**Project Report**

***on***

**FAKE NEWS DETECTION**

***In partial fulfillment of requirements for the degree***

***of***

**BACHELOR OF TECHNOLOGY**

**IN**

**INFORMATION TECHNOLOGY**

***Submitted by:***

ADITI SHARMA [18100BTCSICS02994]

AKSHAT GAWSHINDE [18100BTCSICS02996]

ARNAV SENGAR [18100BTCSICS02998]

UJJWAL THAKRE [18100BTCSICS03043]

***Under the guidance of***

PROF. CHETAN VERMA

PROF. SUMIT NIGAM

Logo, company name

Description automatically generated

Logo, company name

Description automatically generated

Logo, company name

Description automatically generated

Logo, company name

Description automatically generated

Logo, company name

Description automatically generated

**SHRI VAISHNAV VIDYAPEETH VISHWAVIDYALAYA, INDORE**

**SHRI VAISHNAV INSTITUTE OF INFORMATIN TECHNOLOGY**

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**JULY-DEC 2021**

**SHRI VAISHNAV VIDYAPEETH VISHWAVIDYALAYA, INDORE**

**SHRI VAISHNAV INSTITUTE OF INFORMATION TECHNOLOGY**

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**DECLARATION**

We here declare that work which is being presented in the project entitled “**FAKE NEWS DETECTION**” in partial fulfillment of degree of **Bachelor of Technology in Information Technology** is an authentic record of our work carried out under the supervision and guidance of **Mr.** **CHETAN VERMA & Mr. SUMIT NIGAM** Asst. Professor of Information Technology. The matter embodied in this projecthas not been submitted for the award of any other degree.

Aditi Sharma

Akshat Gawshinde

Arnav Sengar

Ujjwal Thakre

Date:

**SHRI VAISHNAV VIDYAPEETH VISHWAVIDYALAYA, INDORE**

**SHRI VAISHNAV INSTITUTE OF INFORMATION TECHNOLOGY**

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**PROJECT APPROVAL SHEEET**

Following team has done the appropriate work related to the “**Fake News Detection**” in partial fulfillment for the award of **Bachelor of Technology in Information** **Technology** of “SHRI VAISHNAV INSTITUTE OF INFORMATION TECHNOLOGY” andis being submitted to SHRI VAISHNAV VIDYAPEETH VISHWAVIDYALAYA, INDORE.

**Team:**

1. **Aditi Sharma**
2. **Akshat Gawshinde**
3. **Arnav Sengar**
4. **Ujjwal Thakre**

**Internal Examiner** **External Examiner**

Date

**SHRI VAISHNAV VIDYAPEETH VISHWAVIDYALAYA, INDORE**

**SHRI VAISHNAV INSTITUTE OF INFORMATION TECHNOLOGY**

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**CERTIFICATE**

This is to certify that **Ms. Aditi Sharma**, **Mr. Akshat Gawshinde**, **Mr. Arnav Sengar** and **Mr. Ujjwal Thakre** working in a team have satisfactorily completed the project entitled “**FAKE NEWS DETECTION**” under the guidance of Mr. CHETAN VERMA & Mr. SUMIT NIGAM in the partial fulfillment of the degree of **Bachelor of Technology in Information Technology** awarded by SHRI VAISHNAV INSTITUTE OF INFORMATION TECHNOLOGY affiliated to SHRI VAISHNAV VIDYAPEETH VISHWAVIDYALAYA, INDORE during the academic year **July** **2021-Dec 2021**.

|  |  |
| --- | --- |
| Mr. Chetan Verma |  |
| Mr. Sumit Nigam | Prof. Manish Kumar |
| **Project Guide** | **Project Coordinator** |

Dr. Jigyasu Dubey

**Head, Department of Information Technology**

**ACKNOWLEDGEMENT**

We are grateful to a number of persons for their advice and support during the time of complete our project work. First and foremost our thanks goes to **Dr. Jigyasu Dubey** Head of the Department of Information Technology and **Mr. Chetan Verma and Mr. Sumit Nigam** the mentor of our project for providing us valuable support and necessary help whenever required and also helping us explore new technologies by the help of their technical expertise. His direction, supervision and constructive criticism were indeed the source of inspiration for us.

We would also like to express our sincere gratitude towards our Director **Dr. Anand Rajavat**  for providing us valuable support.

We are really indebted to **Prof. Manish Kumar,** project coordinator for helping us in each aspect of our academics activities. We also owe our sincere thanks to all the **faculty members** of Information Technology Department who have always been helpful.

We forward our sincere thanks to all **teaching and non-teaching staff** of Information Technology department, SVVV Indore for providing necessary information and there kind co-operation.

We would like to thanks our parents and family members, our classmates and our friends for their motivation and their valuable suggestion during the project. Last, but not the least, we thank all those people, who have helped us directly or indirectly in accomplishing this work. It has been a privilege to study at SHRI VAISHNAV VIDYAPEETH VISHWAVIDYALAYA, INDORE.

**ABSTRACT**

The advent of the World Wide Web and the rapid adoption of social media platforms (such as Facebook and Twitter) paved the way for information dissemination that has never been witnessed in human history before. With the current usage of social media platforms, consumers are creating and sharing more information than ever before, some of which are misleading with no relevance to reality. In this work, we propose to use a machine learning ensemble approach for automated classification of news articles. The extensive spread of fake news has the potential for extremely negative impacts on individuals and society. Therefore, fake news detection on social media has recently become an emerging research that is attracting tremendous attention. Fake news detection on social media presents unique characteristics and challenges that make existing detection algorithms from traditional news media inactive or not applicable. There are two ways of claiming that some news is fake or not: First, an attack on the factual points. Second, the language. The former can only be done through substantial searches into the internet with automated query systems. It could be an overwhelming task, especially for someone who is just getting started with data science and natural language processing. The latter is possible through a natural language processing pipeline followed by a machine learning pipeline. It is how we would implement our fake news detection project in Python. It is another one of the problems that are recognized as a machine learning problem.

**LIST OF FIGURES**

|  |  |
| --- | --- |
| **S.no.** | **Figure** |
| **1** | **Proposed System** |
| **2** | **Survey Question-1** |
| **3** | **Survey Question-2** |
| **4** | **Survey Question-3** |
| **5** | **Survey Question-4** |
| **6** | **Survey Question-5** |
| **7** | **Survey Question-6** |
| **8** | **Survey Question-7** |
| **9** | **Use Case Model** |
| **10** | **Conceptual level activity diagram** |
| **11** | **DFD Level 0** |
| **12** | **DFD Level 1** |
| **13** | **Sequence Diagram** |
| **14** | **State Diagram** |
| **15** | **Activity Diagram** |
| **16** | **Component Diagram** |
| **17** | **Deployment Diagram** |
| **18** | **ML testing 1** |
| **19** | **ML testing 2** |
| **20** | **Front end 1** |
| **21** | **Front end 2** |
| **22** | **Front end 3** |
| **23** | **Front end 4** |
| **24** | **Front end 5** |
| **25** | **Front end 6** |

**TABLE OF CONTENT**

Declaration

Project Approval Sheet

Certificate

Acknowledgment

Abstract

List of Figures

CHAPTER 1 – **INTRODUCTION**

1. Introduction
2. Problem Statement
3. Need for the proper System
4. Objective
5. Modules of the system
6. Scope

CHAPTER 2 - **LITERATURE SURVEY**

1. Existing System
2. Proposed System
3. Feasibility Study
   1. Technical Feasibility
   2. Economical Feasibility
   3. Operational Feasibility

CHAPTER 3 – **REQUIREMENTS ANALYSIS**

1. Method used for Requirement analysis
2. Data Requirements
3. Functional Requirements
4. Non Functional Requirements
5. System Specification
   1. Hardware specification
   2. Software Specification

CHAPTER 4 – **DESIGN**

1. Software Requirements Specification
   1. Glossary
   2. Supplementary Specifications
   3. Use Case Model
2. Conceptual level class diagram
3. Conceptual level activity diagram
4. Data flow Diagram( Level 0,1)
5. Database Design (ER-Diagram)

|  |  |  |
| --- | --- | --- |
| CHAPTER 5 – **SYSTEM MODELING** | |  |
| 5.1 | Detailed Class Diagram |  |
| 5.2 | Interaction Diagram |  |
|  | 5.2.1 Sequence Diagram |  |
|  | 5.2.2 Collaboration Diagram |  |
| 5.3 | State Diagram |  |
| 5.4 | Activity Diagram |  |
| 5.5 | Object Diagram |  |
| 5.6 | Component Diagram |  |
| 5.7 | Deployment Diagram  Test Plans and Implementation Images |  |
| CHAPTER 6 – **CONCLUSION & FUTURE WORK** | |  |
| 6.1 | Limitation of Project |  |
| 6.2 | Future Enhancement |  |
| CHAPTER 7 - **BIBLIOGRAPHY & REFERENCES** | |  |

1. Reference Books
2. Other Documentations & Resources

**CHAPTER 1 – INTRODUCTION**

**1.1 Introduction:**

The term “Fake News” was a lot less unheard of and not prevalent a couple of decades ago but in this digital era of social media, it has surfaced as a huge monster. Fake news, information bubbles, news manipulation and the lack of trust in the media are growing problems within our society. However, in order to start addressing this problem, an in-depth 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. “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.

These days’ fake news is creating different issues from sarcastic articles to fabricated news and planned government propaganda in some outlets. Fake news and lack of trust in the media are growing problems with huge ramifications in our society. Obviously, a purposely misleading story is “fake news “but lately blathering social media’s discourse is changing its definition. Some of them now use the term to dismiss the facts counter to their preferred viewpoints.

The importance of disinformation within American political discourse was the subject of weighty attention, particularly following the American presidential election. The term 'fake news' became common parlance for the issue, particularly to describe factually incorrect and misleading articles published mostly for the purpose of making money through page views. In this paper, it is seeked to produce a model that can accurately predict the likelihood that a given article is fake news.

Detecting fake news on social media poses several new and challenging research problems. The content of fake news is rather diverse in terms of topics, styles and media platforms, and fake news attempts to distort truth with diverse linguistic styles while simultaneously mocking true news. Thus, existing hand-crafted and data-specific textual features are generally not sufficient for fake news detection. Fake news is usually related to newly emerging, time-critical events, which may not have been properly verified by existing knowledge bases due to the lack of corroborating evidence or claims.

**1.2 Problem Statement:**

Social media for news consumption is a double-edged sword. On the one Hand, its low cost, easy access and rapid circulation of information lead people to consume news from social media. On the other Hand, it enables the wide spread of ‘fake news’ i.e., low quality news with intentionally false information. Therefore, Fake News detection on social media has recently become an appearing research that is attracting tremendous attention. Fake news detection on social media presents unique characteristics and challenges that make existing detection algorithms from traditional news media ineffective or not applicable. Fake news is intentionally written to mislead readers to believe false information.

**1.3 Need for the proper System:**

The low cost of creating social media accounts also encourages malicious user accounts, such as social bots, cyborg users, and trolls. Social bots can become malicious entities designed specifically with the purpose to do harm, such as manipulating and spreading fake news on social media. Therefore, fake news detection on social media has recently become an emerging research that is attracting tremendous attention. Fake news detection on social media presents unique characteristics and challenges that make existing detection algorithms from traditional news media ineffective or not applicable. First, fake news is intentionally written to mislead readers to believe false information therefore, we need to include auxiliary information, such as user social engagements on social media, to help make a determination. Second, exploiting this auxiliary information is challenging in and of itself as users’ social engagements with fake news produce data that is big, incomplete, unstructured, and noisy.

**1.4 Objective**:

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 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 to acknowledge this problem and take on this challenge to control the rates of crime, political unrest, grief, and thwart the attempts of spreading fake news.

The main objective is to detect the fake news, which is a classic text classification problem with a straightforward proposition. It is needed to build a model that can differentiate between “Real” news and “Fake” news

**1.5** **Modules of the System:**

This section consists various modules present in our project:

1.Machine learning module:

Main module of our project which will help us to detect between a real news and a fake news. This module consists of a machine learning algorithm which is logistic regression. User gives us a news article and with the help of this module we detect whether its a fake news or a real news.

2.Live News Module:

In this module the feature used is- Embedding with YouTube channel. YouTube has provided a feature for the developers to embed their YouTube videos directly with the link.

3.Interface Module:

This module will always take input from the user and transfer it to machine learning module to give the user output whether if the input is a fake news or a real news. This module also contains certain links of trusted news resources where a user can go and find out what is the trending news.

4.About Us Module:

It consists of information about US-the creators of this project.

5. Result Module:

This is the result module which will show user the prediction and redirect user to trusteed sources where one can find true news.

**1.6 Scope:**

Social media for news consumption is a double-edged sword. On the one hand, its low cost, easy access, and rapid dissemination of information allow users to consume and share the news. On the other hand, it can make viral “fake news”, i.e., low-quality news with intentionally false information. The quick spread of fake news has the potential for calamitous impacts on individuals and society. For [example](https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook?utm&utm_term=.xcGkVBNoxk#.wwwqW6rpmq), the most popular fake news was more widely spread on Facebook than the most popular authentic mainstream news during the U.S. 2016 president election. Therefore, fake news detection on social media has attracted increasing attention from researchers to politicians.

Fake news detection on social media has unique characteristics and presents new challenges. First, fake news is intentionally written to mislead readers to believe false information, which makes it difficult to detect based on news content. Thus, we need to include auxiliary information, such as user social engagements on social media, to help differentiate it from the true news. Second, exploiting this auxiliary information is nontrivial in and of itself as users’ social engagements with fake news produce data that is big, incomplete, unstructured, and noisy.

**CHAPTER 2 - LITERATURE SURVEY**

**2.1 Existing System:**

There exists a large body of research on the topic of machine learning methods for deception detection, most of it has been focusing on classifying online reviews and publicly available social media posts. Particularly since late 2016 during the American Presidential election, the question of determining 'fake news' has also been the subject of particular attention within the literature.

Fake news is not a new concept. Before the era of digital technology, it was spread through mainly yellow journalism with focus on sensational news such as crime, gossip, disasters and satirical news. The prevalence of fake news relates to the availability of mass media digital tools. Since anyone can publish articles via digital media platforms, online news articles include well researched pieces but also opinion-based arguments or simply false information. There is no custodian of credibility standards for information on these platforms making the spread of fake news possible. To make things worse, it is by no means straightforward telling the difference between real news and semi-true or false news. The nature of social media makes it easy to spread fake news, as a user potentially sends fake news articles to friends, who then send it again to their friends and so on. Comments on fake news sometimes fuel its ‘credibility’ which can lead to rapid sharing resulting in further fake news.

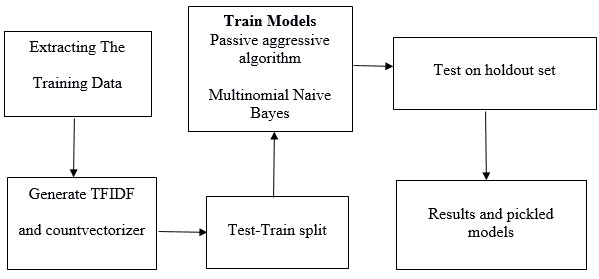
Social bots are also responsible for the spreading of fake news. Bots are sometimes used to target super-users by adding replies and mentions to posts. Humans are manipulated through these actions to share the fake news articles.

Clickbait is another tool encouraging the spread of fake news. Clickbait is an advertising tool used to get the attention of users. Sensational headlines or news are often used as clickbait that navigate the user to advertisements. More clicks on the advert means more money.

Fortunately, tools have been developed for detecting fake news. For example, a tool has been developed to identify fake news that spreads through social media through examining lexical choices that appear in headlines and other intense language structures. Another tool, developed to identify fake news on Twitter, has a component called the Twitter Crawler which collects and stores tweets in a database. When a Twitter user wants to check the accuracy of the news found they can copy a link into this application after which the link will be processed for fake news detection. This process is built on an algorithm called the NER (Named Entity Recognition).

**2.2 Proposed System:**

In this project, a model is build based on the count vectorizer or a tfidf matrix (i.e.) word tallies relatives to how often they are used in other articles in your dataset) can help. Since this problem is a kind of text classification, implementing a Naive Bayes classifier will be best as this is standard for text-based processing. The actual goal is in developing a model which was the text transformation (count vectorizer vs tfidf vectorizer) and choosing which type of text to use (headlines vs full text). Now the next step is to extract the most optimal features for count vectorizer or tfidf-vectorizer, this is done by using a n-number of the most used words, and/or phrases, lower casing or not, mainly removing the stop words which are common words such as “the”, “when”, and “there” and only using those words that appear at least a given number of times in a given text dataset.



*Fig*-1

**2.3 Feasibility Study:**

A feasibility analysis is used to determine the viability of an idea, such as ensuring a project is legally and technically feasible as well as economically justifiable. It tells us whether a project is worth the investment- in some cases, a project may not be doable. There can be many reasons for this, including requiring too many resources, which not only prevents those resources from performing other tasks but also may cost more than an organization would earn back by taking on a project that is not profitable.

**2.3.1 Technical Feasibility:**

The project ‘Fake News detection’ is mainly Machine Learning application project. The various tools and technology used in out project are listed below. The justification for the use of that technology is also discussed.

**Machine learning implementation:**

1-DATA COLLECTION AND ANALYSIS: We can get online news from different sources like social media websites, search engine, homepage of news agency websites or the fact-checking websites. On the Internet, there are a few publicly available datasets for Fake news classification like Buzzfeed News, BS Detector etc. These datasets have been widely used in different research papers for determining the veracity of news.

REAL.CSV: we used the real news dataset for passive aggressive classifier. It contains 3 columns viz 1- Title, 2- Text, 3-Subject.

FAKE.CSV: we used the real news dataset for passive aggressive classifier. It contains 3 columns viz 1- Title, 2- Text, 3-Subject.

2-CLEANING- Cleaning up the text data is necessary to highlight attributes that were going to want our machine learning system to pick up on. Cleaning (or pre- processing) the data typically consists of several steps:

1. Remove punctuation: Punctuation can provide grammatical context to a sentence which supports our understanding. But for our vectorizer which counts the number of words and not the context, it does not add value, so we remove all special characters. eg: How are you? ->How are you
2. Tokenization: Tokenizing separates text into units such as sentences or words. It gives structure to previously unstructured text. eg: Plata o Plomo-> Plata, o,Plomo.
3. Remove stopwords: Stopwords are common words that will likely appear in any text. They do not tell us much about our data, so we remove them. eg: silver or lead is fine for me-> silver, lead, fine.
4. Stemming: Stemming helps reduce a word to its stem form. It often makes sense to treat related words in the same way. It removes suffices, like ing, ly, s, etc. by a simple rule-based approach. It reduces the corpus of words but often the actual words get neglected. eg: Entitling, Entitled -> Entitle. Note: Some search engines treat words with the same stem as synonyms

3-VECTORIZING DATA:

TF-IDF -It computes relative frequency that a word appears in a document compared to its frequency across all documents TF-IDF weight represents the relative. Used for search engine scoring, text summarization, document clustering.

IDF stands for Inverse Document Frequency: A word is not of much use if it is present in all the documents. Certain terms like a, an, the, on, of etc. appear many times in a document but are of little importance. IDF weighs down the importance of these terms and increase the importance of rare ones. The more the value of IDF, the more unique is the word.

TF-IDF is applied on the body text, so the relative count of each word in the sentences is stored in the document matrix.

4- LOGISTIC REGRESSION:

It is a classification not a regression algorithm. It is used to estimate discrete values (Binary values like 0/1, yes/no, true/false) based on given set of independent variable(s). In simple words, it predicts the probability of occurrence of an event by fitting data to a logit function. Hence, it is also known as logit regression. Since, it predicts the probability, its output values lies between 0 and 1 (as expected).

Mathematically, the log odds of the outcome are modelled as a linear combination of the predictor variables.

Odds = p/(1-p) = probability of event occurrence / probability of not event occurrence

ln(odds) = ln(p/(1-p))

logit(p)=ln(p/(1-p))= b0+b1X1+b2X2+b3X3….+bkXk

5-PYTHON LIBRARIES AND PACKAGES:

Some libraries used in python:

• Numpy : Used for mathematical functions

• Pandas: Used for Web Scraping

• Itertools: they form an “iterator algebra” making it possible to construct specialized tools succinctly and efficiently

• Matplotlib: plotting library

• Scikit: it Is a free software machine learning library.

Each of the technologies are freely available and the technical skills required are manageable. Time limitations of product development and the ease of implementing using these technologies are synchronized.

**2.3.2. Economic Feasibility:**

Being a machine learning based web application Fake News Detection does not involve any associated cost. Since the system doesn’t consist of any multimedia data transfer i.e. bandwidth requirement is low and does not follow any software standards, therefore there is no cost-charging of any kind.

From here, Fake News Detection is economically feasible.

**2.3.3 Operational Feasibility:**

The fundamental thought to make a model to foresee the trust worthiness of continuous news affairs going on. The proposed is comprises of following steps:

•Collection Of Data

•Pre-Processing Of Data

•Classification

•Result Analysis

The key expressions of news affairs have been taken in a form that needs to be verified.

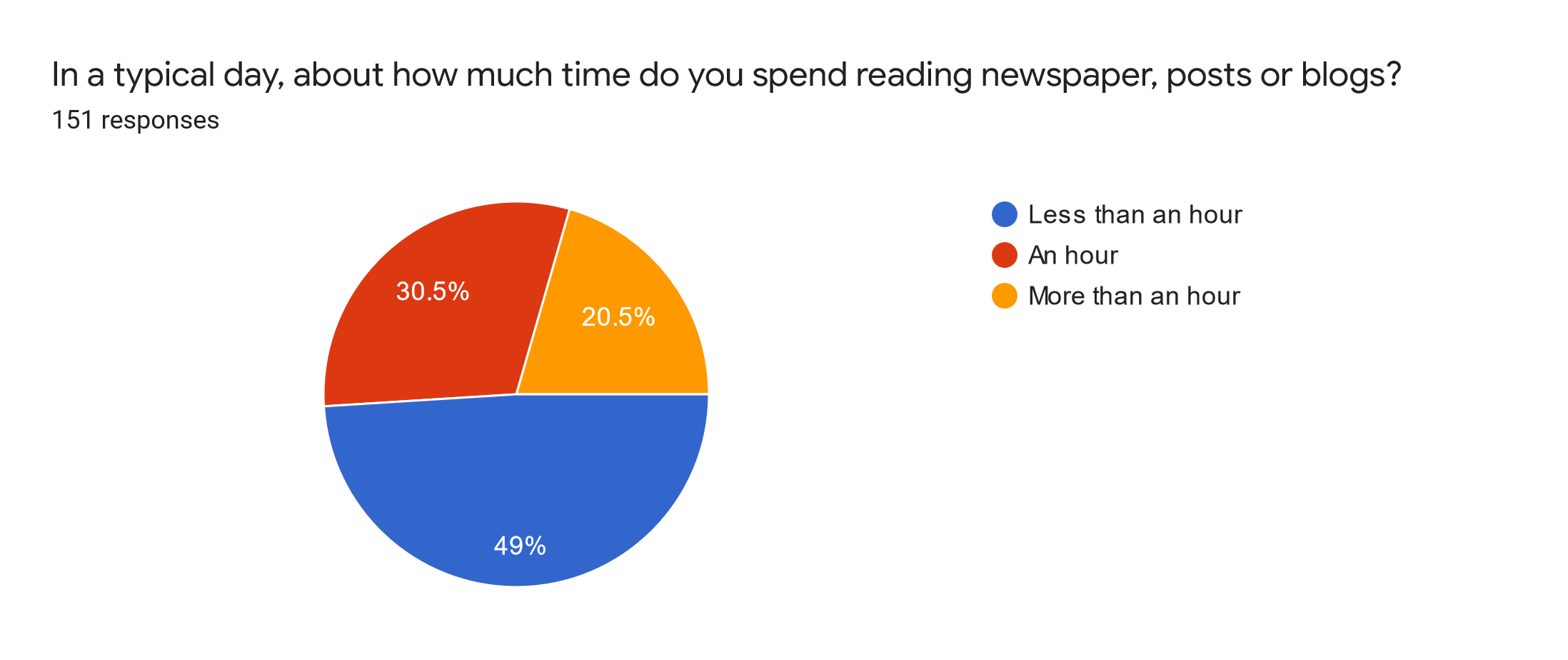
**CHAPTER 3 – REQUIREMENTS ANALYSIS**

**3.1 Method used for Requirement analysis:**

We performed requirement elicitation using the Survey method. For Survey, a set of questions is given to targeted audience to quantify their thoughts. After collecting the responses from them, data is analyzed to identify the area of interest of stakeholders. We performed **Close Ended Survey method** which includes a predefined set of answers for all the questions and the respondent has to choose from those answers. Questions can be multiple choice or can be ranked from not important to very important. The main aim for conducting survey is that by querying about their expectation and requirements from the upcoming system, we got the idea of scope and requirement of our system. Following the ideal approach to this technique that is by making a basic Google Form and offering it to the correct individuals, we found the following result:

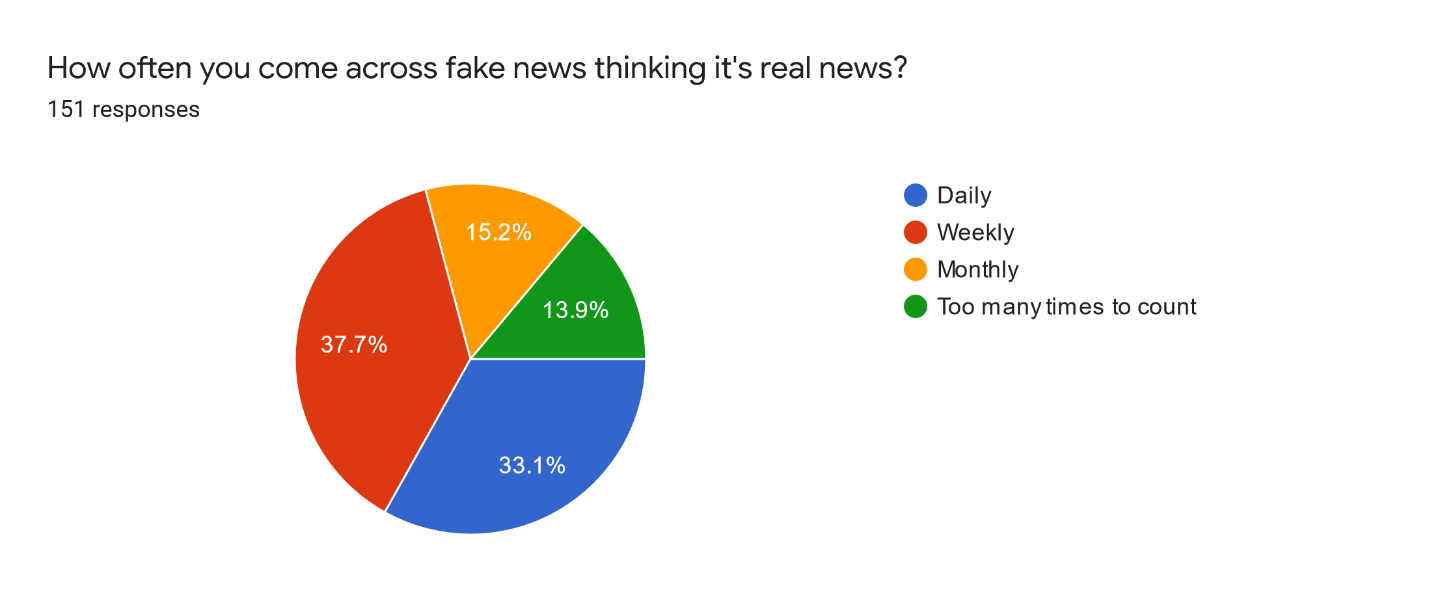
Our google form includes following questions and their respective responses:

1- In a typical day, about how much time do you spend reading newspaper, posts or blogs?



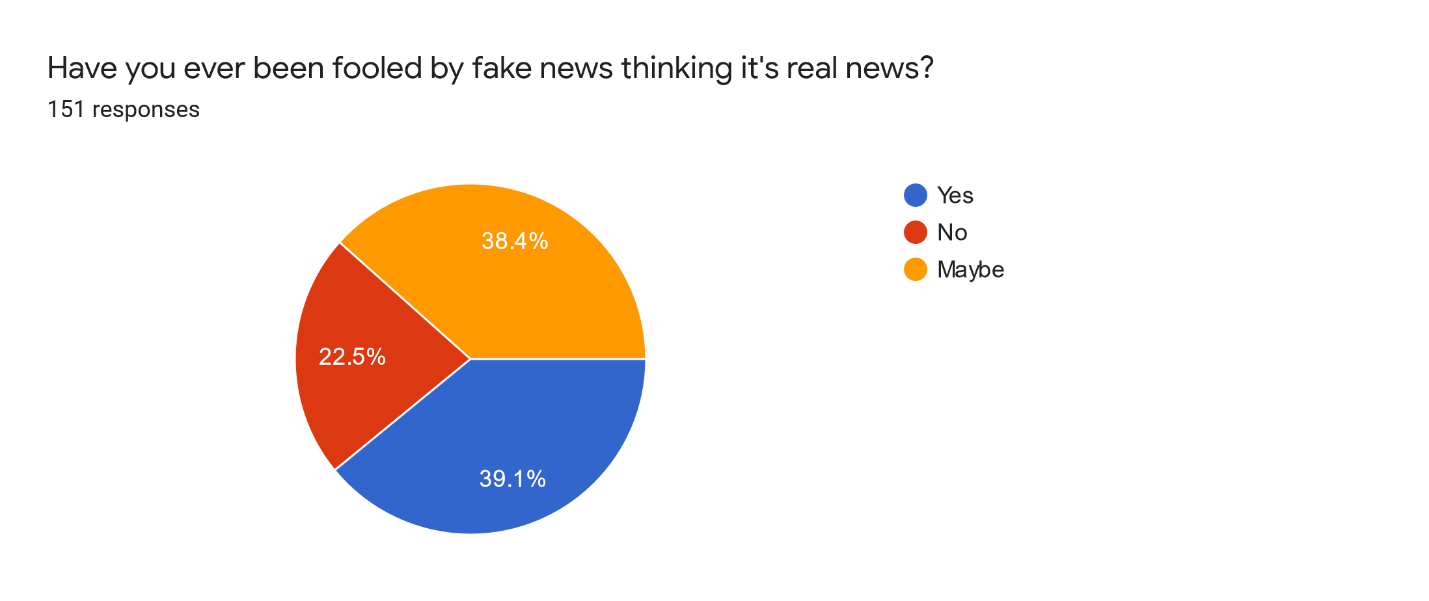
*Fig*-2

2- How often you come across fake news thinking it's real news?



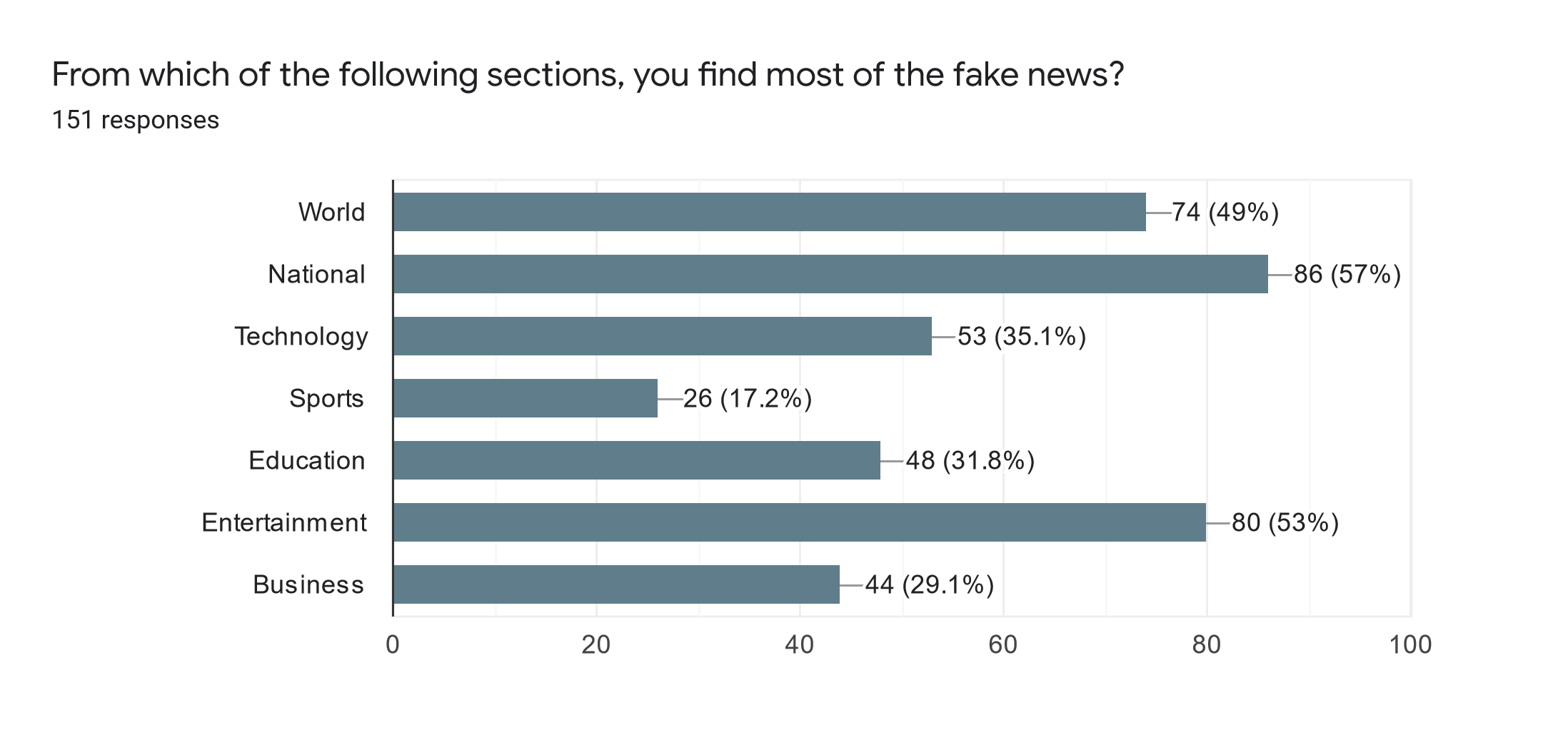
*Fig*-3

3- Have you ever been fooled by fake news thinking it's real news?



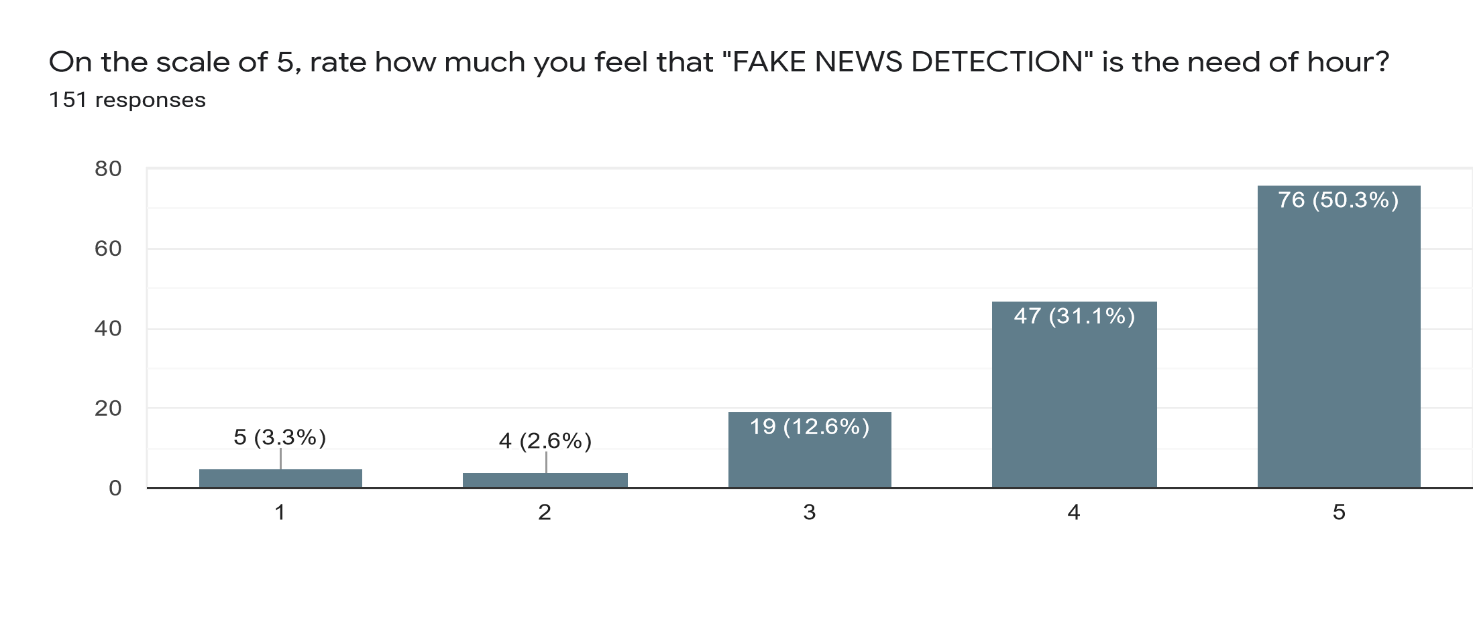
*Fig*-4

4- From which of the following sections, you find most of the fake news?



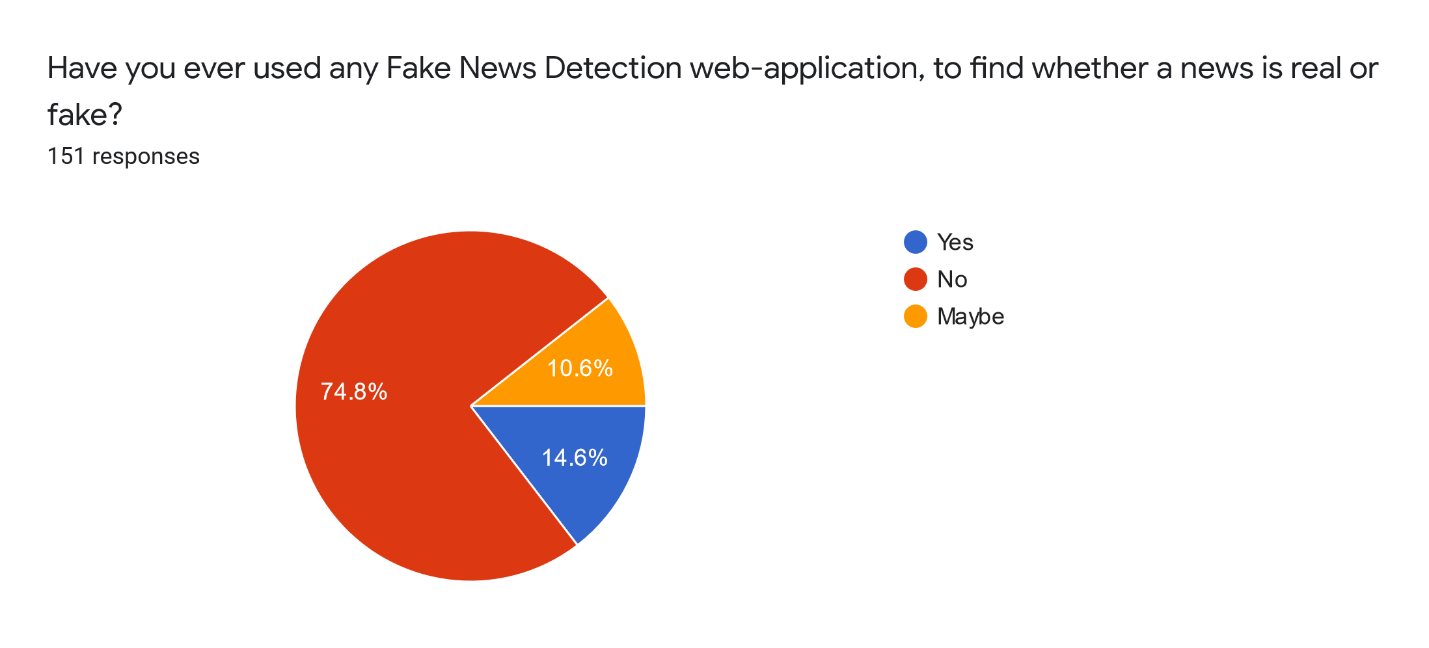
*Fig*-5

5- On the scale of 5, rate how much you feel that "FAKE NEWS DETECTION" is the need of hour?



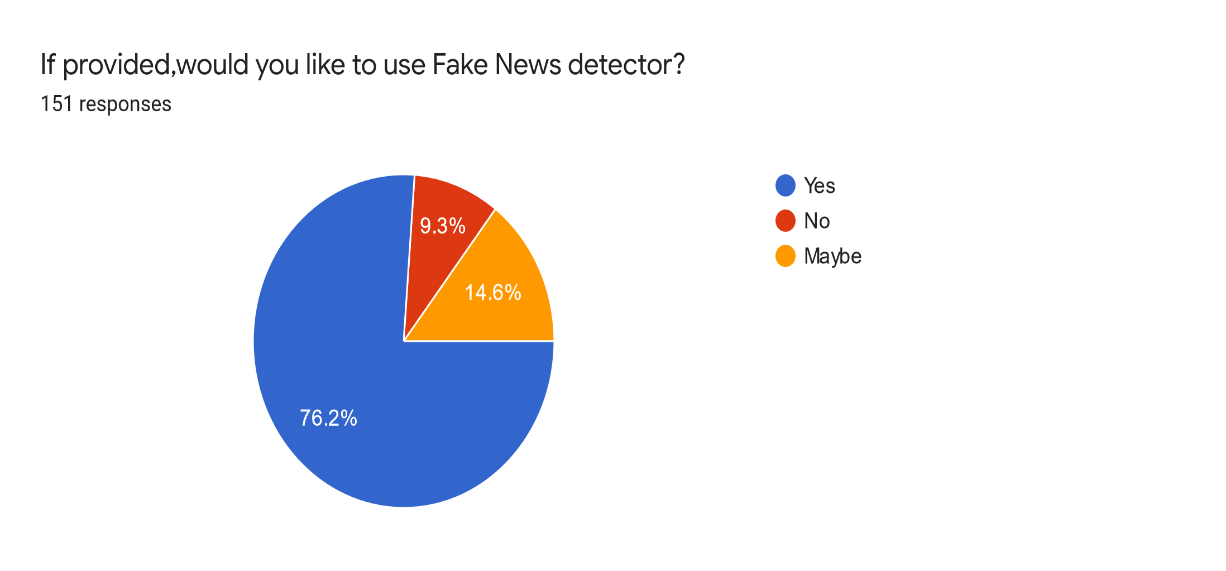
*Fig*-6

6- Have you ever used any Fake News Detection web-application, to find whether a news is real or fake?



*Fig*-7

7- If provided, would you like to use Fake News detector?



*Fig*-8

**3.2 Data Requirements:**

We have used lots of data in our project as we needed to train our machine more and more. The datasets we used in this study are open source and freely available online. The data includes both fake and truthful news articles from multiple domains. The domains it covered include sports, entertainment, and politics. The articles are not limited to a single domain such as politics as they include both fake and true articles from various other domains. The truthful news articles published contain true description of real-world events, while the fake news websites contain claims that are not aligned with facts. With help of websites like Kaggle.com we got various versions of the data what we wanted in CSV (Comma Separated Value) files. We mainly combined all of them into two CSV files.

One for real news and another for fake news. Both of our True news dataset and Fake news dataset contains approx. 27000 rows of news and 5 columns. Later we manipulate both these files and use them to train our model.

**3.3 Functional Requirements:**

* Take the input news from the user.
* Input goes through a certain form of vectorization.
* Correctly classify news input as fake news or credible news using logistic regression machine learning models/algorithm.
* Give the user output.
* Ask user if our machine has predicted it right or wrong.

**3.4 Non-Functional Requirements**

* The system should respond to a user query and return a result in less than 5 seconds.
* Web crawling should be done in fast time.
* Feature extraction must be done in milliseconds.
* Time taken by ML algorithms should be in milliseconds.
* System should be able to handle multiple simultaneous requests.

**3.5 System Specification**

This section includes hardware and software specifications required for the system to run:

**3.5.1 Hardware specification**

Our application is a web application, so it needs to be run on a web browser. The mobile interface is not yet provided so it can only be run on any computer or laptop for now. As it is a web application, we need an internet connection too.

**3.5.2 Software Specification**

This software does not require a particular Operating System (Windows, Mac, Linux, etc.) to run but only a web browser and internet connection because it is a web application.

**CHAPTER 4 – DESIGN**

**4.1 Software Requirements Specification:**

The SRS is a specification for a specific software product, program, or set of applications that perform functions in a specific environment. It serves several goals depending on who is writing it. The production of the requirements stage of the software development process is Software Requirements Specifications (SRS) (also called a requirements document). This report lays a foundation for software engineering activities and is constructing when entire requirements are elicited and analyzed. SRS is a formal report, which acts as a representation of software that enables the customers to review whether it (SRS) is according to their requirements. Also, it comprises user requirements for a system as well as detailed specifications of the system requirements.

**4.1.1 Use Case Model:**

The Use-case model is defined as a model which is used to show how users interact with the system to solve a problem. Here, User has two use cases namely, Enter news and Watch live news. When the user enters news two other use cases <includes> Check news and View & direct to trusted news sources.

Graphical user interface, website

Description automatically generated

*Fig*-9

**4.2 Conceptual level activity diagram:**

Activity diagram is basically a flowchart to represent the flow from one activity to another activity. The activity can be described as an operation of the system. The flow starts and there are two activities; Enter News to predict and Watch Live News Feeds. After entering news, machine detects it and a decision is made whether the news entered is true or false. The system shows results with additional feature of redirection to trusted news source.

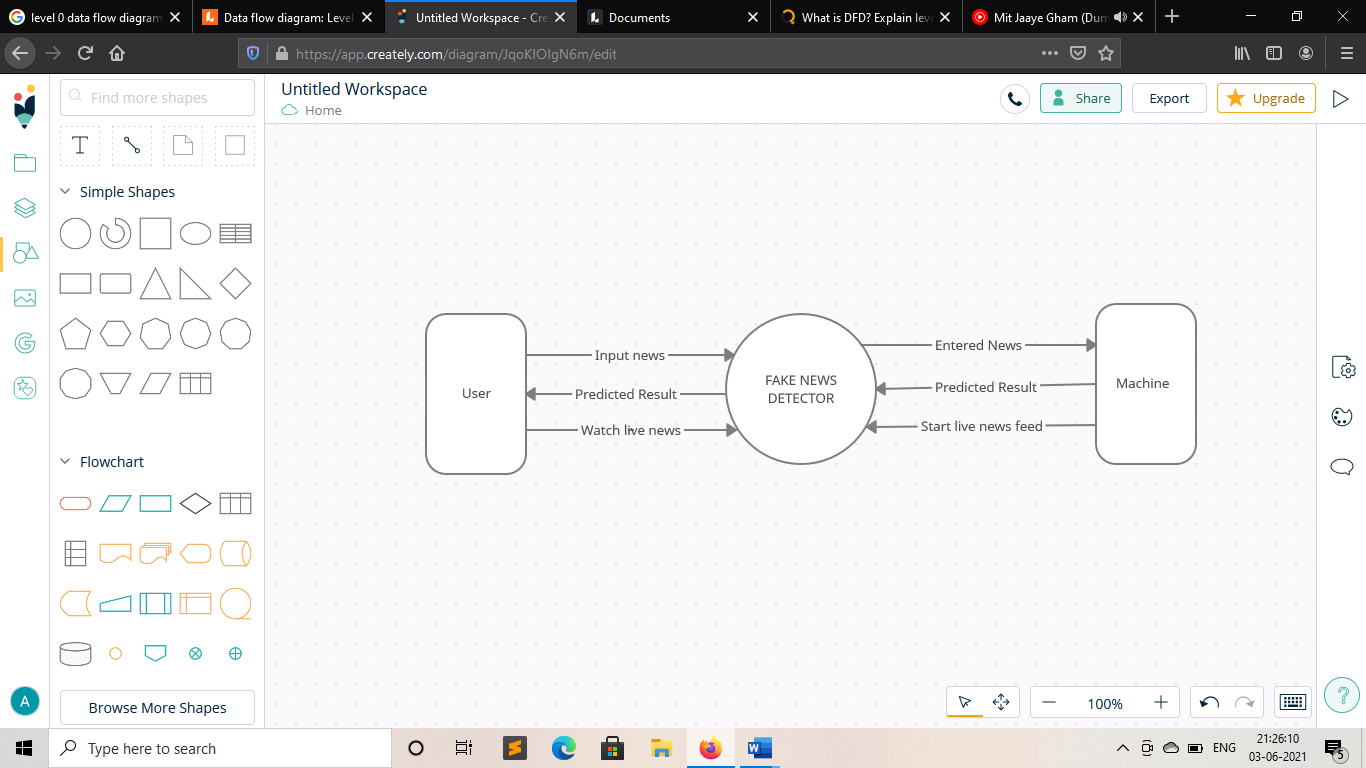
Diagram

Description automatically generated

*Fig*-10

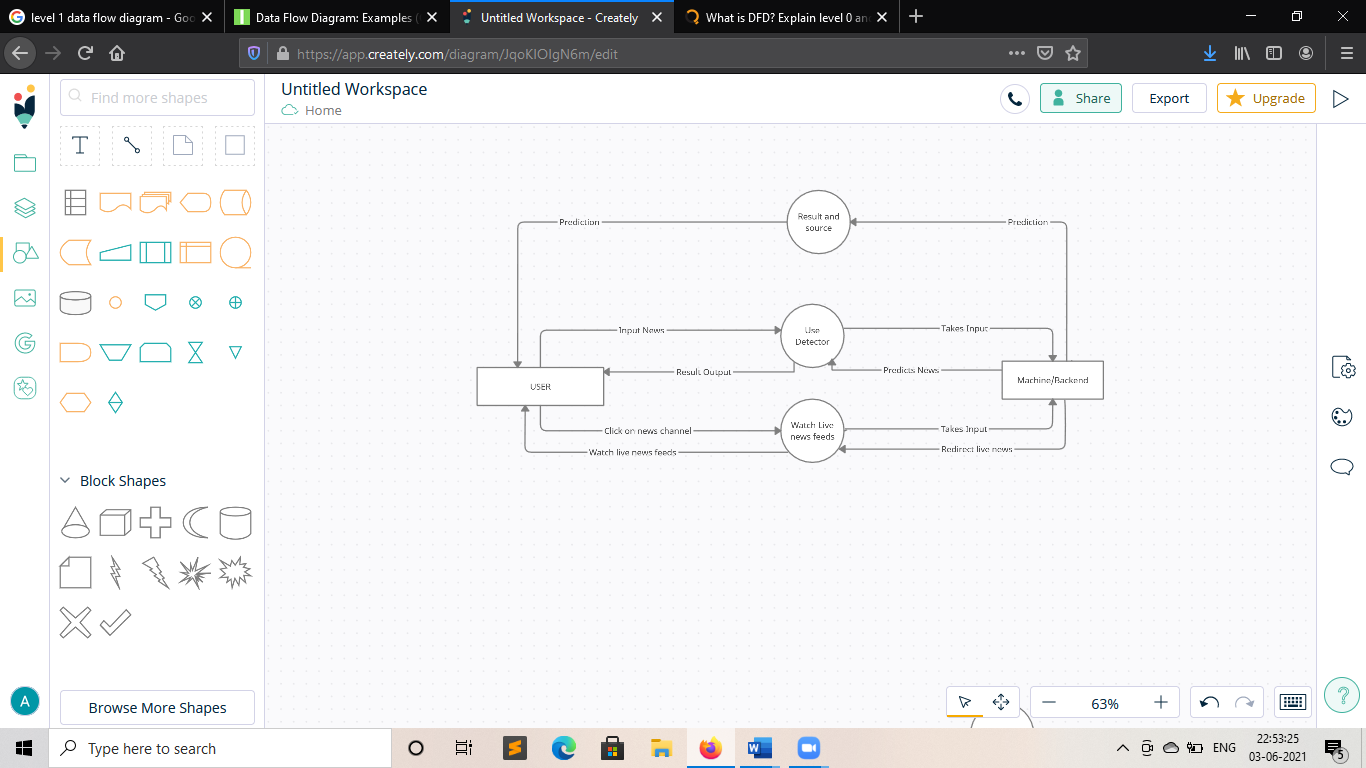
**4.3 Data flow Diagram (Level 0,1):**

DFD describes the processes that are involved in a system to transfer data from the input to the file storage and reports generation.

**DFD- Level 0:**

*Fig*-11

**DFD-Level 1:**



*Fig*-12

**CHAPTER 5- SYSTEM MODELLING**

**5.1 Interaction Diagram:**

The interaction diagram helps to envision the interactive (dynamic) behavior of any system. It portrays how objects residing in the system communicates and connects to each other.

**5.1.1 Sequence Diagram:**

Sequence Diagrams are interaction diagrams that detail how operations are carried out. They capture the interaction between objects in the context of a collaboration.

Diagram

Description automatically generated

*Fig*-13

**5.2 State Diagram:**

A state diagram is used to represent the condition of the system or part of the system at finite instances of time. It’s a behavioral diagram and it represents the behavior using finite state transitions.

Diagram

Description automatically generated

*Fig*-14

**5.3 Activity Diagram:**

Activity diagram is basically a flowchart to represent the flow from one activity to another activity. The activity can be described as an operation of the system. The control flow is drawn from one operation to another. This flow can be sequential, branched, or concurrent. Activity diagrams deal with all type of flow control by using different elements such as fork, join, etc.

Diagram

Description automatically generated

*Fig*-15

**5.4 Component Diagram:**

Component diagrams are used to describe the physical artifacts of a system. This artifact includes files, executables, libraries, etc. Component diagrams are used during the implementation phase of an application. However, it is prepared well in advance to visualize the implementation details.

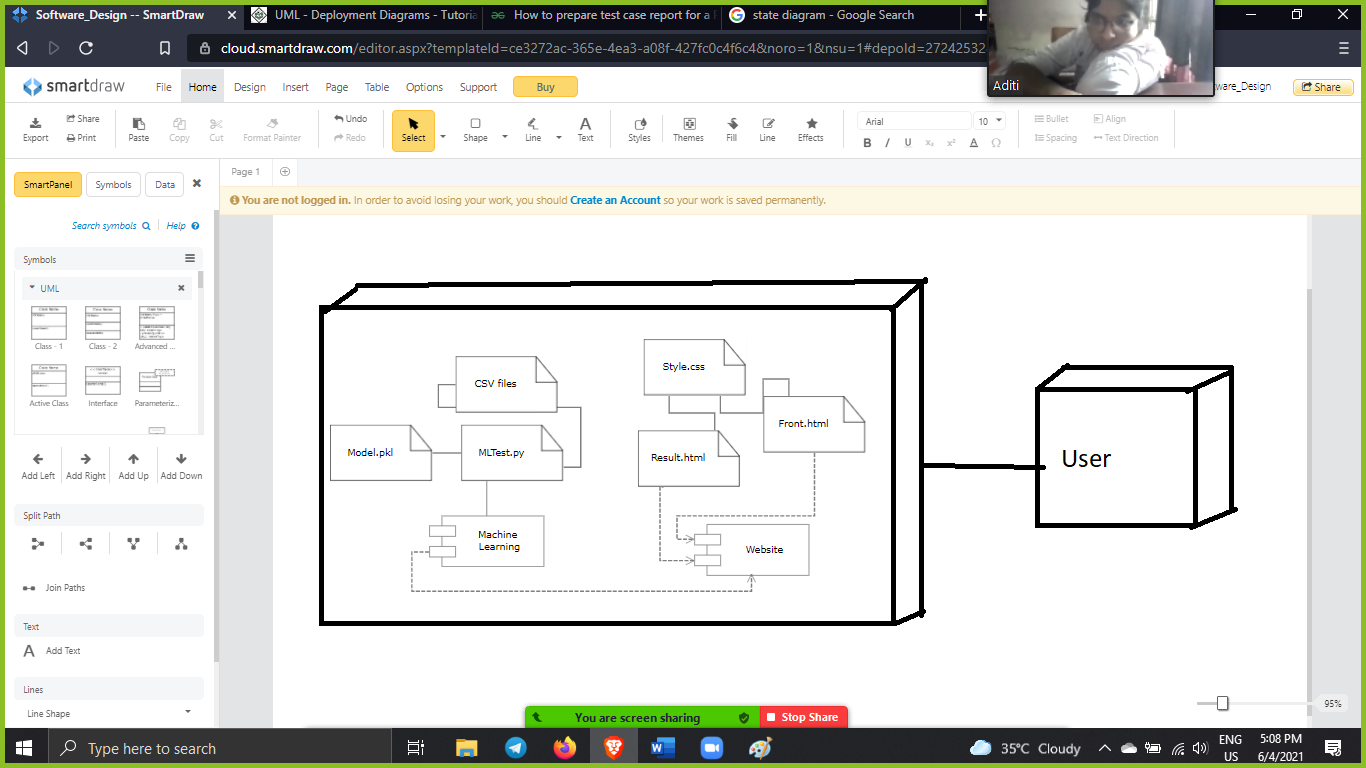
A screenshot of a computer

Description automatically generated with medium confidence

*Fig*-16

**5.5 Deployment Diagram:**

Deployment diagram is a diagram that shows the configuration of run time processing nodes and the components that live on them. It is a kind of structure diagram used in modeling the physical aspects of an object-oriented system.



*Fig*-17

**5.6 Test Plans and Implementation Images:**

We will test two components of our project which are:

* Machine Learning Model
* Front End

1. Machine Learning Model Testing:

We tested out machine learning model with different sets of data and here it is how it went:

Table 1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Test case Id | Test case Objective | Steps | Input Data | Expected Output | Actual Output | Status |
| Tc01 | To test our ml model with self generated testing data. | Input data in the textbox and let the model predict. | n number of fake and real news. | 95-100% accuracy | 98.67% accurate. | PASS |
| Tc02 | To test our ml model with current news from the internet. | Take data from a news website and paste in the textbox. | 10 Real News. | All 10 news are real. | 8 of them were detected real rest 2 fake. | PASS |
| Tc03 | To test our ml model with short amount of data and of different languages. | Take data from internet and paste it. | 10 real news, 10 fake news. | 50% accuracy as we didn’t train our model with diff languages. | 0% accurate. | FAIL |
| Tc04 | To test our ml model with short news data as we have trained the model more now. | Take data from internet and paste it. | 10 real news, 10 fake news. | 70-100% accuracy | 70% accurate. | PASS |

2. Front End Testing:

We implemented Testing on our Frontend using the review method:

1. We gave some users our front end and give us feedback.
2. We got the feedback that it basic and boring.
3. We made some changes and added some additional features to out front end.
4. Then we again gave it to some users and this time we got better feedback that our webpages looking more diligent and additional features are really helpful.

**Implementation Images:**

1- Machine Learning Testing:

A picture containing text, screenshot, computer

Description automatically generated

*Fig*-18

Text

Description automatically generated

*Fig*-19

2- Front End Testing:

i- Initial Implementation-

A picture containing text, monitor, electronics, computer

Description automatically generated

*Fig*-20

A screenshot of a computer

Description automatically generated

*Fig*-21

ii-Final Implementation:

A computer screen capture

Description automatically generated with low confidence

*Fig*-22

A screenshot of a computer

Description automatically generated

*Fig*-23

Graphical user interface, website

Description automatically generated

*Fig*- 24

Graphical user interface, application

Description automatically generated

*Fig*-25

**CHAPTER 6- CONCLUSION & FUTURE WORK**

The task of classifying news manually requires in-depth knowledge of the domain and expertise to identify anomalies in the text. In this research, we discussed the problem of classifying fake news articles using machine learning models and ensemble techniques. The data we used in our work is collected from the World Wide Web and contains news articles from various domains to cover most of the news rather than specifically classifying political news. The model is trained using an appropriate dataset and performance evaluation is also done using various performance measures. The best model, i.e., the model with highest accuracy is used to classify the news headlines or articles. As evident above for static search, our best model came out to be Logistic Regression with an accuracy of 98.57%. Hence, we can say that if a user feed a particular news article or its headline in our model, there are 98.75% chances that it will be classified to its true nature. The user can check the news article or keywords online; he can also check the authenticity of the website. The accuracy for dynamic system is 98% and it increases with every iteration. In the 21st century, the majority of the tasks are done online. Newspapers that were earlier preferred as hard- copies are now being substituted by applications like Facebook, Twitter, and news articles to be read online. WhatsApp forwards are also a major source. The growing problem of fake news only makes things more complicated and tries to change or hamper the opinion and attitude of people towards use of digital technology. When a person is deceived by the real news two possible things happen- People start believing that their perceptions about a particular topic are true as assumed.

**6.1** **Limitation of Project:**

Limitations of our project are obvious as it is the first step towards fake news detection using machine learning.

* Fake news usually spreads on mobile phone, and we only have a pc interface for now.
* Very few people will come to the pc website after looking the news on mobile phone.
* Language is a very big barrier as for now our machine has learned only with English language data.
* People needs a more efficient way to check if the news is fake or not.
* We have trained out data as good as possible but with new cases happen every day in this world our ML model needs to be trained frequently.

**6.2 Future Enhancement:**

* Main goal of using Machine Learning for using Fake news detection is that any machine any device could itself automatically detect a fake news as soon as possible it is shared or sensed.
* Implementing this inside an app.
* Collecting data of different languages and training our model more and more with that data, so that it can work and detect news of different languages.
* Collecting fresh data from users too, to train our model more and more.

**CHAPTER 7- BIBLIOGRAPHY AND REFERENCES**

7.1 **Reference Book & Other Resources**:

* D. M. J. Lazer, M. A. Baum, Y. Benkler et al., “The science of fake news,” Science, vol. 359, no. 6380, pp. 1094–1096, 2018.
* J. Hua and R. Shaw, “Corona virus (covid-19) “infodemic” and emerging issues through a data lens: the case of China,” International Journal of Environmental Research and Public Health, vol. 17, no. 7, p. 2309, 2020.
* Douglas, “News consumption and the new electronic media,” The International Journal of Press/Politics, vol. 11, no. 1, pp. 29–52, 2006.
* J. Wong, “Almost all the traffic to fake news sites is from facebook, new data show,” 2016.
* D. M. J. Lazer, M. A. Baum, Y. Benkler et al., “The science of fake news,” Science, vol. 359, no. 6380, pp. 1094–1096, 2018.
* Wang, W. Y. (2017). " liar, liar pants on fire": A new benchmark dataset for fake news detection.arXiv preprint arXiv:1705.00648.
* Pal, S., Kumar, T. S., & Pal, S. (2019). Applying Machine Learning to Detect Fake News. Indian Journal of Computer Science, 4(1), 7-12.
* Kaggle, Fake News Detection, Kaggle, San Francisco, CA, USA, 2018.
* T. M. Mitchell, The Discipline of Machine Learning, Carnegie Mellon University, Pittsburgh, PA, USA, 2006.
* L. Breiman, J. Friedman, R. Olshen, and C. Stone, Classification and Regression Trees, Springer, Berlin, Germany, 1984