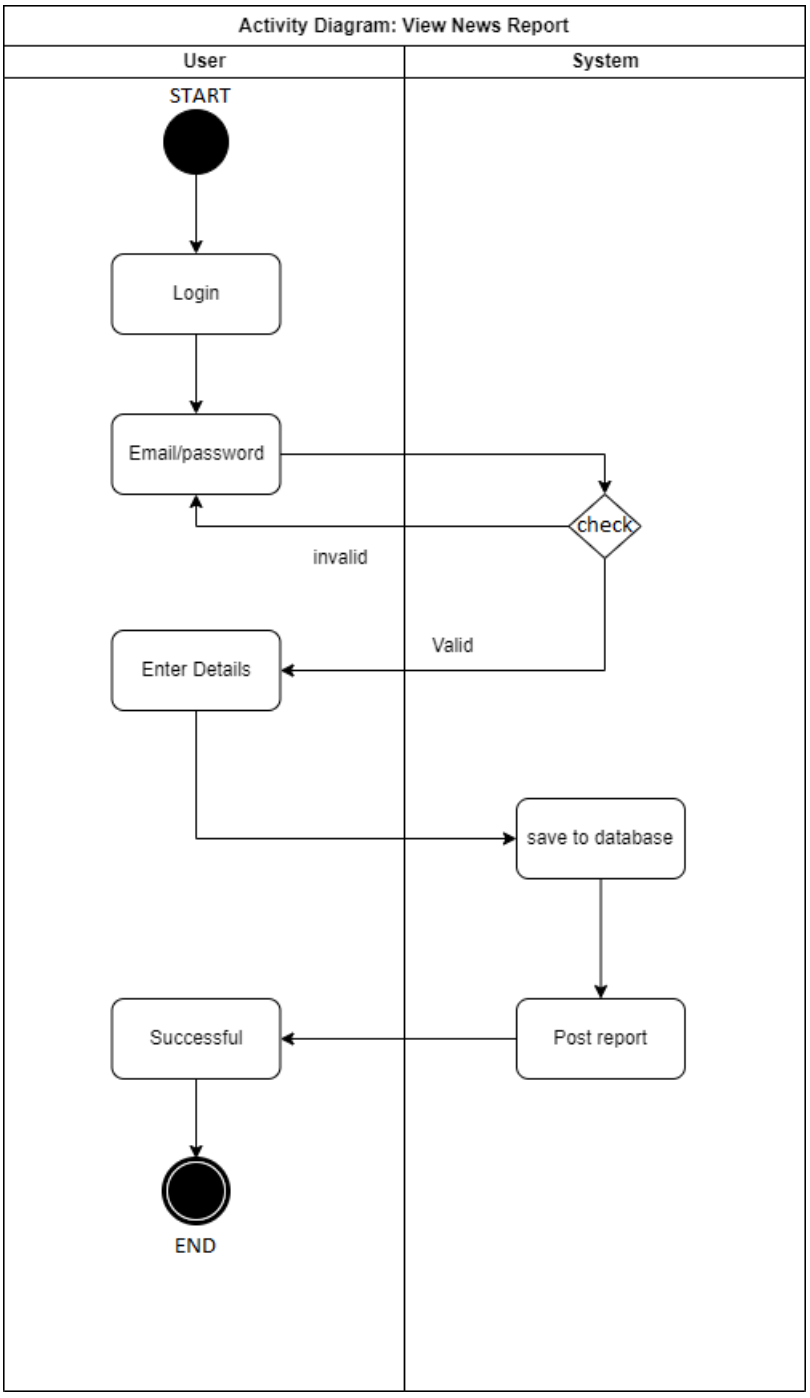

Figure 4.8: Activity diagram for view news report

Chapter 5

System Implementation

5.1 System implementation

System implementation is the process of specifying how the system should be built, as well as verifying that it is operational and fulfils quality requirements.

5.2 System Architecture

A theoretical model known as a system architecture defines the behavior, structure, and other components of a system. A formal definition and depiction of a system is called an architecture description, and it is set up to help with the analysis of its structures and behaviours. Our 3-tier system is as follows:

- Application Layer
- Business Layer
- Database Layer

5.2.1 System architecture diagram

Figure 5.1 is system architecture diagram of our application which is an abstract depiction of the system's component architecture.

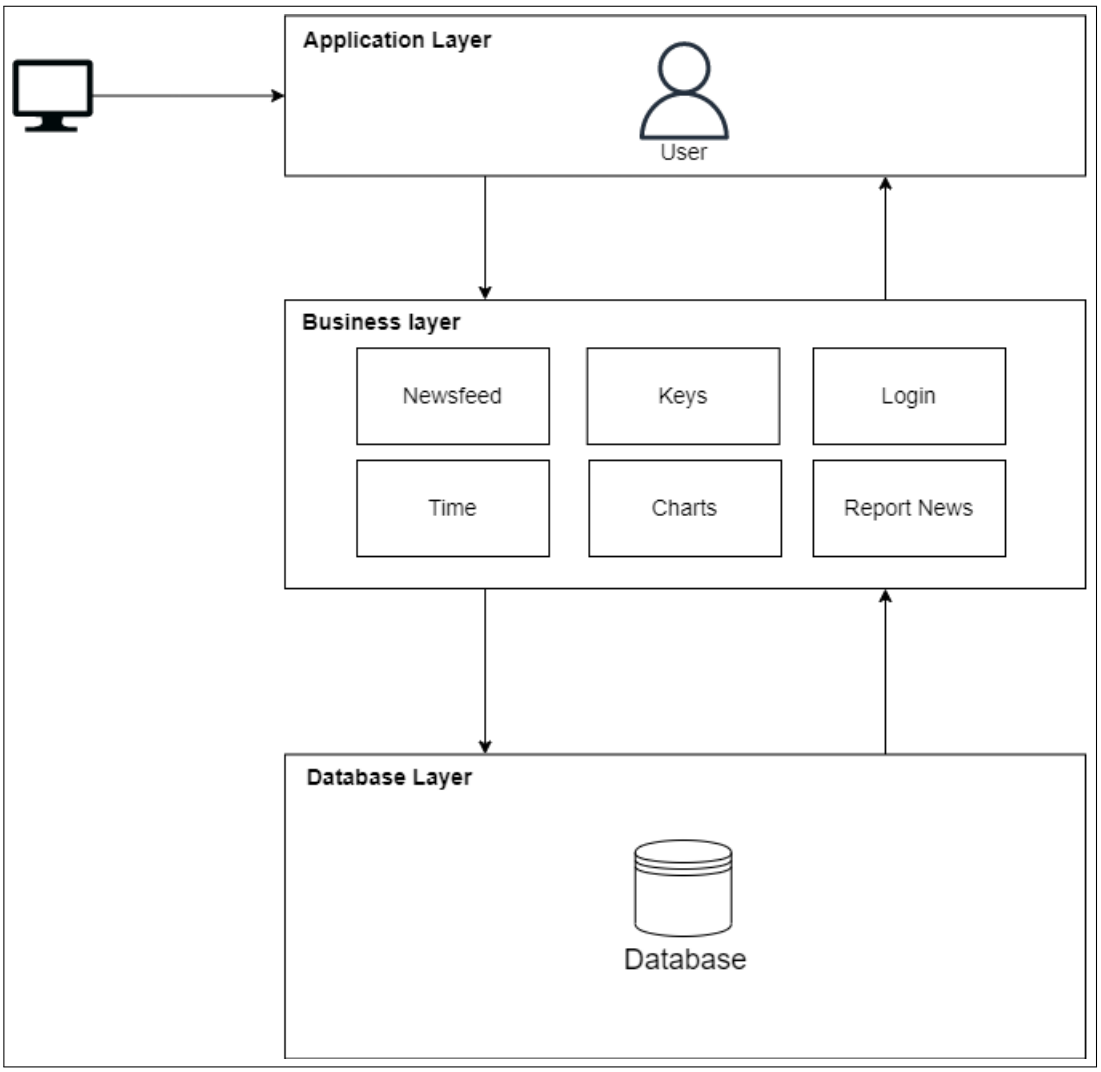

Figure 5.1: System Architecture

5.3 System Internal Component

5.3.1 Register

User can register themselves by using there email or google account as shown in figure 5.2.

Figure 5.2: User Register

5.3.2 Login

After registering their account, System will authenticate their email and password as shown in figure 5.3

Figure 5.3: User sign in

5.4 Navbar

Navbar of our system contain 5 options. Select Category, NewsFeed, Charts, Post news and Login as shown in figure 5.4

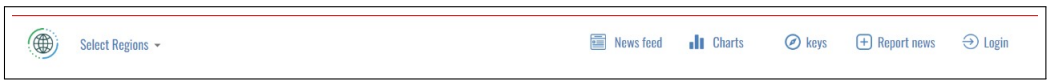

Figure 5.4: Navbar

5.5 Map

Newsverse map contain markers, markers will represent news on map according to the location of occurrence as shown in figure 5.5

Figure 5.5: Map

5.6 Select Category

After selecting the category user will be able to see news of that specific category only as shown in figure 5.6

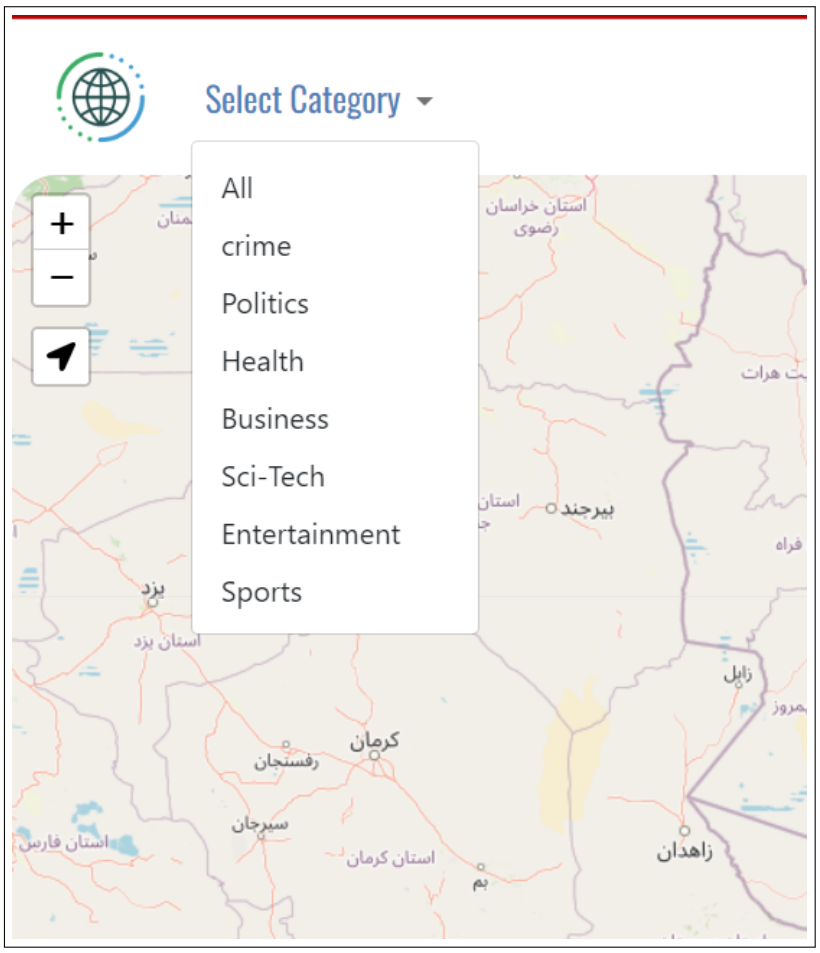

Figure 5.6: Select region

5.7 News feed

News feed will show news and its description in a structured form. And by selecting the specific news from list will show the exact location along with details on map as shown in figure 5.7

Figure 5.7: News feed

5.8 Homepage

Homepage contains map which shows the location of news and header contain 5 options as shown in figure 5.8

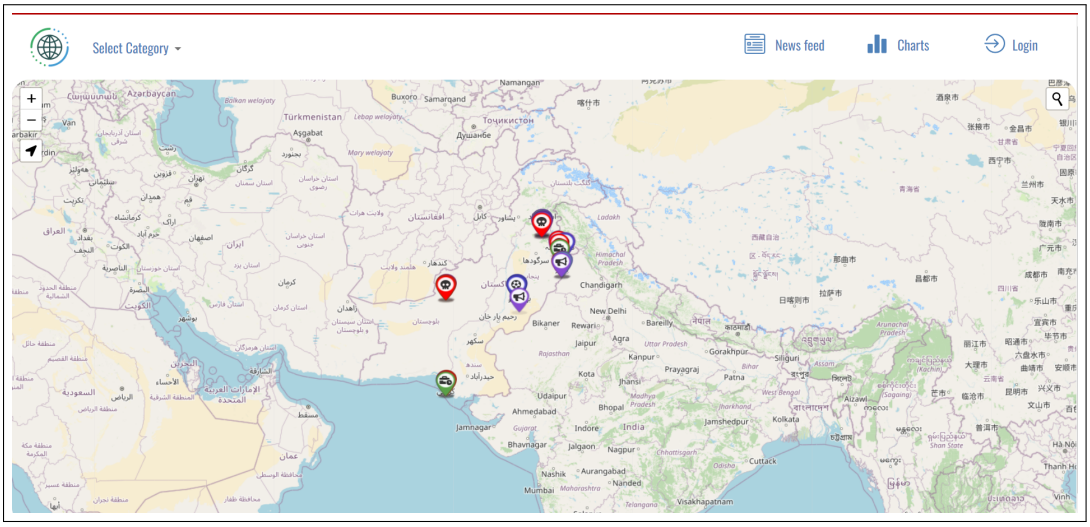

Figure 5.8: Home Page

Chapter 6

System Testing and Evaluation

6.1 Introduction

Investigating the security elements of every program is crucial. When it comes to web app security testing, it is something that must also be addressed seriously. Being prepared and aware of what to look for beforehand is the greatest strategy to avoid weaknesses. Finding errors is done by running a system under test. Testing enhances a system's integrity by discovering design faults and system issues. Testing looks for areas that are susceptible to errors. As a result, the system is able to prevent mistakes. The testing procedures that were employed to evaluate the system are described in this section. The following test fields will be covered in this chapter:

6.2 Graphical User Interface Testing:

The project's GUI was created with the goal of making it appealing to kids. The hues were striking, but not blindingly so. All keys are positioned for maximum usability, and the regular text are large enough for the user to readily read.

Table 6.1: Test case GUI

Test case ID	Test case 1
Description	Test the graphical user interface
Initial Conditions	System is set up as per requirements
steps	Task and results
1	Check the application is opening. (pass)
2	Verifying all the labels are working properly. (pass)
3	verifying all the links are working properly. (pass)
4	Verifying map and markers is working properly (pass)
5	Check all the buttons are working properly(pass)
6	Check all the text fields are accepting data. (pass)
7	Search bar and other map navigation tools working properly. (pass)

6.2.1 Usability testing:

The project's layout was kept simple but entirely operational to make it simple to use while giving the user complete control. The project is maintained as minimal as it can be to ensure that the user has no problems utilizing the web application.

Table 6.2: Usability testing

Test case ID	Test case 2
Description	Test the Usability of system
Initial Conditions	System is set up as per requirements
steps	Task and results
1	Check the application is opening. (pass)
2	check all labels, buttons, bars, icons, texts and fonts are relatable. (pass)
3	Check that user can navigate between different pages. (pass)
4	Check that design is emphasizing on main features of application (pass)
5	Check all tasks can be completed easily(pass)
6	Check that markers are representing data correctly on map. (pass)

6.2.2 Performance testing:

As our application is using scraping to collect data from different resources and its performance may be affected by the speed of internet because it may require stable internet for good performance.

Table 6.3: Performance testing

Test case ID	Test case 3
Description	Test the Performance of system
Initial Conditions	System is set up as per requirements
steps	Task and results
1	Check the application is opening. (pass)
2	Verify that loading time of an application is minimum. (pass)
3	Check the response time of an application is very small. (pass)
4	Check that application can manage multiple users at a time. (pass)
5	Verify that the application can function even with a slower connection. (pass)

6.2.3 Security testing:

The project is free from any vulnerabilities and threats that may cause loss of data. All of the information is encrypted and used and does not allow any invalid user who may cause problems in an application.

Table 6.4: Security testing

Test case ID	Test case 4
Description	Test the security of system
Initial Conditions	System is set up as per requirements
steps	Task and results
1	Check the application is opening. (pass)
2	Verify that only authenticated users allowed to report and see their own details. (pass)
3	Verify the encryption of the passwords. (pass)
4	Make sure that the application does not permit any invalid users. (pass)
5	Check the application to make sure that only authorized users are able to access information. (pass)

6.3 Test Cases:

6.3.1 Logging in application:

Table 6.5: Test case: Logging in Application

Test case ID	TC01
Description	Logging into application by using username and password
Compatibility	Runs on all supported browsers
Use case ID	UC1
Input	Username and password
Output	Logged in
Result	Pass

6.3.2 Signing up in application:

Table 6.6: Test case: Signing up in application

Test case ID	TC02
Description	Signing up into application by using username and password
Compatibility	Runs on all supported browsers
Use case ID	UC2
Input	Username and password
Output	Registered
Result	Pass

6.3.3 View news map:

Table 6.7: Test case: View news map

Test case ID	TC03
Description	Viewing news on map and search for specific news by using filters
Compatibility	Runs on all supported browsers
Use case ID	UC3
Input	Selecting news categories or searching for a location
Output	News showed on map
Result	Pass

6.3.4 Report news:

Table 6.8: Test case: Report news

Test case ID	TC04
Description	Reporting news by users
Compatibility	Runs on all supported browsers
Use case ID	UC4
Input	News headlines, description, location coordinates, user information
Output	News posted showed on map as marker with different color.
Result	Pass

6.3.5 View news trends:

Table 6.9: Test case: View news trends

Test case ID	TC05
Description	View trends and stats of news
Compatibility	Runs on all supported browsers
Use case ID	UC5
Input	Select news category or select location
Output	Graphs were shown.
Result	Pass

6.4 Our Collected Dataset:

Dataset is collected from various local news channels for training of machine learning model based on their categories so that algorithm should be able to classify news headlines the categories of data.

6.4.1 Dataset collection:

In order to collect data from different news website, we use a python library called Beautiful soup. Beautiful Soup is a Python library which help to extract data out of HTML and XML files. First it get HTML text of a website from which we want to scrap data. Then parse the HTML text to collect data. To parse a document, pass it into the BeautifulSoup constructor. It creates parse tree from website source code which help to iterate over a tree to find elements which contain use full information. We extract news titles, dates, links and category. We collect data from different news websites to train our model in which include Geo news, Dunya news, Tribune news, 92News, MM news and BOL News

Table 6.10 shows our data count according to categories of news headlines

Table 6.10: Data count

Categories	**Number of instances**
Business	2457
Sports	2447
Crime	2098
Health	2070
Entertainment	2026
Politics	2010
Sci-Tech	1918
Total instances=15026	

6.5 Visualization of data count in each category

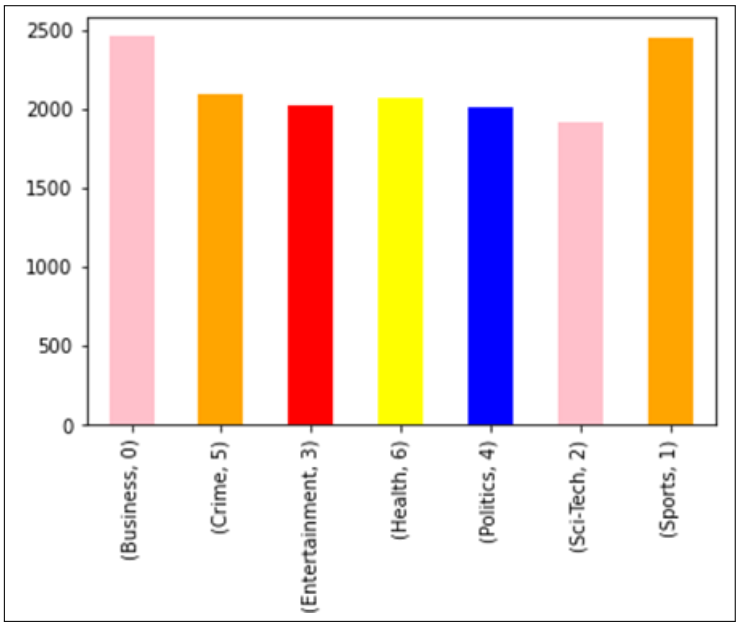

Figure 6.1: Category of data

6.6 Preprocessing of data:

Data preparation is the process of putting raw data into a format that is usable. Given that we cannot work with raw data, it is also a crucial stage in data mining. Before using machine learning or data mining methods, the data's quality should be examined.

- First of all we remove tags, special characters and stop words from our dataset and then convert everything into lower case.
- After that Lemmatization was applied, which is the process to combining a word's several inflected forms into a single unit for analysis. Similar to stemming, lemmatization adds context to the words. As a result, it ties words with related meanings together. Because lemmatization performs morphological analysis on the words, it is preferred to stemming.
- Every algorithm we use in NLP operates on numbers. Our text cannot be entered into the algorithm directly. As a result, the text is first preprocessed using the Bag of Words model, which creates a bag of terms from it and keeps track of how many times the most common words appear overall.

6.7 Evaluation matrix:

Table 6.11: Evaluation matrix

Models	**Model Performance**			
	Accuracy	**Precision**	**Recall**	**F1-Score**
Logistic Regression	89.64%	0.89640	0.8964	0.8964
Random Forest	88.75%	0.8875	0.8875	0.8875
Multinomial Naive Bayes	90.86%	0.9086	0.9086	0.9086
Gaussian Naive Bayes	72.56%	0.7256	0.7256	0.7256

We find that the Multinomial Naive Bayes Classifier model has provided the best accuracy of all machine learning models after training and testing the model.

6.8 Confusion matrix:

Table 6.12: Confusion matrix

Predicted Class	**Actual Class**						
	Business	**Crime**	**Entertainment**	**Health**	**Politics**	**Sci-Tech**	**Sports**
Business	623	13	43	4	20	9	11
Crime	11	644	6	7	5	5	13
Entertainment	47	11	479	9	9	4	19
Health	7	18	7	564	25	10	10
Politics	7	3	5	5	578	15	10
Sci-Tech	1	1	7	3	31	587	11
Sports	9	6	7	2	9	10	588

On the test set, this model's overall accuracy (true positives or accurate predictions/all) was 0.9086. The confusion matrix, however, provides extra information about accuracy per class as well as intuition for accuracy and recall effectiveness. The matrix reveals that the model was extremely accurate in classifying categories. The category of entertainment, however, had lesser accuracy. The confusion matrix's diagonal values, however, should be larger because this indicates a high number of accurate predictions.

Chapter 7

Conclusions

7.1 Conclusion

In conclusion, live news mapping system is great help for all those who are more curious to know about the situation. It allows public to share news which cannot be shown on mainstream media. Also, it is big help for the law enforcement agencies to monitor situation so that they response quickly as possible according to reported news. Moreover, Sources of news can be changed easily so that news analysts analyze news according to their own sources. This application provides many benefits if used in correct way.

7.2 Future Enhancements

Our application is only GIS news system in Pakistan, in future it can be easily extend to the other parts of world. Moreover, many features will also be added to this application so that every user get advantage of it.

Appendix A

User Manual

A.1 Sign in screen:

Figure A.1 shows the login screen of our application where users can sign in into application.

Figure A.1: User Manual: Sign in screen

A.2 Signup screen:

Figure A.2 shows the signup screen where new users can signup.

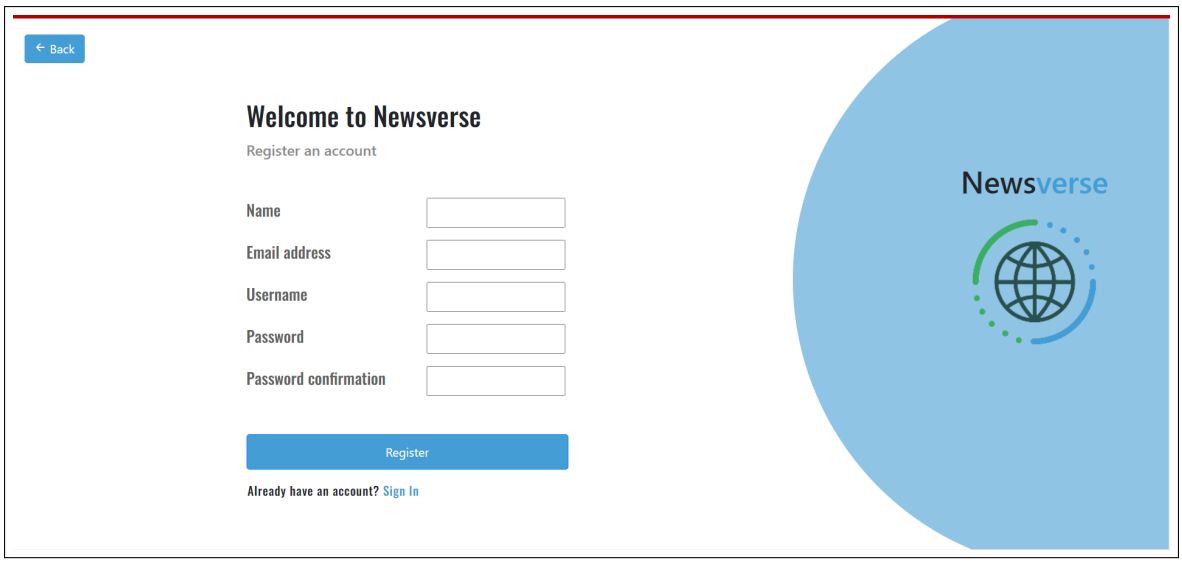

Figure A.2: User Manual: Sign up screen

A.3 Main screen:

Figure A.3 shows the main screen where users can see map and select different options in order to interact with application.

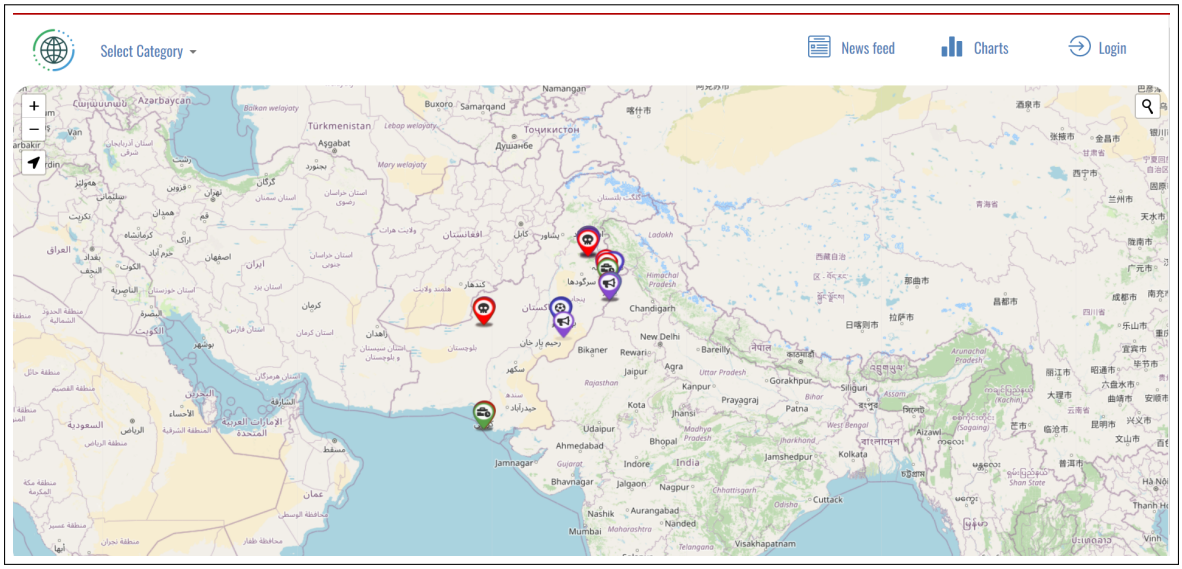

Figure A.3: User Manual: main screen

A.4 Select category option:

Figure A.4 shows the drop-down menu from where users can select specific category and the will be able to see news of that category.

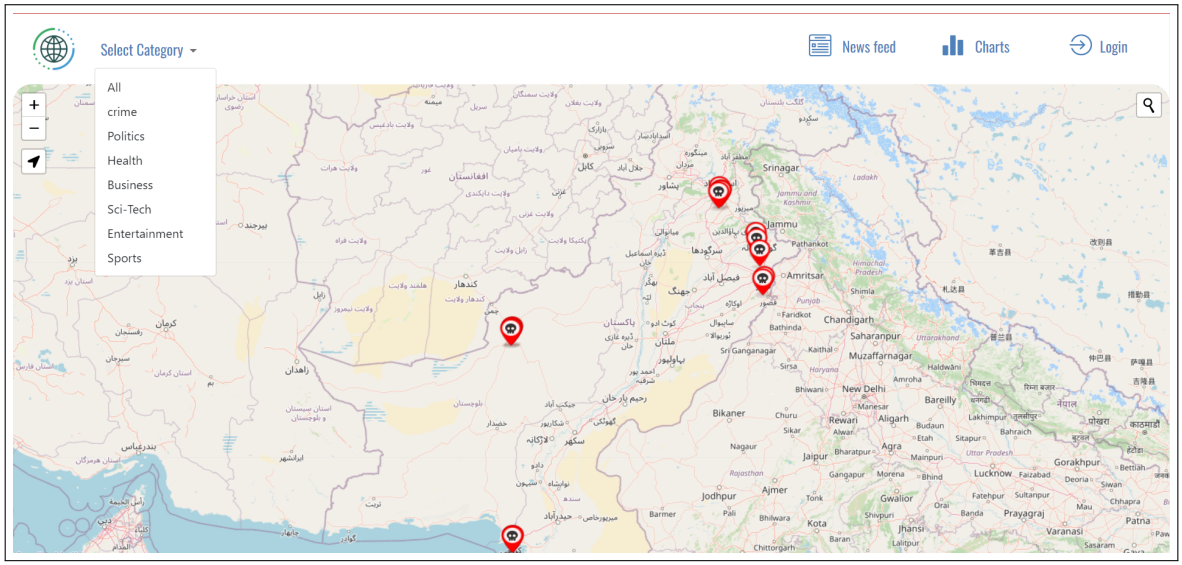

Figure A.4: User Manual: Select category option

A.5 Newsfeed:

Figure A.5 shows newsfeed, After clicking the newsfeed users will be able to see news marked on map.

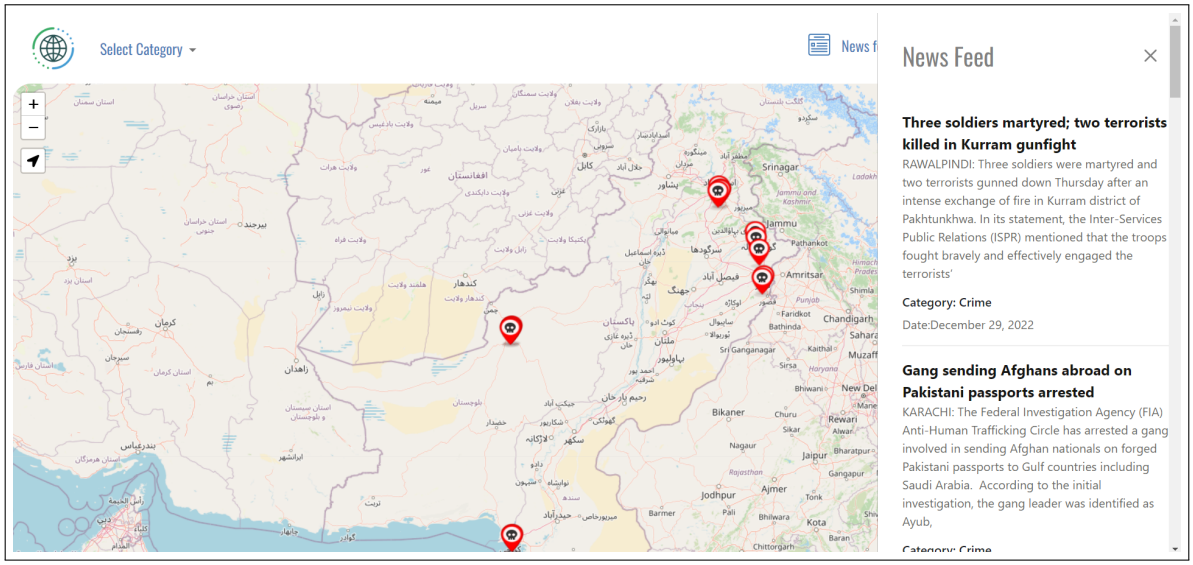

Figure A.5: User Manual: Newsfeed screen

A.6 Report news Screen:

Figure A.6 shows the page which allows users to post news after entering all the required information in fields ,if they are authenticated only.

Figure A.6: User Manual: Report news Screen

A.7 News analysis:

Figure A.7 shows analysis of news where user can see analysis of news.

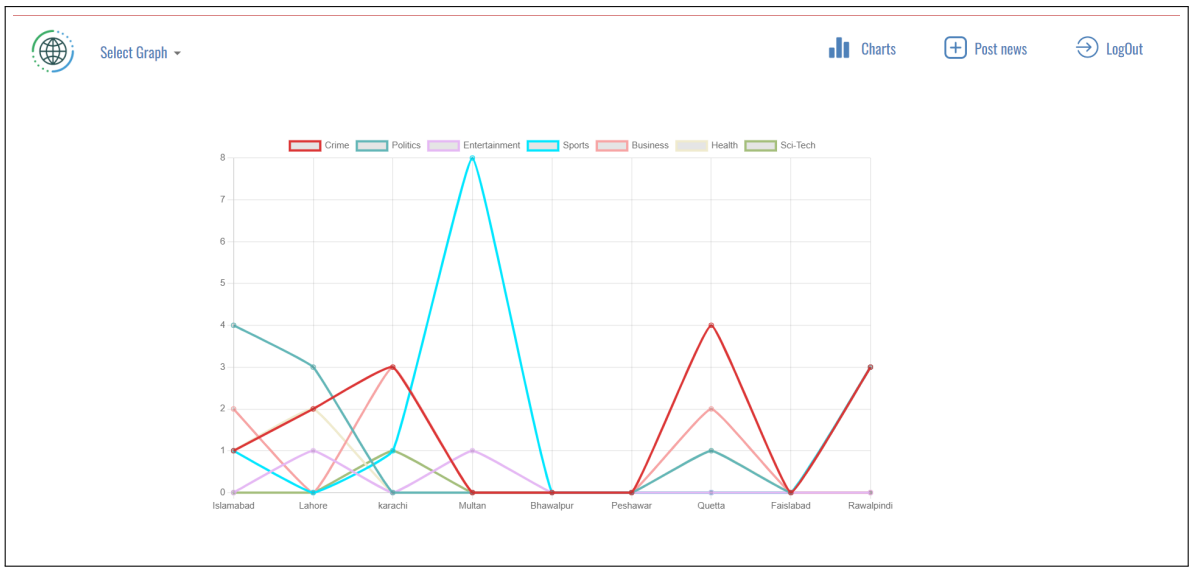

Figure A.7: User Manual: Report news Screen

References

[1] Yukiko Kawai, Yusuke Fujita, Tadahiko Kumamoto, Jianwei Jianwei, and Katsumi Tanaka. Using a sentiment map for visualizing credibility of news sites on the web. In *Proceedings of the 2nd ACM workshop on Information credibility on the web*, pages 53–58, 2008. Cited on p. 6.

[2] April Lindgren and Jon Corbett. The local news map, 2016. Cited on p. 10.

[3] Hanan Samet, Marco D Adelfio, Brendan C Fruin, Michael D Lieberman, and Jagan Sankaranarayanan. Photostand: A map query interface for a database of news photos. *Proceedings of the VLDB Endowment*, 6(12):1350–1353, 2013. Cited on p. 8.

[4] Hanan Samet, Jagan Sankaranarayanan, Michael D Lieberman, Marco D Adelfio, Brendan C Fruin, Jack M Lotkowski, Daniele Panozzo, Jon Sperling, and Benjamin E Teitler. Reading news with maps by exploiting spatial synonyms. *Communications of the ACM*, 57(10):64–77, 2014. Cited on p. 8.

[5] Lei Tang, Yi Zhu, Kunwang Tao, and Yong Feng. Implementation of website news map function based on map world. 2013. Cited on p. 9.

[6] Benjamin E Teitler, Michael D Lieberman, Daniele Panozzo, Jagan Sankaranarayanan, Hanan Samet, and Jon Sperling. Newsstand: A new view on news. In *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, pages 1–10, 2008. Cited on p. 7.