

FAKE NEWS DETECTION USING MACHINE LEARNING

A MINI PROJECT REPORT

Submitted by

KUMAR SHASHWAT [Reg No:

RA2011003011262]

ARYAMAN ADIVYA SINGH [Reg No:

RA2011003011266]

MOHAMMED SHOAIB KHAN [Reg No:

RA2011003011289]

for the course 18CSC206J Software Engineering and Project Management

Under the guidance of
Dr. SUCHITHRA M

(ASSOCIATE PROFESSOR, DEPARTMENT OF
COMPUTING TECHNOLOGIES)

*in partial fulfillment for the award of the
degree of*

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

of

FACULTY OF ENGINEERING AND TECHNOLOGY



S.R.M. Nagar, Kattankulathur, Chengalpattu District



SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
KATTANKULATHUR-603203

BONAFIDE CERTIFICATE

Certified to be the bonafide record of work done by **Kumar Shashwat(RA2011003011262)** of Computer Science and Engineering, B.Tech Degree course in the Practical "**18CSC206J – Software Engineering and Project Management**" in **SRM Institute of Science and Science and Technology**, Kattankulathur during the academic year 2021-2022

Faculty Incharge

Head of the Department

Date:

Submitted for the University Examination held on _____ at **SRM institute of Science and Technology, Kattankulathur**

Date:

Examiner 1

Examiner 2

ABSTRACT

The prevalence of fake news has increased with the recent rise of social media, especially the Facebook News Feed, and this misinformation is gradually seeping into the mainstream media. Several factors have been implicated in the spread of fake news, such as political polarization, post-truth politics, motivated reasoning, confirmation bias, and social media algorithms.

Multiple strategies for fighting fake news are currently being actively researched, for various types of fake news. Politicians in certain autocratic and democratic countries have demanded effective self-regulation and legally-enforced regulation in varying forms, of social media and web search engines.

Through this project we have implemented a project that would detect Fake News from the list already defined. We implemented a Logistic Regression model to detect fake news we hope to achieve an accuracy of above 85% at least. This can be compared to other more complex models that have been used to detect fake news and their accuracy. A result that we hope to observe is how accurately does the model detect fake news that we have not included in our dataset that we generated.

Along with that we expect to gain more insights on what are the recognizing factors of fake news.

CHAPTER NO	TITLE	PAGE NO
	ABSTRACT	3
	LIST OF FIGURES	5
1	PROBLEM STATEMENT	7
2	STAKEHOLDERS & PROCESS MODELS	12 - 13
3	IDENTIFYING REQUIREMENTS	16- 21
4	PROJECT PLAN & EFFORT	24 - 27
5	WORK BREAKDOWN STRUCTURE & RISK ANALYSIS	29- 32
6	SYSTEM ARCHITECTURE, USE CASE & CLASS DIAGRAM	35 - 38
7	ENTITY RELATIONSHIP DIAGRAM	40 - 44
8	DATA FLOW DIAGRAM	47- 49
9	SEQUENCE & COLLABORATION DIAGRAM	51 - 52
10	DEVELOPMENT OF TESTING FRAMEWORK/USER INTERFACE	54 - 57
11	TEST CASES & REPORTING	60 - 63
12	ARCHITECTURE/DESIGN/FRAMEWORK/IMPLEMENTATION	72 -106
	CONCLUSION	107

LIST OF FIGURE

FIGURE NAME	FIGURE PAGE NUMBER
AGILE MODEL	12
WORK BREAKDOWN SYSTEM	29
GANTT CHART	31
RISK ANALYSIS	30
RISK MANAGEMENT	32
RISK CULTURE IMPACT	32
SYSTEM ARCHITECTURE	35
PIPE-FILTER DIAGRAM	35
HYBRID MODEL	36
SYSTEM USE CASE	37
CLASS DIAGRAM	38
ENTITY RELATIONSHIP DIAGRAM	42
DATA FLOW DIAGRAM	47-48
SEQUENCE DIAGRAM	50
COLLABORATION	51



School of Computing

SRM IST, Kattankulathur – 603 203

Course Code: 18CSC206J

Course Name: Software Engineering and Project Management

Experiment No	1
Title of Experiment	To identify the Software Project, Create Business Case, Arrive at a Problem Statement
Name of the candidate	KUMAR SHASHWAT
Team Members	MOHAMMED SHOAIB KHAN ARYAMAN ADIVYA SINGH
Register Number	RA2011003011262
Date of Experiment	15-03-2022

Mark Split Up

S.No	Description	Maximum Mark	Mark Obtained
1	Exercise	5	
2	Viva	5	
Total		10	

Staff Signature with date

Aim - To Frame a project team, analyze and identify a Software project. To create a business case and Arrive at a Problem Statement for the <title of the project>

Team Members:

S. No	Register No	Name	Role
1	RA2011003011266	ARYAMAN ADIVYA SINGH	Lead/Rep
2	RA2011003011262	KUMAR SHASHWAT	Member
3	RA2011003011289	MOHAMMED SHOAIB KHAN	Member

Project Title:

FAKE NEWS DETECTION using Python and Machine Learning

Project Description:

Basic Description:

A pioneering and state-of-the-art Machine Learning Model that detects Fake News being circulated across the internet, using it as a propaganda against an individual, society, organization or a political party. Such news perhaps contain false claims or deceptive articles, spreading virally across social platforms, with ending up users in a false bubble. Fake news publishers often have malicious intent to spread distorted and misleading information and influence large communities of consumers, not seen in true news articles. Specifically, the model will introduce the definition of Fake News Detection through various approaches and orientations.

Business Case

ONE PAGE BUSINESS CASE TEMPLATE

DATE	15-03-2022
SUBMITTED BY	ARYAMAN ADIVYA SINGH
TITLE / ROLE	THE LEADER

THE PROJECT

In bullet points, describe the problem this project aims to solve or the opportunity it aims to develop.

- Fake News Model **detects the misleading information**, articles and posts being circulated across the internet
- The Model provides an **effective approach for big organizations** and communities to prevent propaganda being spread across social media platforms
- The Model **tags the authenticity** of Authors and Articles and debar Fake Write-ups being spread on social media or any internet platforms

THE HISTORY

In bullet points, describe the current situation.

- Fake and misleading bulletins lead to communal harm and important social & political issues.
- Fake Stories can lead a user to make harmful decisions for their health or personal life.
- Inaccurate and false information can cause tremendous real world impacts and cause mass hysteria among the masses
- Affect of Fake News specially on Social media such as Facebook and Instagram can cause Global Muddles, as big as fluctuations in GDP per capita or even War.

LIMITATIONS

List what could prevent the success of the project, such as the need for expensive equipment, bad weather, lack of special training, etc.

- **Lack of updated Real World Data** can cause it to work slowly, limiting its effectiveness
- **Non-Implementation into Big Data** of Large Industries may cause Model Training to be outdated
- **Absence of 24/7 online servers** functioning properly can cause crashing of trained model making it inoperative.

APPROACH

List what is needed to complete the project.

- An Initial Fake News Dataset to build our trainable model.
- Using tools such as Sklearn,TfidVectorizer and PassiveAggressive Classifier to build our trained and reusable model.
- Using real world web development skills and tools to build an actual operative Fake News Detection Model.

BENEFITS

In bullet points, list the benefits that this project will bring to the organization.

- Automated Index Scoring on Authors on the articles they publish
- Stopping the upload of any fake news article on a social media through integrated tools
- Blocking all Spammers and Bot Detection
- Trending Rumor Classification and Clickbait detection

Result

Thus, the project team formed, the project is described, the business case was prepared and the problem statement was arrived.



School of Computing

SRM IST, Kattankulathur – 603 203

Course Code: 18CSC206J

Course Name: Software Engineering and Project Management

Experiment No	2
Title of Experiment	Identification of Process Methodology and Stakeholder Description
Name of the candidate	KUMAR SHASHWAT
Team Members	MOHAMMED SHOAIB KHAN ARYAMAN ADIVYA SINGH
Register Number	RA2011003011262
Date of Experiment	30-03-2022

Mark Split Up

S.No	Description	Maximum Mark	Mark Obtained
1	Exercise	5	
2	Viva	5	
Total		10	

Staff Signature with date

Aim

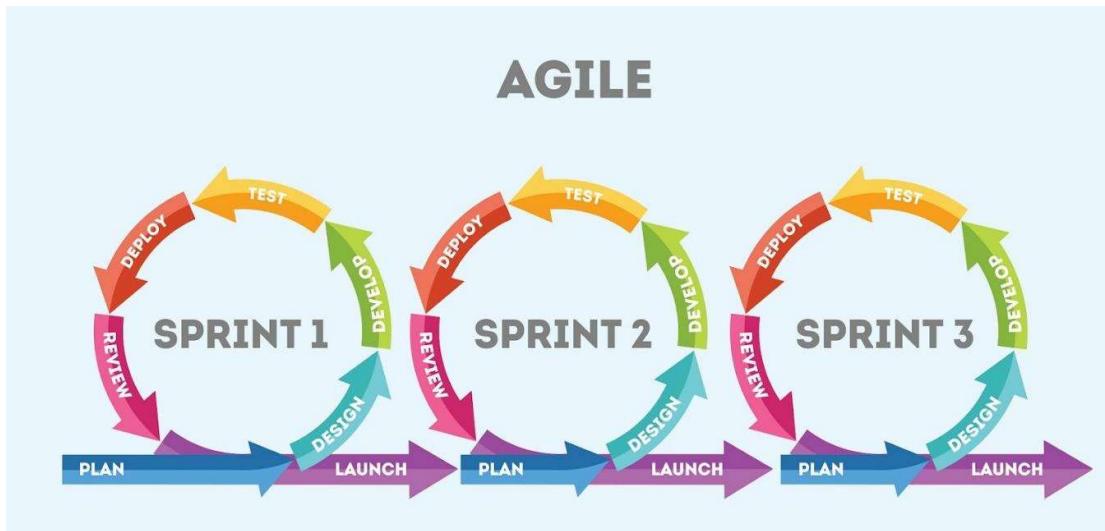
To identify the appropriate Process Model for the project and prepare Stakeholder and User Description.

Team Members:

SI No	Register No	Name	Role
1	RA201100301266	ARYAMAN ADIVYA SINGH	Rep/Member
2	RA2011003011262	KUMAR SHASHWAT	Member
3	RA2011003011289	MOHAMMED SHOAIB KHAN	Member

FAKE NEWS DETECTION SYSTEM

Selection of Methodology



Agile is change of rapid development and deployment, meaning that starting with the basic of software programming that is planning. Now, since we know what we are about to build **Fake News Detection application**, we breakdown all of the activity or process into *small chunks of services* and then we work on those small services, one at a time which ensures that we follow the *micro services model* and at the same time we don't really affect the entire application in general.

So, we plan, we design, we develop or architect, we test, we deploy and then we review it. We notice that launch is actually outside the entire circle, meaning every time we make a change, (*it could be as simple as a one line of code, say changing a variable name*) no matter how big or small that change is, it has to be first deployed in a **DevOps** environment to get a constant feedback over what's happening with the code that is updated or added.

Benefit of working with agile is that we are able to *work in iterations or sprint*, in which we need not worry about the whole application all at once rather we can focus on micro services so that work gets faster and easier. It's basically like creating a multiple function in any programming language rather than writing all big codes in the main function itself. Like even if all micro services are in synergy yet they are completely different tasks.

STAKEHOLDER REGISTER

Stakeholder Name	Activity/ Area /Phase	Interest	Influence	Priority (High/ Medium/ Low)
Enterprise Manager John McCavern	Success for whole project/Increased development capability	High	High	1
Board of Directors Albert Russ	Success of project by Oversight of the Enterprise, and Establishing policies	High	High	1
National and International Authorities (at all levels)	Legislative framework	High	Low	2
Corporate Mr. Alex	Funding the Project	High	Low	2
Head of Dep. IT Ms. Suiz	Responsible in Team Management and Product Delivery	High	High	1
Development Team Mr. Lionel	Directly involved in delivering Project	High	Low	2
Local Community Mr. Carl Johnson	New features and Ease of use for the whole system	High	Low	1
Communication Media Mr. Suarez	General Public concern and Media Impact through Transparency	Low	High	2

Result

Thus the *Project Methodology* was identified and the *Stakeholders* were described.



School of Computing

SRM IST, Kattankulathur – 603 203

Course Code: 18CSC206J

Course Name: Software Engineering and Project Management

Experiment No	3
Title of Experiment	System, Functional and Non-Functional Requirements of the Project
Name of the candidate	KUMAR SHASHWAT
Team Members	MOHAMMED SHOAIB KHAN, ARYAMAN ADIVYA SINGH
Register Number	RA2011003011262
Date of Experiment	13-04-2022

Mark Split Up

S.No	Description	Maximum Mark	Mark Obtained
1	Exercise	5	
2	Viva	5	
Total		10	

Staff Signature with date

Aim:

To identify the system, functional and non-functional requirements for the project.

Team Members:

S No	Register No	Name	Role
1	RA2011003011266	ARYAMAN ADIVYA SINGH	Rep/Member
2	RA2011003011289	MOHAMMED SHOAIB KHAN	Member
3	RA2011003011262	KUMAR SHASHWAT	Member

FAKE NEWS DETECTION using Python and Machine Learning

System Requirements

Hardware Requirements:

Minimum Requirements

CPU: Single Core 2.4 GHz

RAM: 512 MB

Graphics Card: Nvidia GeForce 5xxx series or equivalent

Operating System: Windows XP

Hard Drive: 5 Gigabytes

Network: Broadband Recommended

Recommended Requirements

CPU: Duo Core or higher

RAM: 4 GB

Graphics Card: Nvidia GeForce 7xxx series or equivalent

Operating System: Windows 7 or higher

Hard Drive: 6 Gigabytes or more

Network: Broadband Recommended

Operating System Requirements

Windows

To use the Application on the browser of Windows, you'll need:

- Windows 7, Windows 8, Windows 8.1, Windows 10 or later
- An Intel Pentium 4 processor or later that's SSE3 capable

Note: Servers require Windows Server 2008 R2, Windows Server 2012, Windows Server 2012 R2, or Windows Server 2016.

Mac

To use the Application on the browser of Mac, you'll need:

- OS X El Capitan 10.11 or later

Linux

To use the Application on the browser of Linux, you'll need:

- 64-bit Ubuntu 18.04+, Debian 10+, openSUSE 15.2+, or Fedora Linux 32+
- An Intel Pentium 4 processor or later that's SSE3 capable

Android

To use Web Application on the browser of Android, you'll need:

- Android Marshmallow 6.0 or later

Web Browser Requirements

Web / Mobile Browser	Infrastructure Classic Console and Applications Console
Microsoft Internet Explorer	11 or later
Mozilla Firefox	52 or later
Google Chrome	63 or later
Apple Safari	10 or later
Microsoft Edge	35 or later
Safari, Chrome, Firefox on iOS (iPad and iPhone)	Latest
Chrome, Firefox on Android (Phone and Tablet)	Latest

Functional Requirements

Login:

- The system should provide a login page for the user to log into the system.
- If login is successful then the user can access the system.
- else it would display “Login Credentials Incorrect”.

Search:

- The system should display a search option where users can search for news.
- If the news is already in the database then it will provide the article and also display if the news is fake or real.

Upload:

- If the news is not present in the database then, the user will get the option to upload the news title, author, and article.
- The system will then provide the data to the machine learning model which then displays if the news is real or fake.

Update:

- After the processing of the news, the system will add the news article to its database so that in the future no need to process the article again.

Verification:

- The system will provide a feedback system to users which will allow the user to tell if the news is real or fake and matches the result with the result provided by the algorithm.
- If the result is not correct then the system will update the result and its accuracy score.

Logout:

- The system will provide a logout option to log out of the system.

Non-Functional Requirements

Performance and scalability.

The application returns the result instantly provided it meets the user system requirement. There won't be any changes in the performance or response time with the higher workload.

Portability and compatibility.

Minimum hardware devices are iPhone 6 and above, Android devices that support Android Kitkat and above including android one. Browsers with which this application will be compatible are Internet Explorer 10 and above. Google Chrome V19.0.1084 and above. Safari 5.1.10 and above. Firefox 10 and above. The Operating System for this application to run is Windows. It is an application that runs independently without conflicting with other processes.

Reliability, availability, maintainability.

Weekly maintenance is required for this application, to segregate the news that has been fetched by the user for it to be organized on the basis of its accuracy score, also modularize based upon the user's feedback, and also update the firewall and security options to prevent any cyber-attacks.

We would try to lessen the downtime as per the maintenance requirement but it will certainly not be anything too long.

Security.

All communication should be using HTTPS over TLS 1.2. All APIs will have auth token (JWT token with SHA256 algorithm). All Personal Identifiable data would be always encrypted. Files on servers would be encrypted. All documents generated should be password protected, display an error message in case of the wrong password with the date it was set, and a hint of standard documents – like '*this is the same password as all your other documents*'. Do not add sensitive data openly to request payload. Use OAuth 2.0 for authentication using popular IDPs such as Google, MS ADFS, etc.

Usability.

UX is crisp and consistent across all major devices, standard and non-standard accepted viewports and resolutions are a standard requirement. The navigation is smooth without halts or glitches and follows standard navigation guidelines:

- Modular UI architecture
- Load data fast (<2 seconds)
- Use as little data as possible
- Use static assets from a local cache
- Separate content from navigation
- Retrieve and display page-specific content (HTML, JSON, etc.)
- Optionally, cache dynamic content

Result:

Thus the requirements were identified and accordingly described.



School of Computing

SRM IST, Kattankulathur – 603 203

Course Code: 18CSC206J

Course Name: Software Engineering and Project Management

Experiment No	4
Title of Experiment	Prepare Project Plan based on scope, Calculate Project effort based on resources and Job roles and responsibilities
Name of the candidate	KUMAR SHASHWAT
Team Members	MOHAMMED SHOAIB KHAN ARYAMAN ADIVYA SINGH
Register Number	RA2011003011262
Date of Experiment	11-05-2022

Mark Split Up

S.No	Description	Maximum Mark	Mark Obtained
1	Exercise	5	
2	Viva	5	
Total		10	

Staff Signature with date

Aim

To Prepare Project Plan based on scope, Calculate Project effort based on resources, Find Job roles and responsibilities

Team Members:

Sl No	Register No	Name	Role
1	RA2011003011266	ARYAMAN ADIVYA SINGH	Lead
2	RA2011003011262	KUMAR SHASHWAT	Member
3	RA2011003011289	MOHAMMED SHOAIB KHAN	Member

1. Project Management Plan

Describe the key issues driving the project. [Min 3 Focus Areas]

Focus Area	Details
Integration Management	Governance Framework Project Team Structure Roles & Responsibilities of Team Change Management (Change Control, Issue Management) Project Closure
Scope Management	Scope Statement Requirement Management (Gathering, Control, Assumption, Constraint Stakeholder) Define Deliverable Requirement Change Control Activities and Sub-Tasks
Schedule Management	Define Milestones Schedule Control
Cost Management	Estimate Effort Assign Team Budget Control
Quality Management	Quality Assurance: Quality assurance will be managed including governance, roles and responsibilities, tools and techniques and reporting Quality Control: Specify the mechanisms to be used to measure and control the quality of the work products
Resource Management	Estimate and Manage the need People: People & Skills Required Finance: Budget Required Physical: Facilities, IT Infrastructure
Stakeholder	Identifying, Analyzing, Engaging Stakeholders
Communication Management	Determine communication requirements, roles and responsibilities, tools and techniques. [Type of Communication, Schedule, Mechanism Recipient]
Risk Management	Identifying, analysing, and prioritizing project risks
Procurement Management	Adhering to organization procurement process

Estimation

2.1. Effort and Cost Estimation

Activity Description	Sub-Task	Sub-Task Description	Effort (in hours)	Cost in INR
Design the user screen	E1R1A1T1	Human efforts	$2 \times 10 \times 3 = 60$	60000
	E1R1A1T2	Customer Design	6	6000
	E1R1A1T3	Version Updates	4	4000
Technology Used	E1R1A2T1	Various Software and Firmware Setup	3	3000
	E1R1A2T2	Powerful and Efficient Hardware assembly	6	6000
Documentation	E1R1A3T1	Historical Data, Test Cases and DataSets' Sources	10	10000

Effort (hr)	Cost (INR)
1	1000

2.2. Infrastructure/Resource Cost [CapEx]

< OneTime Infra requirements >

Infrastructure Requirement	Qty	Cost per qty	Cost per item
Laptop	3	80000	240000
Internet Connectivity/Router	1	1000	1000

2.3 Maintenance and Support Cost [OpEx]

Category	Details	Qty	Cost per qty per annum	Cost per item
People	Network, System, Middleware and DB admin Developer , Support Consultant	3	2,000,000	6,000,000
License	Operating System Database Middleware IDE	10	10000	100,000
Infrastructures	Server, Storage and Network	20	20000	400,000

3. Project Team Formation

3.1. Identification Team members

Name	Role	Responsibilities
ARYAMAN ADIVYA SINGH	Key Business User (Product Owner)	Provide clear business and user requirements
ARYAMAN ADIVYA SINGH	Project Manager	Manage the project
ARYAMAN ADIVYA SINGH	Business Analyst	Discuss and Document Requirements
MOHAMMED SHOAIB KHAN	Technical Lead	Design the end-to-end ML architecture, Perform data preprocessing
KUMAR SHASHWAT	UX Designer	Design the user experience
KUMAR SHASHWAT	Frontend Developer	Develop user interface
ARYAMAN ADIVYA SINGH KUMAR SHASHWAT	Backend Developer	Design, Develop and Unit Test Services/API/DB
ARYAMAN ADIVYA SINGH	Cloud Architect, Operator and Deployment Manager	Design the cost effective, highly available and scalable architecture and Provision required Services
Mohammed Shoaib Khan	Tester	Define Test Cases and Perform Testing

3.2. Responsibility Assignment Matrix

RACI Matrix		Team Members		
Activity	Name (BA)	Name (Developer)	Name (Project Manager)	Key Business User
User Requirement Documentation	ARYAMAN ADIVYA SINGH	ARYAMAN ADIVYA SINGH	ARYAMAN ADIVYA SINGH	ARYAMAN ADIVYA SINGH
Software Coding	ARYAMAN ADIVYA SINGH	ARYAMAN ADIVYA SINGH	ARYAMAN ADIVYA SINGH	ARYAMAN ADIVYA SINGH
	KUMAR SHASHWAT	KUMAR SHASHWAT		
	MOHAMMED SHOAIB KHAN	MOHAMMED SHOAIB KHAN		
Model Training	MOHAMMED SHOAIB KHAN	ARYAMAN ADIVYA SINGH MOHAMMED SHOAIB KHAN	MOHAMMED SHOAIB KHAN	MOHAMMED SHOAIB KHAN
UI/UX Design	KUMAR SHASHWAT	ARYAMAN ADIVYA SINGH	KUMAR SHASHWAT	KUMAR SHASHWAT
Cloud Application Management (Heroku)	ARYAMAN ADIVYA SINGH	ARYAMAN ADIVYA SINGH	ARYAMAN ADIVYA SINGH	ARYAMAN ADIVYA SINGH
	KUMAR SHASHWAT			

A	Accountable
R	Responsible
C	Consult
I	Inform

Reference

1. <https://www.pmi.org/>
2. <https://www.projectmanagement.com/>
3. <https://www.tpsgc-pwgsc.gc.ca/biens-property/snpg-npms/ti-it/ervcpgrm-dsfvpmpeng.html>

Result:

Thus, the Project Plan was documented successfully



School of Computing

SRM IST, Kattankulathur – 603 203

Course Code: 18CSC206J

Course Name: Software Engineering and Project Management

Experiment No	5
Title of Experiment	Prepare Work breakdown structure, Timeline chart, Risk identification table
Name of the candidate	KUMAR SHASHWAT
Team Members	ARYAMAN ADIVYA SINGH MOHAMMED SHOAIB KHAN
Register Number	RA2011003011262
Date of Experiment	18/05/2022

Mark Split Up

S.No	Description	Maximum Mark	Mark Obtained
1	Exercise	5	
2	Viva	5	
Total		10	

Staff Signature with date

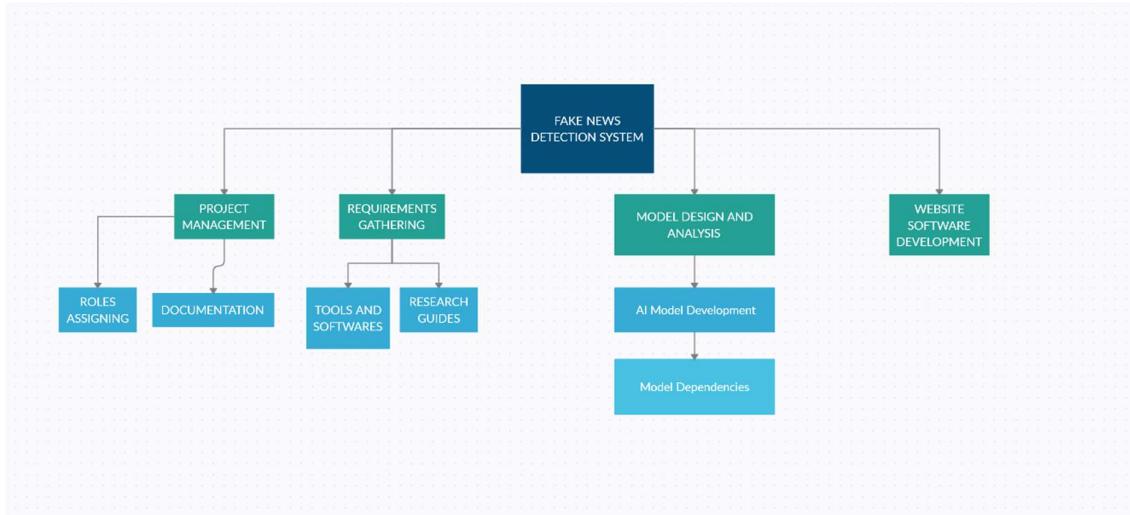
Aim

To Prepare Work breakdown structure, Timeline chart and Risk identification table

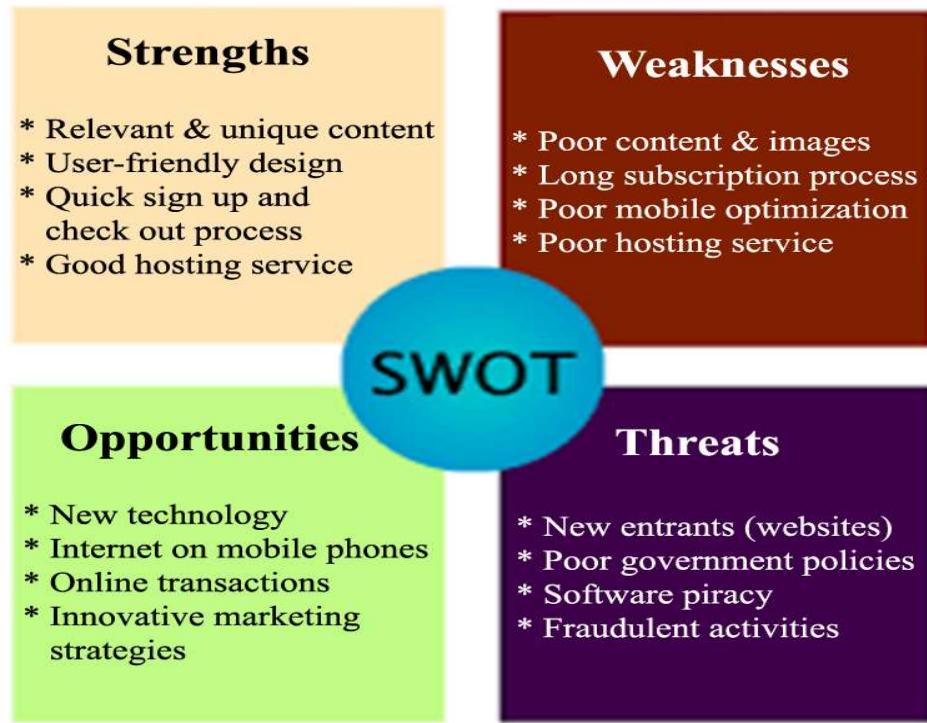
Team Members:

Sl No	Register No	Name	Role
1	RA2011003011266	ARYAMAN ADIVYA SINGH	Rep
2	RA2011003011262	KUMAR SHASHWAT	Member
3	RA2011003011289	MOHAMMED SHOAIB KHAN	Member

Work Breakdown System Diagram



RISK ANALYSIS – SWOT & RMMM



TIMELINE – GANTT CHART



Risk Management Framework- Risks And Mitigation ...

Response	Strategy	Examples
Avoid	Risk avoidance is a strategy where the project team takes action to remove the threat of the risk or protect from the impact	<ul style="list-style-type: none"> Extending the schedule Reducing/removing scope Change the execution strategy
Transfer	Risk transference involves shifting or transferring the risk threat and impact to a third party. Rather transfer the responsibility and ownership	<ul style="list-style-type: none"> Purchasing insurance Performance bonds Warranties Contract issuance (lump sum)
Mitigate	Risk mitigation is a strategy where the project team takes a action to reduce the probability of the risk occurring. This does not risk or potential impact , but rather reduces the likelihood of it becoming real.	<ul style="list-style-type: none"> Increasing testing Changing suppliers to a more stable one Reducing process complexity
Accept	Risk acceptance means the team acknowledges the risk and its potential impact, but decides not to take any preemptive action to prevent it. It is dealt with only if it occurs.	<ul style="list-style-type: none"> Contingency reserve budgets Management schedule float Event contingency

Slide 1 of 5

Safeguarding our reputation in creating long-term value for our stakeholders



Result:

Thus, WBS Gantt CHART AND RISK MITIGATION DIAGRAM WAS PRINTED SUCCESSFULLY.



School of Computing

SRM IST, Kattankulathur – 603 203

Course Code: 18CSC206J

Course Name: Software Engineering and Project Management

Experiment No	6
Title of Experiment	Design a System Architecture, Use Case and Class Diagram
Name of the candidate	KUMAR SHASHWAT
Team Members	ARYAMAN ADIVYA SINGH MOHAMMED SHOAIB KHAN
Register Number	RA2011003011262
Date of Experiment	20-05-2022

Mark Split Up

S.No	Description	Maximum Mark	Mark Obtained
1	Exercise	5	
2	Viva	5	
Total		10	

Staff Signature with date

Aim

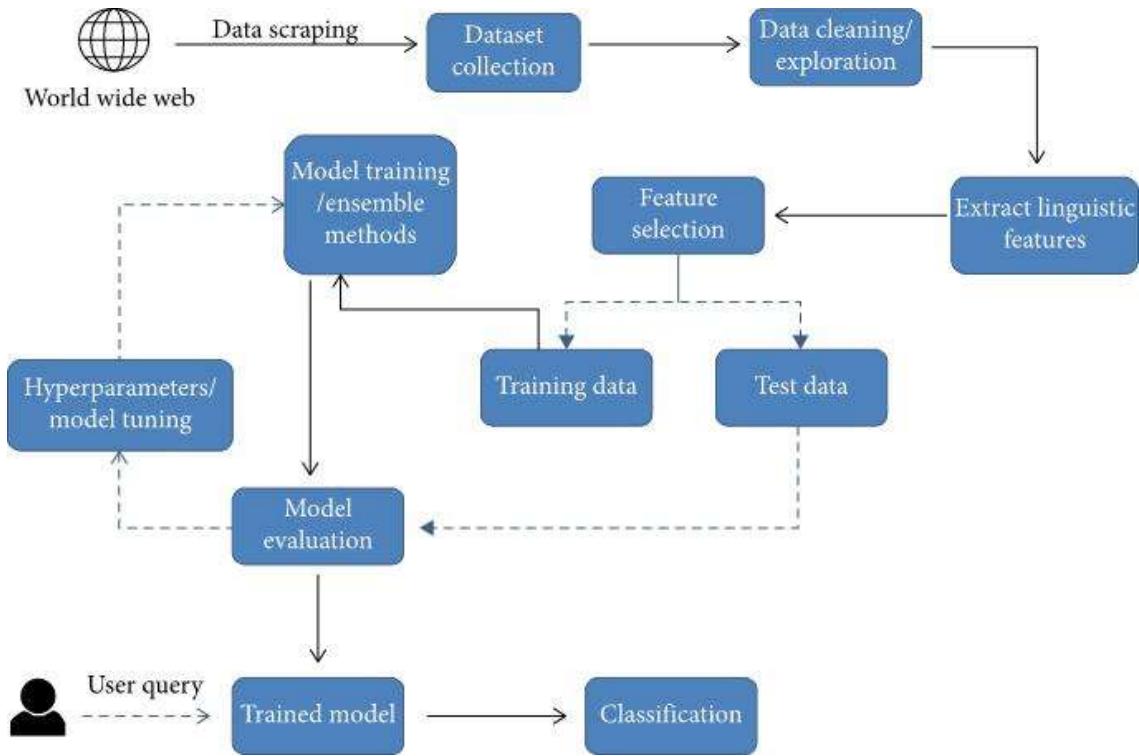
To Design a System Architecture, Use case and Class Diagram

Team Members:

Sl No	Register No	Name	Role
1	RA2011003011266	ARYAMAN ADIVYA SINGH	Rep
2	RA2011003011262	KUMAR SHASHWAT	Member
3	RA2011003011289	MOHAMMED SHOAIB KHAN	Member

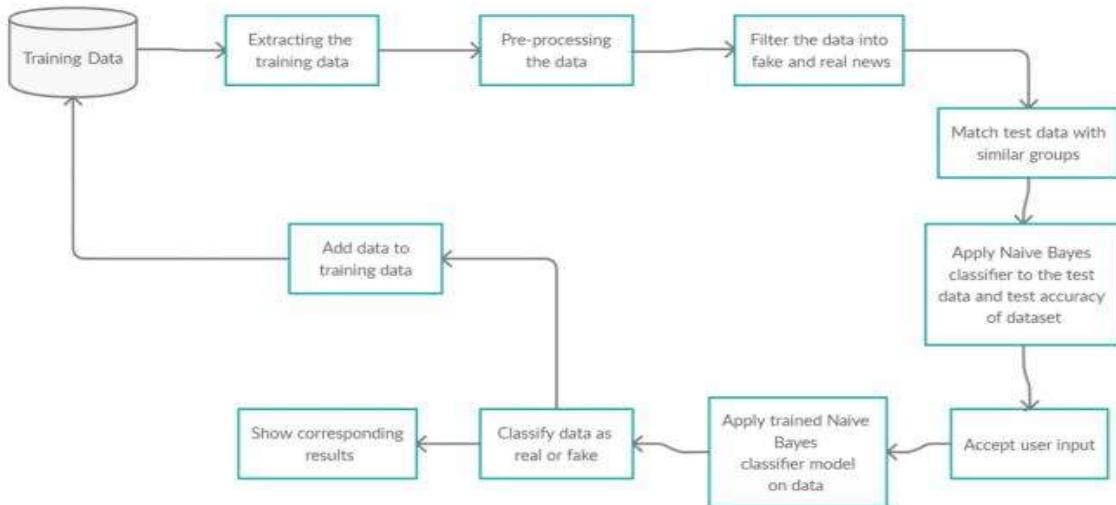
SYSTEM ARCHITECTURE

Fundamental Structure

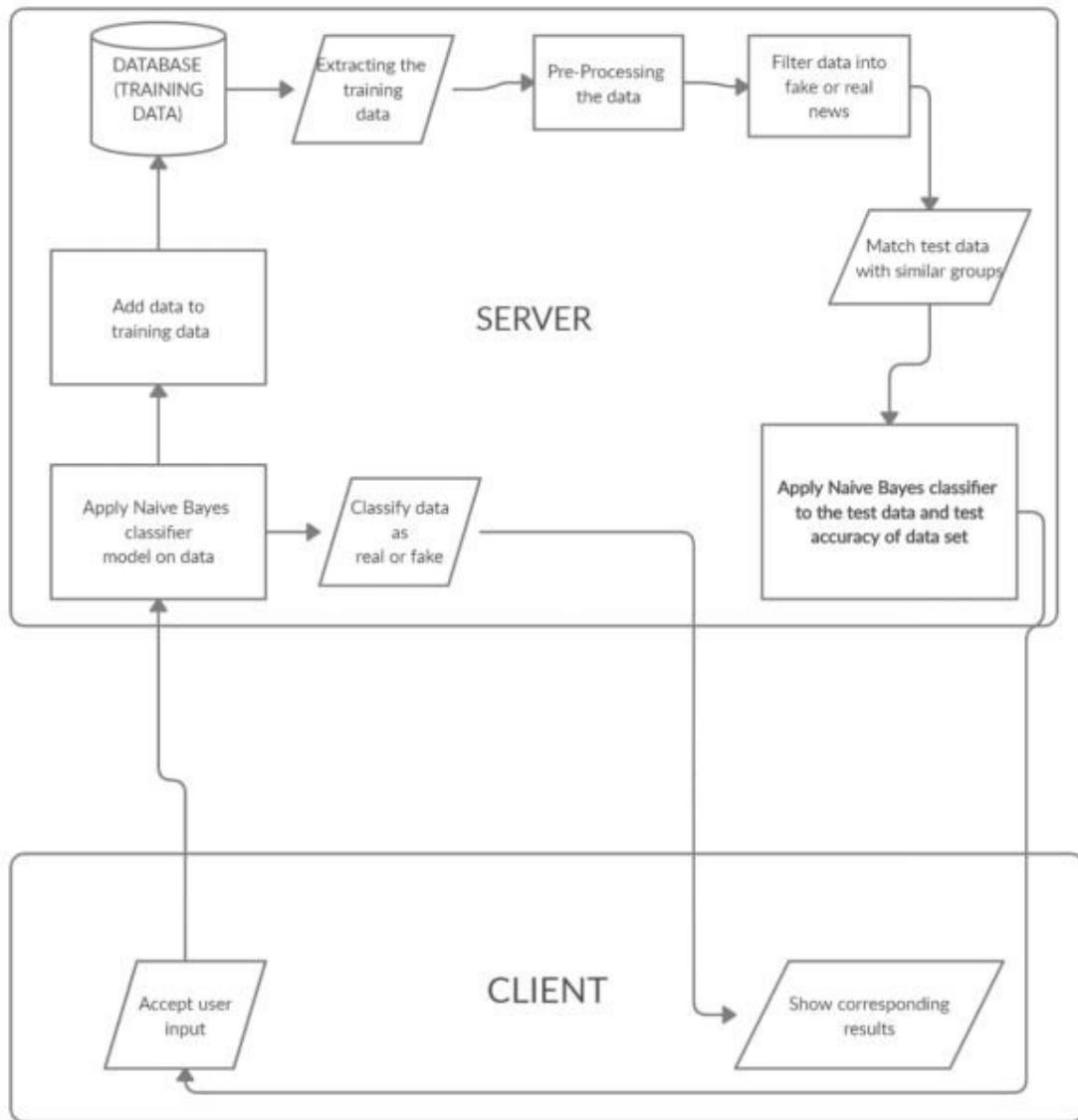


EXHAUSTIVE STRUCTURE DIAGRAM

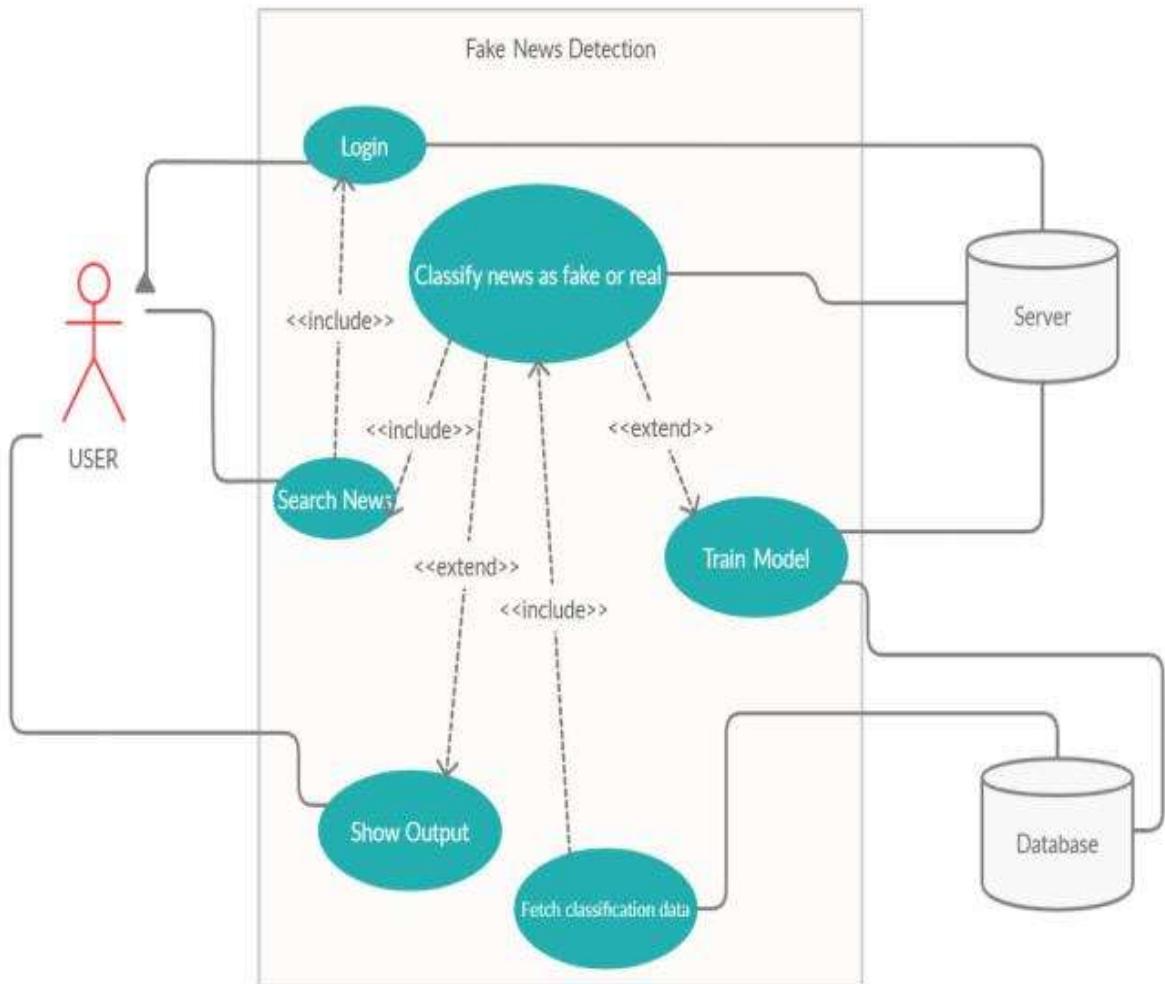
1.PIPE AND FILTER PATTERN:



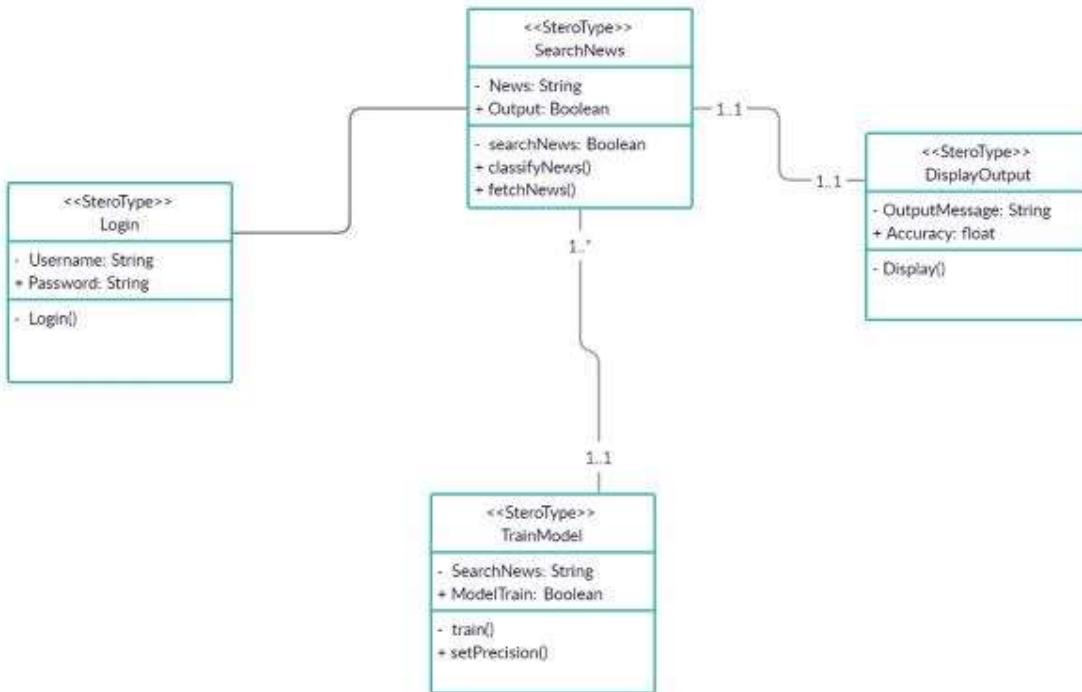
2. HYBRID MODEL WITH CLIENT-SERVER AND PIPE FILTER PATTERN:



3.SYSTEM USE CASE DIAGRAM



CLASS DIAGRAM



Result:

Thus, the system architecture, use case and class diagram created successfully.



School of Computing

SRM IST, Kattankulathur – 603 203

Course Code: 18CSC206J

Course Name: Software Engineering and Project Management

Experiment No	7
Title of Experiment	Design a Entity relationship diagram
Name of the candidate	KUMAR SHASHWAT
Team Members	MOHAMMED SHOAIB KHAN, ARYAMAN ADIVYA SINGH
Register Number	RA2011003011262
Date of Experiment	23-05-2022

Mark Split Up

S. No	Description	Maximum Mark	Mark Obtained
1	Exercise	5	
2	Viva	5	
	Total	10	

Staff Signature with date

Aim

To create the Entity Relationship Diagram

Team Members:

S No	Register No	Name	Role
1	RA2011003011266	ARYAMAN ADIVYA SINGH	Rep
2	RA2011003011262	KUMAR SHASHWAT	Member
3	RA2011003011289	MOHAMMED SHOAIB KHAN	Member

***/ ER Diagram, Notation and Example**

What is ER Diagram?

- ER Diagram stands for Entity Relationship Diagram, also known as ERD is a diagram that displays the relationship of entity sets stored in a database. In other words, ER diagrams help to explain the logical structure of databases. ER diagrams are created based on three basic concepts: entities, attributes and relationships.
- ER Diagrams contain different symbols that use rectangles to represent entities, ovals to define attributes and diamond shapes to represent relationships.
- At first look, an ER diagram looks very similar to the flowchart. However, ER Diagram includes many specialized symbols, and its meanings make this model unique. The purpose of ER Diagram is to represent the entity framework infrastructure.

What is ER Model?

- ER Model stands for Entity Relationship Model is a high-level conceptual data model diagram. ER model helps to systematically analyze data requirements to produce a well-designed database.
- ER Model represents real-world entities and the relationships between them. Creating an ER Model in DBMS is considered as a best practice before implementing your database.
- ER Modeling helps you to analyze data requirements systematically to produce a well-designed database. So, it is considered a best practice to complete ER modeling before implementing your database.

Why use ER Diagrams?

Here, are prime reasons for using the ER Diagram

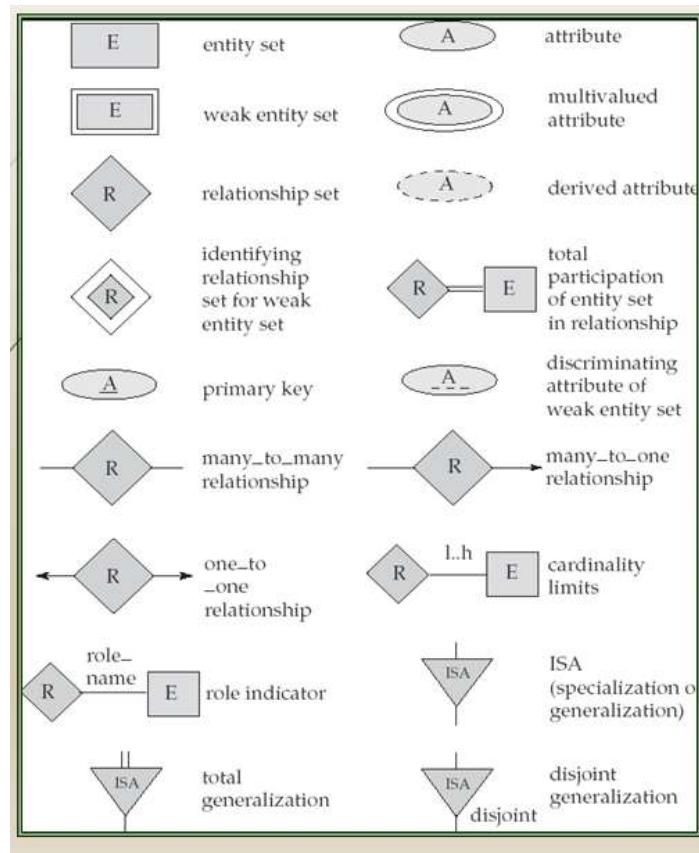
- Helps you to define terms related to entity relationship modeling
- Provide a preview of how all your tables should connect, what fields are going to be on each table
- Helps to describe entities, attributes, relationships
- ER diagrams are translatable into relational tables which allows you to build databases quickly
- ER diagrams can be used by database designers as a blueprint for implementing data in specific software applications
- The database designer gains a better understanding of the information to be contained in the database with the help of ERP diagram
- ERD Diagram allows you to communicate with the logical structure of the database to users

Components of the ER Diagram

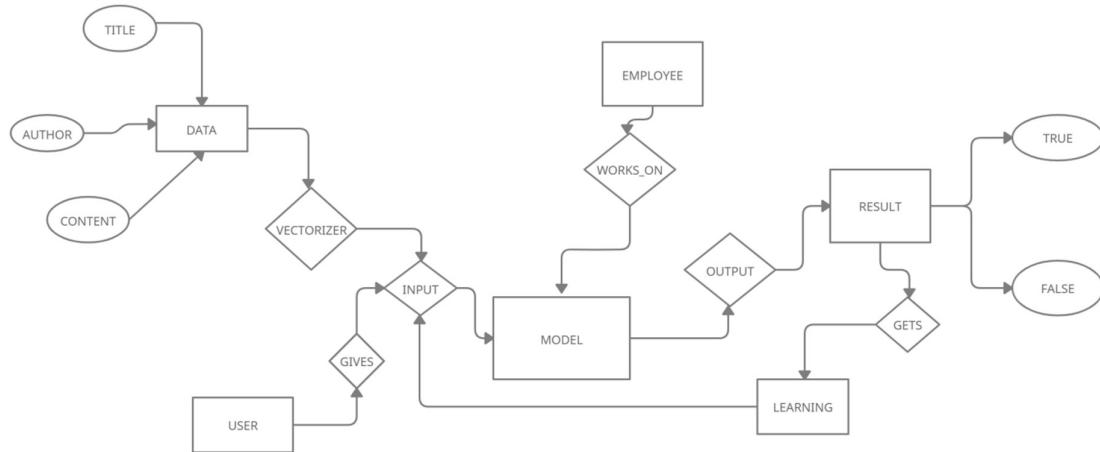
This model is based on three basic concepts: Entities, Attributes, Relationships

ER Diagram – Notations

- Rectangles represent entity sets.
- Diamonds represent relationship sets.
- Lines link attributes to entity sets and entity sets to relationship sets.
- Ellipses represent attributes
- Double ellipses represent multivalued attributes.
- Dashed ellipses denote derived attributes.
- Underline indicates primary key attributes



ER Diagram



ADDITIONAL NOTES

- A database can be modeled as a collection of entities, relationship among entities.
- An entity is an object that exists and is distinguishable from other objects.

Example: specific person, company, event, plant

- Entities have attributes.

Example: people have names and addresses

- An entity set is a set of entities of the same type that share the same properties.

Example: set of all persons, companies, trees, holidays

- Express the number of entities to which another entity can be associated via a relationship set.
- Most useful in describing binary relationship sets.
- We express cardinality constraints by drawing either a directed line (\rightarrow), signifying "one," or an undirected line ($-$), signifying "many," between the relationship set and the entity set.

- An entity is represented by a set of attributes, that is descriptive properties possessed by all members of an entity set.

Example: customer = (customer-id, customer-name, customer-street, customer-city)
loan = (loan-number, amount)

- Domain – the set of permitted values for each attribute

- Attribute types:

1. Simple and composite attributes.
2. Single-valued and multi-valued attributes

E.g. multivalued attribute: phone-numbers

3. Derived attributes-Can be computed from other attributes

E.g. age, given date of birth

Cardinality

- For a binary relationship set the mapping cardinality must be one of the following types:

1. One to one

A customer is associated with at most one loan via the relationship borrower. A loan is associated with at most one customer via borrower

2. One to many

A loan is associated with at most one customer via borrower, a customer is associated with several (including 0) loans via borrower

3. Many to one

A loan is associated with several (including 0) customers via borrower, a customer is associated with at most one loan via borrower

4. Many to many

A loan is associated with several (including 0) customers via borrower, a customer is associated with several loans (including 0) via borrower

Weak Entity Set

- An entity set that does not have a primary key is referred to as a weak entity set and represented by double outlined box in E-R diagram.

Example : Consider the entity set payment which got three attributes : payment_number, payment_date and payment_amount. Payment numbers are sequential starting from 1 generally separately for each loan. Although each payment entity is distinct, payments for different loans may share the same payment number. Thus this entity set does not have a primary key.

Discriminator

- The discriminator (or partial key) of a weak entity set is the set of attributes that distinguishes among all the entities of a weak entity set

Example: discriminator of weak entity set payment is the attribute payment_number since for each loan a payment number uniquely identifies one single payment for that loan.

Specialization-Generalization-ISA

- E-R model provides means of representing these distinctive entity groupings

- Process of designating subgroupings within an entity set is called specialization depicted by triangle component labelled ISA ("is a")

- Bottom up design process in which multiple entity sets are synthesized into higher level entity set - Generalization

- ISA relationship may also be referred to as superclass-subclass relationship

- Higher and lower level entity sets are designated by the terms superclass and subclass.

- Specialization and generalization are simple inversions of each other; they are represented in an E-R diagram in the same way.

Total & Partial Participation

- Total participation (indicated by double line): every entity in the entity set participates in at least one relationship in the relationship set

E.g. participation of loan in borrower is total, every loan must have a customer associated to it via borrower

- Partial participation: some entities may not participate in any relationship in the relationship set

Example: participation of customer in borrower is partial

Cardinality limits

- Cardinality limits can also express participation constraints
- Minimum and maximum cardinality is expressed as l..h where l is the minimum and h is the maximum cardinality
- Minimum value of 1 indicates total participation of entity set in relationship set
- Maximum value of 1 indicates entity participates in atmost one relationship set.
- Maximum value of entity * indicates no limit

Role indicator

- Entity sets of a relationship need not be distinct
- The labels “manager” and “worker” are called roles; they specify how employee entities interact via the works-for relationship set.
- Roles are indicated in E-R diagrams by labeling the lines that connect diamonds to rectangles.
- Role labels are optional, and are used to clarify semantics of the relationship

Disjoint Generalization

- Disjointness constraint requires that an entity belong to more than one lower level entity set.
- Example: account entity can satisfy only one condition for account_type attribute ; entity can either be savings or chequing account but not both.

Result:

Thus, the entity relationship diagram was created successfully.



School of Computing

SRM IST, Kattankulathur – 603 203

Course Code: 18CSC206J

Course Name: Software Engineering and Project Management

Experiment No	8
Title of Experiment	Develop a Data Flow Diagram (Process-Up to Level 1)
Name of the candidate	KUMAR SHASHWAT
Team Members	MOHAMMED SHOAIB KHAN ARYAMAN ADIVYA SINGH
Register Number	RA2011003011262
Date of Experiment	05-30-2022

Mark Split Up

S. No	Description	Maximum Mark	Mark Obtained
1	Exercise	5	

2	Viva	5	
	Total	10	

Staff Signature with date

Aim

To develop the data flow diagram up to level 1 for the <project name>

Team Members:

S No	Register No	Name	Role
1	RA2011003011266	ARYAMAN ADIVYA SINGH	Rep
2	RA2011003011289	MOHAMMED SHOAIB KHAN	Member
3	RA2011003011262	KUMAR SHASHWAT	Member

Data Flow Diagram

The DFD takes an input-process-output view of a system. That is, data objects flow into the software, are transformed by processing elements, and resultant data objects flow out of the software. Data objects are represented by labeled arrows, and transformations are represented by circles (also called bubbles). The DFD is presented in a hierarchical fashion. That is, the first data flow model (sometimes called a level 0 DFD or context diagram) represents the system as a whole. Subsequent data flow diagrams refine the context diagram, providing increasing detail with each subsequent level.

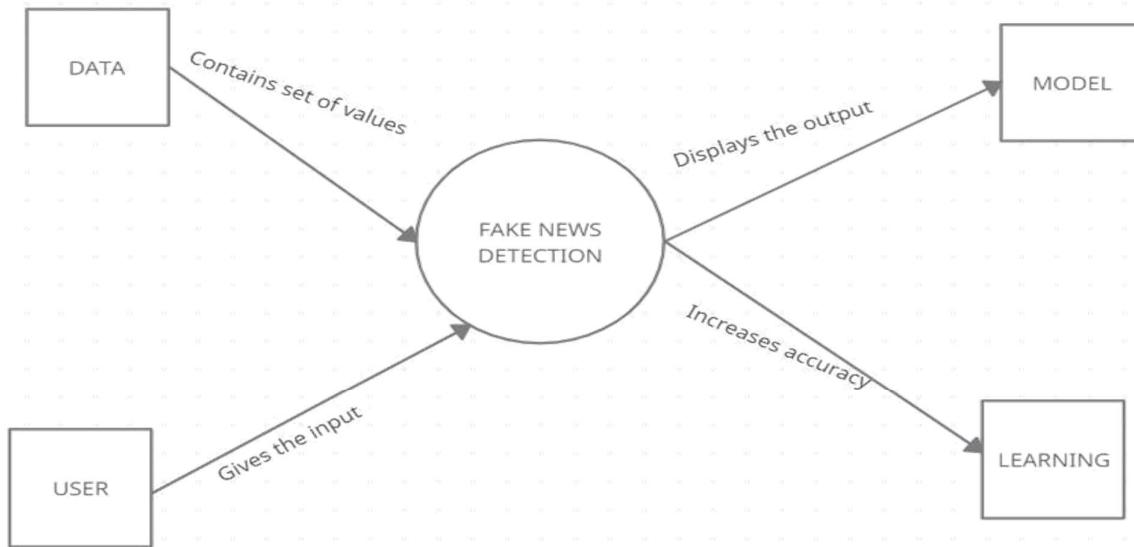
The data flow diagram enables you to develop models of the information domain and functional domain. As the DFD is refined into greater levels of detail, you perform an implicit functional decomposition of the system. At the same time, the DFD refinement results in a corresponding refinement of data as it moves through the processes that embody the application.

A few simple guidelines can aid immeasurably during the derivation of a data flow diagram:

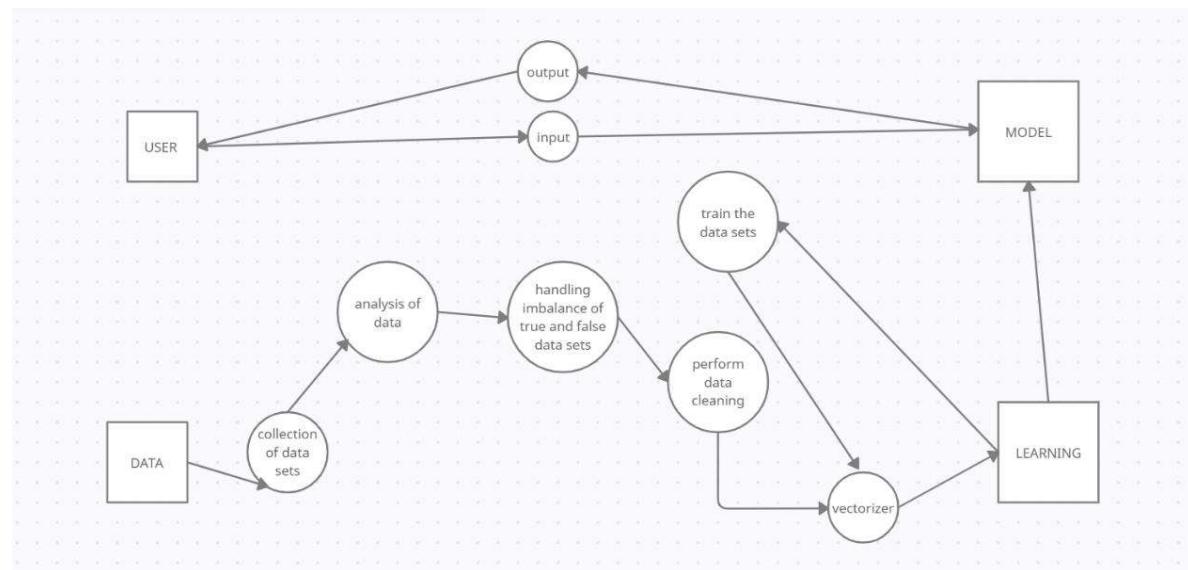
3. Level 0 data flow diagram should depict the software/system as a single bubble;
4. Primary input and output should be carefully noted;
5. Refinement should begin by isolating candidate processes, data objects, and data stores to be represented at the next level;
6. All arrows and bubbles should be labeled with meaningful names;
7. Information flow continuity must be maintained from level to level and
8. One bubble at a time should be refined. There is a natural tendency to overcomplicate the

data flow diagram. This occurs when you attempt to show too much detail too early or represent procedural aspects of the software in lieu of information flow.

DFD Level 0



DFD Level 1



Result:

Thus, the data flow diagrams have been created for the FAKE NEWS DETECTION.



School of Computing

SRM IST, Kattankulathur – 603 203

Course Code: 18CSC206J

Course Name: Software Engineering and Project Management

Experiment No	9
Title of Experiment	Design a Sequence and Collaboration Diagram
Name of the candidate	KUMAR SHASHWAT
Team Members	ARYAMAN ADIVYA SINGH MOHAMMED SHOAIB KHAN
Register Number	RA2011003011262
Date of Experiment	30/05/2022

Mark Split Up

S. No	Description	Maximum Mark	Mark Obtained
1	Exercise	5	
2	Viva	5	
Total		10	

Staff Signature with date

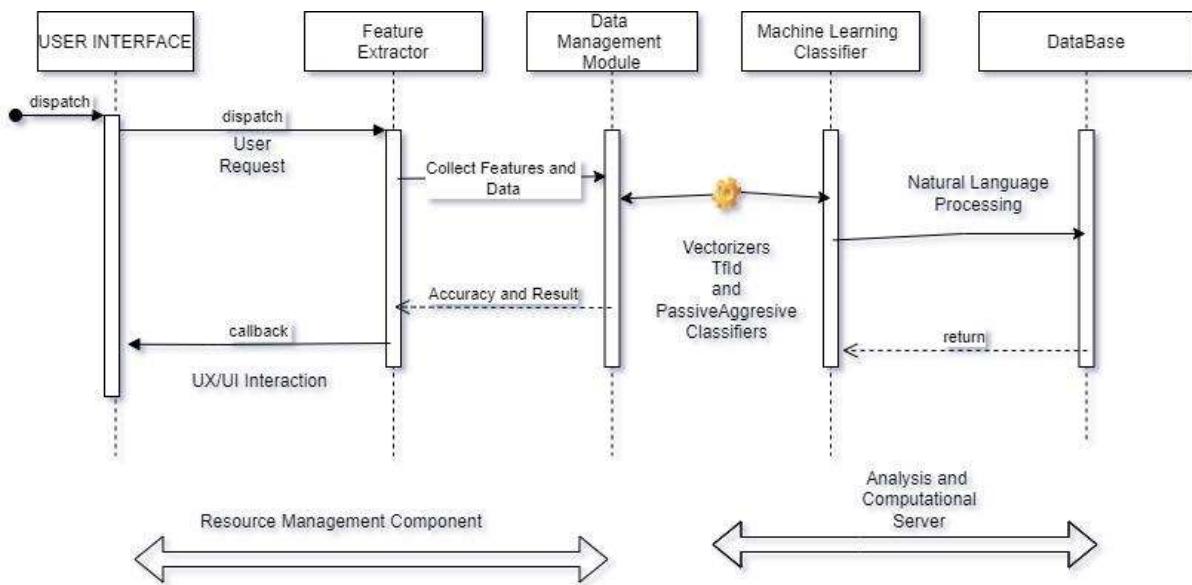
Aim

To create the sequence and collaboration diagram for the FAKE NEWS DETECTION SYSTEM

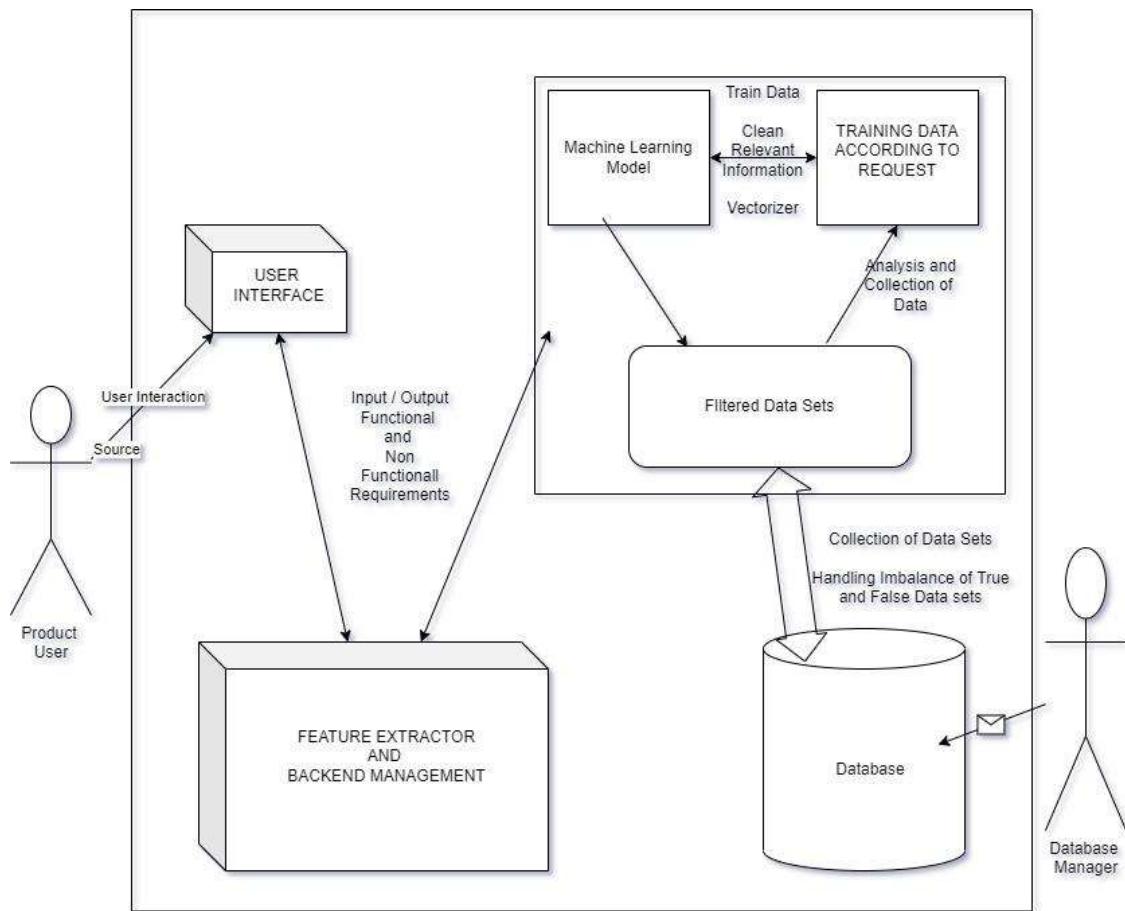
Team Members:

S No	Register No	Name	Role
1	RA2011003011266	ARYAMAN ADIVYA SINGH	Rep/Member
2	RA2011003011289	MOHAMMED SHOAIB KHAN	Member
3	RA2011003011262	KUMAR SHASHWAT	Member

Sequence Diagram



Collaboration Diagram



Result:

Thus, the sequence and collaboration diagrams were created for the FAKE NEWS DETECTION SYSTEM.



School of Computing

SRM IST, Kattankulathur – 603 203

Course Code: 18CSC206J

Course Name: Software Engineering and Project Management

Experiment No	10
Title of Experiment	Develop a Testing Framework/User Interface
Name of the candidate	KUMAR SHASHWAT
Team Members	MOHAMMED SHOAIB KHAN ARYAMAN ADIVYA SINGH
Register Number	RA2011003011262
Date of Experiment	07-06-2022

Mark Split Up

S. No	Description	Maximum Mark	Mark Obtained
1	Exercise	5	
2	Viva	5	
Total		10	

Staff Signature with date

Aim

To develop the testing framework and/or user interface framework for the Fake News Detection.

Team Members:

S No	Register No	Name	Role
1	RA2011003011266	ARYAMAN ADIVYA SINGH	Rep/Member
2	RA2011003011262	KUMAR SHASHWAT	Member
3	RA2011003011289	MOHAMMED SHOAIB KHAN	Member

Test Plan

Scope of Testing

As every aspect of the application is to be tested, we first decide on the types of testing to be performed. When a certain type of testing is to be performed, test cases are generated for that type of testing. And the generated test cases are performed on the application and based on the test report the machine learning application will be further improved.

Types of Testing, Methodology, Tools

Testing Framework

A Test Automation Framework is a set of guidelines like coding standards, test-data handling, object repository treatment etc... which when followed during automation scripting produces beneficial outcomes like increased code re-usage, higher portability, reduced script maintenance cost etc. These are just guidelines and not rules; they are not mandatory and you can still script without following the guidelines. But you will miss out on the advantages of having a Framework.

Creating a test script that creates a placeholder to accommodate the data (content that's normally present in a file upload) from the test file, and then having different values called in this test script will execute this test framework.

Machine learning in this model will be seamlessly implemented using Data Driven Testing with Agile Methodology.

Data-Driven Testing Framework

This framework design uses an execution strategy where the test scripts read test data or trained data(as in the case of MACHINE LEARNING) from data sources or databases such as ADO objects, CSV files,etc rather than hard-coded values. Test scripts are prepared either using Linear Scripting,Web Scraping or Test Library Framework.

Testing Methodology

Software testing methodologies are the various strategies or approaches used to test an application to ensure it behaves and looks as expected.

The goal of utilizing numerous testing methodologies in your development process is to make sure your software can successfully operate in multiple environments and across different platforms. These can typically be broken down between functional and non-functional testing. Functional testing involves testing the application against the business requirements. It incorporates all test types designed to guarantee each part of a piece of software behaves as expected by using use cases provided by the design team or business analyst. These testing methods are usually conducted in order and include:

Functional Requirements Testing :

- Unit Testing
- Integration Testing
- System Testing
- Acceptance Testing

Non-Functional Requirements Testing :

- Load Testing
- Performance Testing

- Security Testing
- Stress Testing

Testing Toolkit

Software testing tools are used by software developers, testers, and quality assurance (QA) teams to ensure that software products perform as intended and look as expected. Testing tools are used throughout the software development lifecycle (SDLC) to test everything from individual units of code to entire software applications. When using effective software testing tools, companies can ensure that the software they are developing meets specifications and requirements, provides a superior user experience, and is relatively free of flaws and defects.

The Project “FAKE NEWS DETECTION USING MACHINE LEARNING” is developed using the Proof Of Concept or Natural Language Processing Toolkit.

Natural Language Processing

Natural language processing (NLP) is a subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data.

The Main Tools and Libraries that the model will be utilized, impeding the absence of technologies otherwise in normal Machine Learning Models are :

1. NLTK - entry-level open-source NLP Tool

Natural Language Toolkit (AKA NLTK) is an open-source software powered with Python NLP. From this point, the NLTK library is a standard NLP tool developed for research and education.

NLTK provides users with a basic set of tools for text-related operations. It is a good starting point for beginners in Natural Language Processing.

Natural Language Toolkit features include:

- Text classification
- Part-of-speech tagging
- Entity extraction
- Tokenization
- Parsing
- Stemming
- Semantic reasoning

NLTK interface includes text corpora and lexical resources.

They include:

- Penn Treebank Corpus
- Open Multilingual Wordnet
- Problem Report Corpus
- Lin’s Dependency Thesaurus

2. Stanford Core NLP - Data Analysis, Sentiment Analysis, Conversational UI

We can say that the Stanford NLP library is a multi-purpose tool for text analysis. Like NLTK, Stanford CoreNLP provides many different natural language processing software. But if you need more, you can use custom modules.

The main advantage of Stanford NLP tools is scalability. Unlike NLTK, Stanford Core NLP is a perfect choice for processing large amounts of data and performing complex operations.

With its high scalability, Stanford CoreNLP is an excellent choice for:

- information scraping from open sources (social media, user-generated reviews)
- sentiment analysis (social media, customer support)
- conversational interfaces(chatbots)
- text processing, and generation(customer support)

Result:

Thus, the testing framework/user interface framework has been created for the Fake News Detection.



School of Computing

SRM IST, Kattankulathur – 603 203

Course Code: 18CSC206J

Course Name: Software Engineering and Project Management

Experiment No	11
Title of Experiment	Test Cases
Name of the candidate	KUMAR SHASHWAT
Team Members	MOHAMMED SHOAIB KHAN ARYAMAN ADIVYA SINGH
Register Number	RA2011003011262
Date of Experiment	06/06/2022

Mark Split Up

S. No	Description	Maximum Mark	Mark Obtained
1	Exercise	5	
2	Viva	5	
Total		10	

Staff Signature with date

Aim

To develop the test cases manual for the Fake News Detection.

Team Members:

S No	Register No	Name	Role
1	RA2011003011266	ARYAMAN ADIVYA SINGH	Rep
2	RA2011003011289	MOHAMMED SHOAIB KHAN	Member
3	RA2011003011262	KUMAR SHASHWAT	Member

Test Cases

Functional Test Cases

Test ID (#)	Test Scenario	Test Case	Execution Steps	Expected Outcome	Actual Outcome	Status	Remarks
TC-01	Verify User Registration	Accept Valid Username	<ul style="list-style-type: none"> User click on sign up User enter credentials Click sign up 	User should view a message "Registration successful"	User should view a message "Registration successful"	Pass	Sign up successful
TC-02	Verify User Registration	Don't Accept Username	<ul style="list-style-type: none"> User click on sign up User enter credentials Click sign up 	User should view a message "Registration successful"	User should view a message "User name already taken"	Fail	Invalid username
TC-03	Verify Login Credentials	Accept login credentials	<ul style="list-style-type: none"> User click login User enter username and password Click login 	User can see his username at top right corner	User can see his username at top right corner	Pass	Login successful
TC-04	Verify login Credentials	Reject login credentials	<ul style="list-style-type: none"> User click login User enter username and password Click login 	User can see his username at top right corner	User sees a message "invalid username or password"	Fail	Invalid Username or Password
TC-05	Verify Relevance of article	Accept input and show result	<ul style="list-style-type: none"> User will either type or paste article User click on predict 	User can see Real/Fake as a result	User can see Real/Fake as a result	Pass	Model Predicted successfully
TC-06	Verify relevance of article	Reject Input	<ul style="list-style-type: none"> User will either type or paste article User click on predict 	User can see Real/Fake as a result	User will see Invalid Input	Fail	Input is either blank or just numbers

							or symbols
TC-07	Verify Search option	Accept Input shows result	<ul style="list-style-type: none"> ● User will type the title of a news ● User will click search 	User can see “News is present is database and it is Real/Fake”	User can see “News is present is database and it is Real/Fake”	Pass	Search successful
TC-08	Verify Search option	Accept Input and doesn't show result	<ol style="list-style-type: none"> 1. User will type the tile of a news 2. User will click search 	User can see “News is present is database and it is Real/Fake”	User will see the article is not present in the database	Fail	Search Unsuccessful

Non-Functional Test Cases

Test ID (#)	Test Scenario	Test Case	Execution Steps	Expected Outcome	Actual Outcome	Status	Remarks
TC-9	Load Testing	Request is sent successfully and data is received	<ol style="list-style-type: none"> 1. The browser will send DNS request to the DNS Server 2. Server replies with the hostname to the DNS server 3. The browser will send HTTPS request 4. The server responds 5. The data is send to the user 	Website will load under 3 seconds	Website will load under 3 seconds	Success	Website loaded successfully
TC-10	Load Testing	Request failed to sent	<ol style="list-style-type: none"> 1. The browser will send the request to the server 2. The server doesn't respond 	Website will load within 3 seconds	The user sees the prompt "Site can't be reached"		
TC-11	Performance Testing	Model predicts results within 5 seconds	<ol style="list-style-type: none"> 1. The User enters the article and press predict 	The model will display Real/Fake under 5 seconds	The model will display Real/Fake under 5 seconds	Success	Model predicted successfully

TC-12	Security Testing	Successfully authenticate user based on user credentials	<ol style="list-style-type: none"> 1. The user will enter the login credentials 2. The system will encrypt the data and send it to the server 3. The server decipts the data and authentic ate the data 4. The server sends the data back to the user 	Login verified and user can access the website	Login verified and user can access the website	Success	System is secure
TC-13	Stress testing	Website should work even for thousand users at a time	<ol style="list-style-type: none"> 1. Access the website from more than thousand different systems 2. get response from each user if the website is working from each user 3. Verify the results 	Website should work for more than thousand users	Website should work for more than thousand users	Success	System can handle at least a thousand users

Result:

Thus, the test case manual has been created for the <project name>.



School of Computing

SRM IST, Kattankulathur – 603 203

Course Code: 18CSC206J

Course Name: Software Engineering and Project Management

Experiment No	12
Title of Experiment	Manual Test Case Reporting
Name of the candidate	KUMAR SHASHWAT
Team Members	ARYAMAN ADIVYA SINGH MOHAMMED SHOAIB KHAN
Register Number	RA2011003011262
Date of Experiment	06/06/2022

Mark Split Up

S. No	Description	Maximum Mark	Mark Obtained
1	Exercise	5	
2	Viva	5	
Total		10	

Staff Signature with date

Aim

To prepare the manual test case report for the Fake News Detection

Team Members:

S No	Register No	Name	Role
1	RA2011003011266	ARYAMAN ADIVYA SINGH	Rep/Member
2	RA2011003011289	MOHAMMED SHOAIB KHAN	Member
3	RA2011003011262	KUMAR SHASHWAT	Member

Test ID (#)	Test Scenario	Test Case	Execution Steps	Expected Outcome	Actual Outcome	Status	Remarks	Priority
TC-01	Verify User Registration	Accept Valid Username	1. User click on sign up 2. User enter credentials 3. Click sign up	User should view a message "Registration successful"	User should view a message "Registration successful"	Pass	Sign up successful	Low
TC-02	Verify User Registration	Don't Accept Username	1. User click on sign up 2. User enter credentials 3. Click sign up	User should view a message "Registration successful"	User should view a message "Username already taken"	Fail	Invalid username	Low
TC-03	Verify Login Credentials	Accept login credentials	1. User click login 2. User enter username and password 3. Click login	User can see his username at top right corner	User can see his username at top right corner	Pass	Login successful	Low

TC-04	Verify login credentials	Reject login credentials	4. User click login 5. User enter username and password 6. Click login	User can see his username at top right corner	User sees a message "invalid username or password"	Fail	Invalid Username or Password	Low
TC-05	Verify Relevance of article	Accept input and show result	1. User will either type or paste article 2. User click on predict	User can see Real/Fake as a result	User can see Real/Fake as a result	Pass	Model Predicted successfully	Medium
TC-06	Verify relevance of article	Reject Input	3. User will either type or paste article 4. User click on predict	User can see Real/Fake as a result	User will see Invalid Input	Fail	Input is either blank or just numbers or symbols	Medium
TC-07	Verify Search option	Accept Input shows result	1. User will type the title of a news 2. User will click search	User can see "News is present in database and it is Real/Fake"	User can see "News is present in database and it is Real/Fake"	Pass	Search successful	Low
TC-08	Verify Search option	Accept Input and doesn't show result	1. User will type the tile of a news 2. User will click search	User can see "News is present in database and it is Real/Fake"	User will see the article is not present in the database	Fail	Search Unsuccessful	Low

Test ID (#)	Test Scenario	Test Case	Execution Steps	Expected Outcome	Actual Outcome	Status	Remarks	Priority
TC-9	Load Testing	Request is sent successfully and data is received	<ol style="list-style-type: none"> The browser will send DNS request to the DNS Server Server replies with the hostname to the DNS server The browser will send HTTPS request The server responds The data is send to the user 	Website will load under 3 seconds	Website will load under 3 seconds	Success	Website loaded successfully	High
TC-10	Load Testing	Request failed to sent	<ol style="list-style-type: none"> The browser will send the request to the server The server doesn't respond 	Website will load within 3 seconds	The user sees the prompt "Site can't be reached"			High
TC-11	Performance Testing	Model predicts results within 5 seconds	<ol style="list-style-type: none"> The User enters the article and press predict 	The model will display Real/Fake under 5 seconds	The model will display Real/Fake under 5 seconds	Success	Model predicted successfully	High
TC-12	Security Testing	Succesfully	<ol style="list-style-type: none"> The user will enter 	Login verified and user can	Login verified and user can	Success	System is secure	High

		authenticate user based on user credentials	the login credentials 2. The system will encrypt the data and send it to the server 3. The server decipts the data and authentic ate the data 4. The server sends the data back to the user	access the website	access the website			
TC-13	Stress testing	Website should work even for thousand users at a time	1. Access the website from more than thousand different systems 2. get response from each user if the website is working from each user 3. Verify the results	Website should work for more than thousand users	Website should work for more than thousand users	Success	System can handle at least a thousand users	High

SUMMARY OF CURRENT STATUS OF TESTING

The Testing and Model Formation has been succeeded and the outputs are obtained with an accuracy of 97%.

OBSTACLES TO PROCEED FURTHER

Category	Progress Against Plan	Status
Functional Testing	Green	Completed
Non-Functional Testing	Green	Completed

Result:

Thus, the test case report has been created for the Fake News Detection



School of Computing

SRM IST, Kattankulathur – 603 203

Course Code: 18CSC206J

Course Name: Software Engineering and Project Management

Experiment No	13
Title of Experiment	Provide the details of Architecture Design/Framework/Implementation
Name of the candidate	KUMAR SHASHWAT
Team Members	MOHAMMED SHOAIB KHAN ARYAMAN ADIVYA SINGH
Register Numbers	RA2011003011262
Date of Experiment	06/06/2022

Mark Split Up

S. No	Description	Maximum Mark	Mark Obtained
1	Exercise	5	
2	Viva	5	
	Total	10	

Staff Signature with date

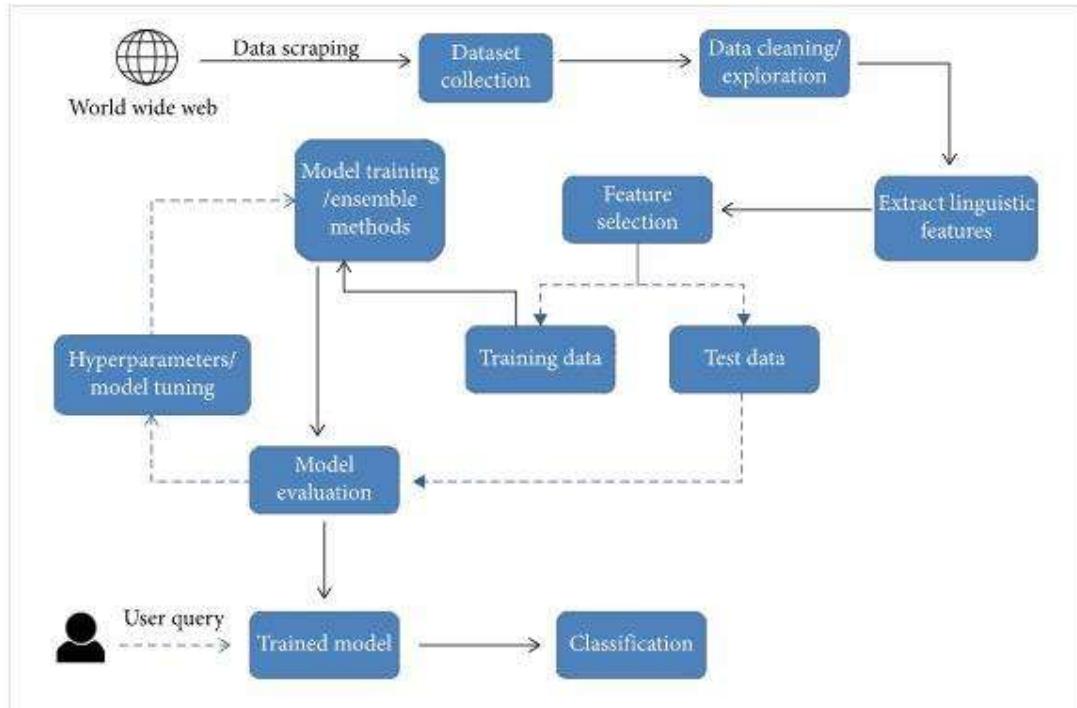
Aim

To provide the details of architectural design/framework/implementation

Team Members:

S No	Register No	Name	Role
1	RA2011003011266	ARYAMAN ADIVYA SINGH	Rep/Member
2	RA2011003011289	MOHAMMED SHOAIB KHAN	Member
3	RA2011003011262	KUMAR SHASHWAT	Member

Module Architecture



INTRODUCTION

PROBLEM

Fake news detection and analysis

Social media is a dominant part of our lives right now, hence, it has become more critical than ever to distinguish fake news from real news. Especially since fake news spreads like wildfire and often always has harmful effects. Fake news has created communal unrest, spread hate and fear and given false information to people that could potentially be fatal and hence needs to be identified and removed quickly.

JUSTIFICATION

It is imperative that a relatively accurate model is built to detect fake news and prevent it from influencing anyone. We need to identify the untrustworthy sources and common phrases and words used in typical fake news posts on social media. We plan to do that by combining different datasets of fake and real news and creating a model using Logistic Regression to classify the piece of information as fake or real news.

ABSTRACT

PURPOSE

Social media is a dominant part of our lives right now, hence, it has become more critical than ever to distinguish fake news from real news. The project aims at detecting fake news from social media platforms. We need to identify the untrustworthy sources and common phrases and words used in typical fake news posts on social media.

METHODS USED

We need to identify the untrustworthy sources and common phrases and words used in typical fake news posts on social media. We plan to do that by combining different datasets of fake and real news and creating a model using logistic regression to classify the piece of information as fake or real news. We plan to build a model that can be used on real-time data and can immediately classify it. The next step was to determine the most optimal parameters for the algorithm and compare the results based on those parameters. From this we can choose those parameters which give

the highest accuracy in detecting fake news.

DATA/OBSERVATIONS

We are going to collect data from different social media. This data will mainly be any post that includes news related to any topic. Data for fake news and real news will be assembled separately. Datasets can be collected from Kaggle as it has released a fake news and real news dataset consisting of 13,000 articles in 2016 as part of its fake news detection challenge.

SELECTION OF METHODOLOGY

Agile is a change of rapid development and deployment, meaning that starting with the basics of software programming that is planning. Now, since we know what we are about to build **Fake News Detection application**, we breakdown all of the activity or process into *small chunks of services* and then we work on those small services, one at a time which ensures that we follow the *micro services model* and at the same time we don't really affect the entire application in general.

So, we plan, we design, we develop or architect, we test, we deploy and then we review it. We notice that launch is actually outside the entire circle, meaning every time we make a change, (*it could be as simple as a one line of code, say changing a variable name*) no matter how big or small that change is, it has to be first deployed in a **DevOps** environment to get a constant feedback over what's happening with the code that is updated or added.

Benefit of working with agile is that we are able to *work in iterations or sprint*, in which we need not worry about the whole application all at once; rather we can focus on micro services so that work gets faster and easier. It's basically like creating a multiple function in any programming language rather than writing all big codes in the main function itself. Like even if all micro services are in synergy yet they are completely different tasks.

System Requirements

Hardware Requirements:

Minimum Requirements

CPU: Single Core 2.4 GHz
RAM: 512 MB
Graphics Card: Nvidia GeForce 5xxx series or equivalent
Operating System: Windows XP
Hard Drive: 5 Gigabytes
Network: Broadband Recommended

Recommended Requirements

CPU: Duo Core or higher
RAM: 4 GB
Graphics Card: Nvidia GeForce 7xxx series or equivalent
Operating System: Windows 7 or higher
Hard Drive: 6 Gigabytes or more
Network: Broadband Recommended

Operating System Requirements

Windows

To use the Application on the browser of Windows, you'll need:

- Windows 7, Windows 8, Windows 8.1, Windows 10 or later
- An Intel Pentium 4 processor or later that's SSE3 capable

Note: Servers require Windows Server 2008 R2, Windows Server 2012, Windows Server 2012 R2, or Windows Server 2016.

Mac

To use the Application on the browser of Mac, you'll need:

- OS X El Capitan 10.11 or later

Linux

To use the Application on the browser of Linux, you'll need:

- 64-bit Ubuntu 18.04+, Debian 10+, openSUSE 15.2+, or Fedora Linux 32+
- An Intel Pentium 4 processor or later that's SSE3 capable

Android

To use Web Application on the browser of Android, you'll need:

- Android Marshmallow 6.0 or later

Web Browser Requirements

Web / Mobile Browser	Infrastructure Classic Console and Applications Console
Microsoft Internet Explorer	11 or later
Mozilla Firefox	52 or later
Google Chrome	63 or later
Apple Safari	10 or later
Microsoft Edge	35 or later
Safari, Chrome, Firefox on iOS (iPad and iPhone)	Latest
Chrome, Firefox on Android (Phone and Tablet)	Latest

Functional Requirements

Login:

- The system should provide a login page for the user to log into the system.
- If login is successful then the user can access the system. • else it would display "Login Credentials Incorrect".

Search:

- The system should display a search option where users can search for news.
- If the news is already in the database then it will provide the article and also display if the news is fake or real.

Upload:

- If the news is not present in the database then, the user will get the option to upload the news title, author, and article.
- The system will then provide the data to the machine learning model which then displays if the news is real or fake.

Update:

- After the processing of the news, the system will add the news article to its database so that in the future no need to process the article again.

Verification:

- The system will provide a feedback system to users which will allow the user to tell if the news is real or fake and matches the result with the result provided by the algorithm.
- If the result is not correct then the system will update the result and its accuracy score.

Logout:

- The system will provide a logout option to log out of the system.

Non-Functional Requirements

Performance and scalability.

The application returns the result instantly provided it meets the user system requirement. There won't be any changes in the performance or response time with the higher workload.

Portability and compatibility.

Minimum hardware devices are iPhone 6 and above, Android devices that support Android Kitkat and above including android one. Browsers with which this application will be compatible are Internet Explorer 10 and above. Google Chrome V19.0.1084 and above. Safari 5.1.10 and above. Firefox 10 and above. The Operating System for this application to run is Windows. It is an application that runs independently without conflicting with other processes.

Reliability, availability, maintainability.

Weekly maintenance is required for this application, to segregate the news that has been fetched by the user for it to be organized on the basis of its accuracy score, also modularize based upon the user's feedback, and also update the firewall and security options to prevent any cyber-attacks.

We would try to lessen the downtime as per the maintenance requirement but it will certainly not be anything too long.

Security.

All communication should be using HTTPS over TLS 1.2. All APIs will have auth token (JWT token with SHA256 algorithm). All Personal Identifiable data would be always encrypted. Files on servers would be encrypted. All documents generated should be password protected, display an error message in case of the wrong password with the date it was set, and a hint of standard documents – like '*this is the same password as all your other documents*'. Do not add

sensitive data openly to request payload. Use OAuth 2.0 for authentication using popular IDPs such as Google, MS ADFS, etc.

Usability.

UX is crisp and consistent across all major devices, standard and non-standard accepted viewports and resolutions are a standard requirement. The navigation is smooth without halts or glitches and follows standard navigation guidelines:

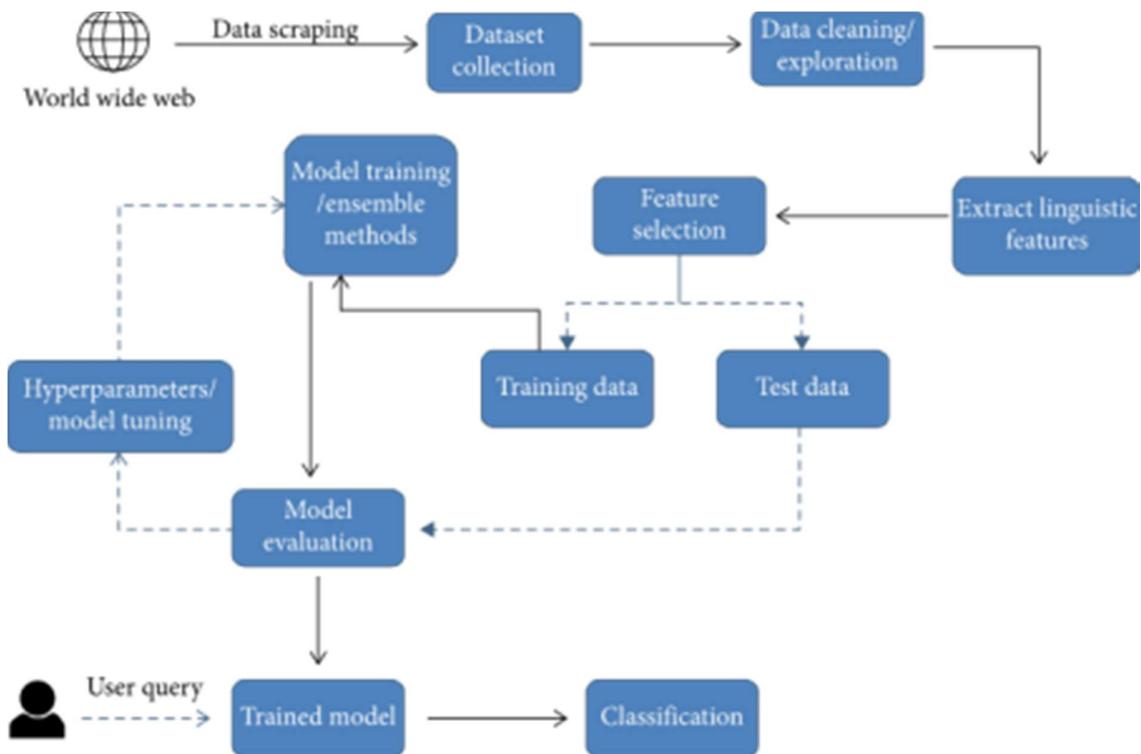
- Modular UI architecture

- Load data fast (<2 seconds)
 - Use as little data as possible
 - Use static assets from a local cache
 - Separate content from navigation
 - Retrieve and display page-specific content (HTML, JSON, etc.) • Optionally, cache dynamic content

PROJECT PLANNING

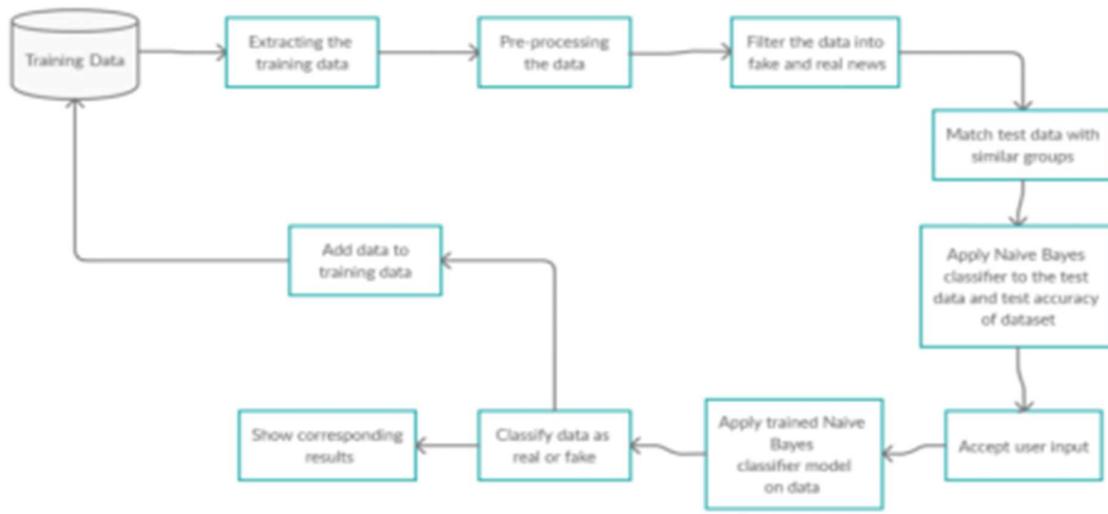
SYSTEM ARCHITECTURE

Fundamental Structure

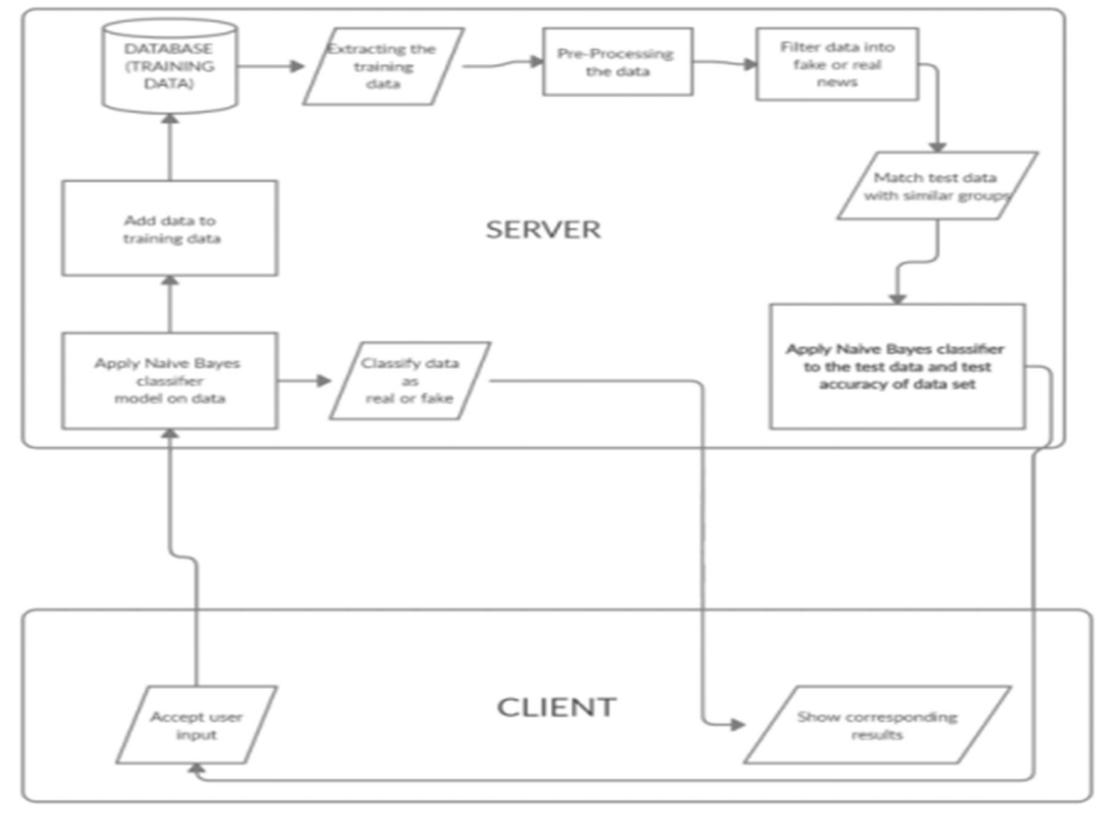


EXHAUSTIVE STRUCTURE DIAGRAM

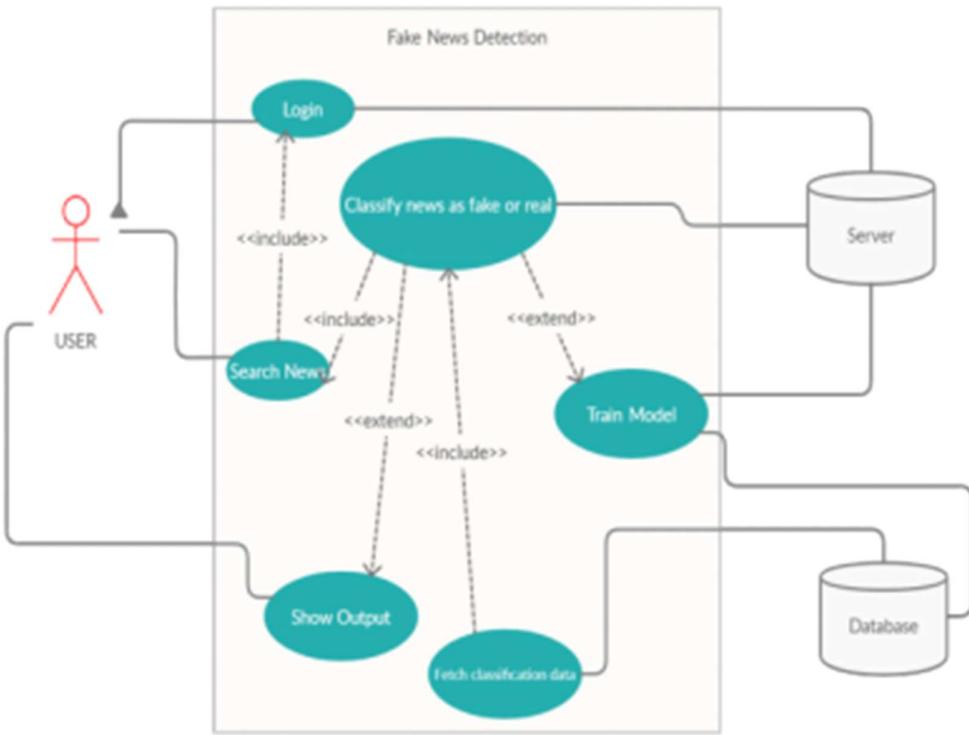
1. PIPE AND FILTER PATTERN:



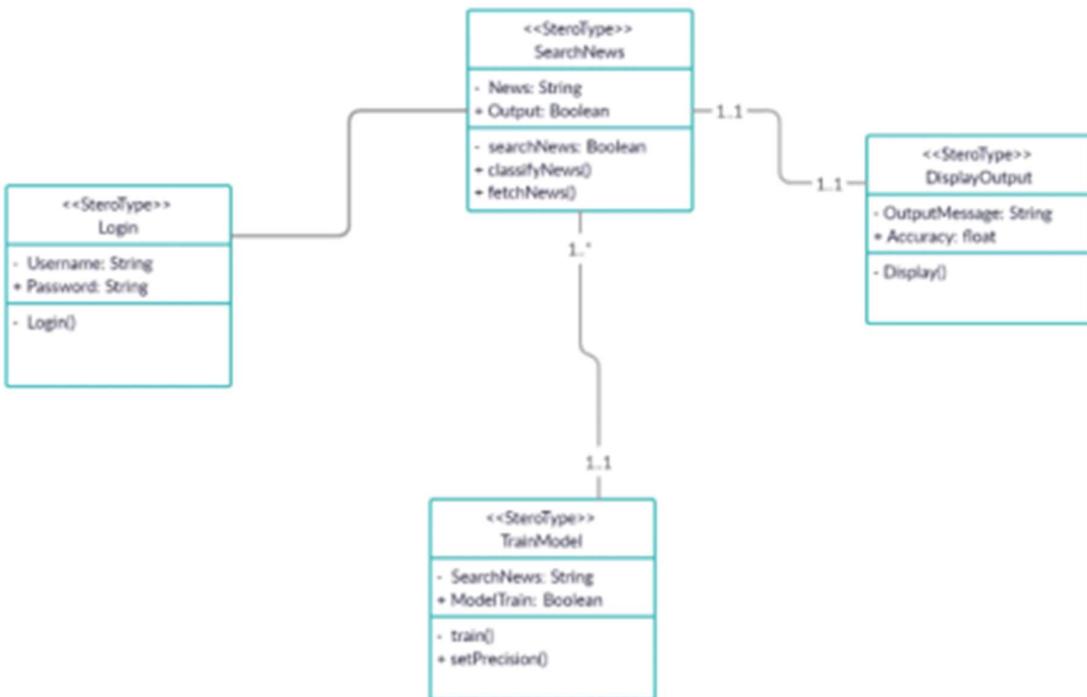
2. HYBRID MODEL WITH CLIENT-SERVER AND PIPE FILTER PATTERN:



3.SYSTEM USE CASE DIAGRAM



CLASS DIAGRAM



CONSTRUCTION

The code has been implemented using Google Colab

In [1]:

```
import pandas as pd
import numpy as np
import itertools
```

In [2]:

```
from google.colab import drive
```

In [3]:

```
drive.mount('/content/drive')
```

Mounted at /content/drive

In [4]:

```
#Data set with only id and label
data = pd.read_csv('/content/drive/MyDrive/FNdata/news.csv')
```

In [5]:

```
data.head(5)
```

Out[5]:

	Unnamed: 0	title	text	label
0	8476	You Can Smell Hillary's Fear	Daniel Greenfield, a Shillman Journalism Fello...	FAKE
1	10294	Watch The Exact Moment Paul Ryan Committed Pol...	Google Pinterest Digg Linkedin Reddit Stumble...	FAKE
2	3608	Kerry to go to Paris in gesture of sympathy	U.S. Secretary of State John F. Kerry said Mon...	REAL
3	10142	Bernie supporters on Twitter erupt in anger ag...	— Kaydee King (@KaydeeKing) November 9, 2016 T...	FAKE
4	875	The Battle of New York: Why This Primary Matters	It's primary day in New York and front-runners...	REAL

In [6]:

```
data.shape
#check data rows and columns
```

Out[6]:

(6335, 4)

In [7]:

```
#Check the blanks and null spaces in data set
data.isnull().sum()
```

Out[7]:

```
Unnamed: 0      0
title          0
text           0
label          0
dtype: int64
```

In [8]:

```
#To categorize labels
labels = data.label
labels
```

Out[8]:

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

In [9]:

```
#Kaggle Data Set
data2 = pd.read_csv('/content/drive/MyDrive/FNdata/train.csv')
```

In [10]:

```
data.columns = ['id', 'title', 'text','label']
data.head(5)
```

Out[10]:

	id	title	text	label
0	8476	You Can Smell Hillary's Fear	Daniel Greenfield, a Shillman Journalism Fello...	FAKE

	id	title	text	label
1	10294	Watch The Exact Moment Paul Ryan Committed Pol...	Google Pinterest Digg LinkedIn Reddit Stumble...	FAKE
2	3608	Kerry to go to Paris in gesture of sympathy	U.S. Secretary of State John F. Kerry said Mon...	REAL
3	10142	Bernie supporters on Twitter erupt in anger ag...	— Kaydee King (@KaydeeKing) November 9, 2016 T...	FAKE
4	875	The Battle of New York: Why This Primary Matters	It's primary day in New York and front-runners...	REAL

In [11]:

data2.head(5)

Out[11]:

	id	title	author	text	label
0	0	House Dem Aide: We Didn't Even See Comey's Let...	Darrell Lucas	House Dem Aide: We Didn't Even See Comey's Let...	1
1	1	FLYNN: Hillary Clinton, Big Woman on Campus - ...	Daniel J. Flynn	Ever get the feeling your life circles the rou...	0
2	2	Why the Truth Might Get You Fired	Consortiumnews.com	Why the Truth Might Get You Fired October 29, ...	1
3	3	15 Civilians Killed In Single US Airstrike Hav...	Jessica Purkiss	Videos 15 Civilians Killed In Single US Airstr...	1

id	title	author	text label
4	Iranian woman jailed for fictional unpublished...	Howard Portnoy	Print \nAn Iranian woman has been sentenced to...

In [12]:

data2.pop('author')

Out[12]:

0		Darrell
Lucus		
1		Daniel J.
Flynn		
2		Consortiumne
ws.com		
3		Jessica P
urkiss		
4		Howard P
ortnoy		
	...	
20795		Jerome
Hudson		
20796		Benjamin H
offman		
20797	Michael J. de la Merced and Rachel	
Abrams		
20798		Alex
Ansary		
20799		David S
wanson		
Name: author, Length: 20800, dtype: object		

In [13]:

data2.head(5)

Out[13]:

id	title	text label
----	-------	------------

	id	title	text	label
0	0	House Dem Aide: We Didn't Even See Comey's Let...	House Dem Aide: We Didn't Even See Comey's Let...	1
1	1	FLYNN: Hillary Clinton, Big Woman on Campus - ...	Ever get the feeling your life circles the rou...	0
2	2	Why the Truth Might Get You Fired	Why the Truth Might Get You Fired October 29, ...	1
3	3	15 Civilians Killed In Single US Airstrike Hav...	Videos 15 Civilians Killed In Single US Airstri...	1
4	4	Iranian woman jailed for fictional unpublished...	Print \nAn Iranian woman has been sentenced to...	1

In [14]:

display(data,data2)

	id	title	text	label
0	8476	You Can Smell Hillary's Fear	Daniel Greenfield, a Shillman Journalism Fellow...	FAKE
1	10294	Watch The Exact Moment Paul Ryan Committed Pol...	Google Pinterest Digg LinkedIn Reddit Stumble...	FAKE
2	3608	Kerry to go to Paris in gesture of sympathy	U.S. Secretary of State John F. Kerry said Mon...	REAL
3	10142	Bernie supporters on Twitter erupt in anger ag...	— Kaydee King (@KaydeeKing) November 9, 2016 T...	FAKE
4	875	The Battle of New York: Why This Primary Matters	It's primary day in New York and front-runners...	REAL
...
6330	4490	State Department says it can't find emails fro...	The State Department told the Republican Natio...	REAL
6331	8062	The 'P' in PBS Should Stand for 'Plutocratic' ...	The 'P' in PBS Should Stand for 'Plutocratic' ...	FAKE

	id	title	text	label
6332	8622	Anti-Trump Protesters Are Tools of the Oligarc...	Anti-Trump Protesters Are Tools of the Oligar...	FAKE
6333	4021	In Ethiopia, Obama seeks progress on peace, se...	ADDIS ABABA, Ethiopia —President Obama convene...	REAL
6334	4330	Jeb Bush Is Suddenly Attacking Trump. Here's W...	Jeb Bush Is Suddenly Attacking Trump. Here's W...	REAL

6335 rows × 4 columns

	id	title	text	label
0	0	House Dem Aide: We Didn't Even See Comey's Let...	House Dem Aide: We Didn't Even See Comey's Let...	1
1	1	FLYNN: Hillary Clinton, Big Woman on Campus - ...	Ever get the feeling your life circles the rou...	0
2	2	Why the Truth Might Get You Fired	Why the Truth Might Get You Fired October 29, ...	1
3	3	15 Civilians Killed In Single US Airstrike Hav...	Videos 15 Civilians Killed In Single US Airstri...	1
4	4	Iranian woman jailed for fictional unpublished...	Print \nAn Iranian woman has been sentenced to...	1
...
20795	20795	Rapper T.I.: Trump a 'Poster Child For White S...	Rapper T. I. unloaded on black celebrities who...	0
20796	20796	N.F.L. Playoffs: Schedule, Matchups and Odds - ...	When the Green Bay Packers lost to the Washing...	0
20797	20797	Macy's Is Said to Receive Takeover Approach by...	The Macy's of today grew from the union of sev...	0
20798	20798	NATO, Russia To Hold Parallel Exercises In Bal...	NATO, Russia To Hold Parallel Exercises In Bal...	1
20799	20799	What Keeps the F-35 Alive	David Swanson is an author, activist, journa...	1

20800 rows × 4 columns

In [15]:

```
#Converting values in 1 (REAL -> 1 , FAKE -> 0)
data['label'] = data['label'].replace(['REAL','FAKE'],[1,0])
```

In [16]:

```
display(data,data2)
```

	<u>id</u>	<u>title</u>	<u>text</u>	<u>label</u>
0	8476	You Can Smell Hillary's Fear	Daniel Greenfield, a Shillman Journalism Fellow...	0
1	10294	Watch The Exact Moment Paul Ryan Committed Pol...	Google Pinterest Digg LinkedIn Reddit Stumble...	0
2	3608	Kerry to go to Paris in gesture of sympathy	U.S. Secretary of State John F. Kerry said Mon...	1
3	10142	Twitter erupt in anger ag...	— Kaydee King (@KaydeeKing) November 9, 2016 T...	0
4	875	The Battle of New York: Why This Primary Matters	It's primary day in New York and front-runners...	1
...
6330	4490	State Department says it can't find emails fro...	The State Department told the Republican Natio...	1
6331	8062	Stand for 'Plutocratic' ...	The 'P' in PBS Should Stand for 'Plutocratic' ...	0
6332	8622	Anti-Trump Protesters Are Tools of the Oligarc...	Anti-Trump Protesters Are Tools of the Oligar...	0
6333	4021	In Ethiopia, Obama seeks progress on peace, se...	ADDIS ABABA, Ethiopia — President Obama convene...	1
6334	4330	Attacking Trump. Here's W...	Jeb Bush Is Suddenly Attacking Trump. Here's W...	1

6335 rows × 4 columns

	<u>id</u>	<u>title</u>	<u>text</u>	<u>label</u>

	id	title	text	label
0	0	House Dem Aide: We Didn't Even See Comey's Let...	House Dem Aide: We Didn't Even See Comey's Let...	1
1	1	FLYNN: Hillary Clinton, Big Woman on Campus - ...	Ever get the feeling your life circles the rou...	0
2	2	Why the Truth Might Get You Fired	Why the Truth Might Get You Fired October 29, ...	1
3	3	15 Civilians Killed In Single US Airstrike Hav...	Videos 15 Civilians Killed In Single US Airstrike ...	1
4	4	Iranian woman jailed for fictional unpublished...	Print \nAn Iranian woman has been sentenced to...	1
...
20795	20795	Rapper T.I.: Trump a 'Poster Child For White S...	Rapper T. I. unloaded on black celebrities who...	0
20796	20796	N.F.L. Playoffs: Schedule, Matchups and Odds -...	When the Green Bay Packers lost to the Washing...	0
20797	20797	Macy's Is Said to Receive Takeover Approach by...	The Macy's of today grew from the union of sev...	0
20798	20798	NATO, Russia To Hold Parallel Exercises In Bal...	NATO, Russia To Hold Parallel Exercises In Bal...	1
20799	20799	What Keeps the F-35 Alive	David Swanson is an author, activist, journa...	1

20800 rows × 4 columns

In [17]:

```
#Merging the two Data Sets
frame = [data,data2]
final_data = pd.concat(frame)
final_data
```

Out[17]:

	id	title	text	label
0	8476	You Can Smell Hillary's Fear	Daniel Greenfield, a Shillman Journalism Fello...	0

	id	title	text	label
1	10294	Watch The Exact Moment Paul Ryan Committed Pol...	Google Pinterest Digg LinkedIn Reddit Stumble...	0
2	3608	Kerry to go to Paris in gesture of sympathy	U.S. Secretary of State John F. Kerry said Mon...	1
3	10142	Bernie supporters on Twitter erupt in anger ag...	— Kaydee King (@KaydeeKing) November 9, 2016 T...	0
4	875	The Battle of New York: Why This Primary Matters	It's primary day in New York and front-runners...	1
...
20795	20795	Rapper T.I.: Trump a 'Poster Child For White S...	Rapper T. I. unloaded on black celebrities who...	0
20796	20796	N.F.L. Playoffs: Schedule, Matchups and Odds - ...	When the Green Bay Packers lost to the Washing...	0
20797	20797	Macy's Is Said to Receive Takeover Approach by...	The Macy's of today grew from the union of sev...	0
20798	20798	NATO, Russia To Hold Parallel Exercises In Bal...	NATO, Russia To Hold Parallel Exercises In Bal...	1
20799	20799	What Keeps the F-35 Alive	David Swanson is an author, activist, journa...	1

27135 rows × 4 columns

In [18]:

```
final_data2 = final_data.drop(columns='text', axis=1)
```

In [19]:

```
final_data2.head(5)
```

Out[19]:

	id	title	label
0	8476	You Can Smell Hillary's Fear	0
1	10294	Watch The Exact Moment Paul Ryan Committed Pol...	0
2	3608	Kerry to go to Paris in gesture of sympathy	1
3	10142	Bernie supporters on Twitter erupt in anger ag...	0

	id		title	label
4	875	The Battle of New York: Why This Primary Matters		1

In [20]:

```
final_data2.pop('id')
```

Out[20]:

0	8476
1	10294
2	3608
3	10142
4	875
	...
20795	20795
20796	20796
20797	20797
20798	20798
20799	20799

Name: id, Length: 27135, dtype: int64

In [21]:

```
final_data2.head(5)
```

Out[21]:

		title	label
0	You Can Smell Hillary's Fear		0
1	Watch The Exact Moment Paul Ryan Committed Pol...		0
2	Kerry to go to Paris in gesture of sympathy		1
3	Bernie supporters on Twitter erupt in anger ag...		0
4	The Battle of New York: Why This Primary Matters		1

In [22]:

```
final_data2.shape
```

Out[22]:

(27135, 2)

In [23]:

```
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
```

```
from sklearn.metrics import accuracy_score
import re
```

In [24]:

```
import nltk
nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to
/root/nltk_data...
[nltk_data]  Unzipping corpora/stopwords.zip.
```

Out[24]:

True

In [25]:

```
print(stopwords.words('english'))
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ou
rs', 'ourselves', 'you', "you're", "you've",
"you'll", "you'd", 'your', 'yours', 'yoursel
f', 'yourselves', 'he', 'him', 'his', 'himse
lf', 'she', "she's", 'her', 'hers', 'hersel
f', 'it', "it's", 'its', 'itself', 'they',
'them', 'their', 'theirs', 'themselves', 'wh
at', 'which', 'who', 'whom', 'this', 'that',
"that'll", 'these', 'those', 'am', 'is', 'ar
e', 'was', 'were', 'be', 'been', 'being', 'h
ave', 'has', 'had', 'having', 'do', 'does',
'did', 'doing', 'a', 'an', 'the', 'and', 'bu
t', 'if', 'or', 'because', 'as', 'until', 'w
hile', 'of', 'at', 'by', 'for', 'with', 'abo
ut', 'against', 'between', 'into', 'throug
h', 'during', 'before', 'after', 'above', 'b
elow', 'to', 'from', 'up', 'down', 'in', 'ou
t', 'on', 'off', 'over', 'under', 'again',
'further', 'then', 'once', 'here', 'there',
'when', 'where', 'why', 'how', 'all', 'any',
'both', 'each', 'few', 'more', 'most', 'othe
r', 'some', 'such', 'no', 'nor', 'not', 'onl
```

```
y', 'own', 'same', 'so', 'than', 'too', 'ver
y', 's', 't', 'can', 'will', 'just', 'don',
"don't", 'should', "should've", 'now', 'd',
'll', 'm', 'o', 're', 've', 'y', 'ain', 'are
n', "aren't", 'couldn', "couldn't", 'didn',
"didn't", 'doesn', "doesn't", 'hadn', "had
n't", 'hasn', "hasn't", 'haven', "haven't",
'isn', "isn't", 'ma', 'mightn', "mightn't",
'mustn', "mustn't", 'needn', "needn't", 'sha
n', "shan't", 'shouldn', "shouldn't", 'was
n', "wasn't", 'weren', "weren't", 'won', "wo
n't", 'wouldn', "wouldn't"]
```

In [26]:

```
port_stem = PorterStemmer()
```

In [27]:

```
def stemming(content):
    stemmed_content = re.sub('[^a-zA-Z]', ' ', str(content))
    stemmed_content = stemmed_content.lower()
    stemmed_content = stemmed_content.split()
    stemmed_content = [port_stem.stem(word) for word in stemmed_content if not word in stopwords.words('english')]
    stemmed_content = ' '.join(stemmed_content)
    return stemmed_content
```

In [28]:

```
final_data2['title']
```

Out[28]:

0 You Can Smell
 Hillary's Fear
 1 Watch The Exact Moment Paul Ryan Co
 mmitted Pol...
 2 Kerry to go to Paris in gestu
 re of sympathy
 3 Bernie supporters on Twitter erupt
 in anger ag...
 4 The Battle of New York: Why This P
 rimary Matters
 ...
 20795 Rapper T.I.: Trump a 'Poster Child
 For White S...
 20796 N.F.L. Playoffs: Schedule, Matchups

and Odds -....

20797 Macy's Is Said to Receive Takeover Approach by...

20798 NATO, Russia To Hold Parallel Exercises In Bal...

20799 What Keeps the F-35 Alive

Name: title, Length: 27135, dtype: object

In [29]:

```
final_data2['title'] = final_data2['title'].apply(stemming)
```

In [30]:

```
final_data2['title']
```

Out[30]:

0 smel
1 hillari fear
1 watch exact moment paul ryan commit
polit suic...
2 kerri go pari g
estur sympathi
3 bernes support twitter erupt ange
r dnc tri warn
4 battl new york
primari matter

...

20795 rapper trump poster child w
hite supremaci

20796 n f l playoff schedul matchup odd
new york time

20797 maci said receiv takeov approach hu
dson bay ne...

20798 nato russia hold parallel
exercis balkan

20799

keep f aliv

Name: title, Length: 27135, dtype: object

In [31]:

```
X = final_data2['title'].values
Y = final_data2['label'].values
```

In [32]:

```
print(X)
```

['smell hillari fear'
 'watch exact moment paul ryan commit polit
 suicid trump ralli video'
 'kerri go pari gestur sympathi' ...
 'maci said receiv takeov approach hudson ba
 y new york time'
 'nato russia hold parallel exercis balkan'
 'keep f aliv']

In [33]:

```
print(Y)
```

[0 0 1 ... 0 1 1]

In [34]:

```
vectorizer = TfidfVectorizer()
vectorizer.fit(X)

X = vectorizer.transform(X)
```

In [35]:

```
print(type(X[227]))
```

<class 'scipy.sparse.csr.csr_matrix'>

In [36]:

```
v = TfidfVectorizer(decode_error='replace', encoding='utf-8')
X = v.fit_transform(final_data2['title'].values.astype('U'))
```

In [37]:

```
print(X)
```

(0, 4803)	0.550944729105872
(0, 6195)	0.3249900893280307
(0, 12588)	0.7686620501293244
(1, 14661)	0.2193578349045018
(1, 10956)	0.2926686044143114
(1, 14022)	0.12825773413403818

(1, 13222)	0.3405981035359624
(1, 10396)	0.25305701991200313
(1, 2694)	0.34968578671445955
(1, 11747)	0.28887143515813046
(1, 9981)	0.27303572830011025
(1, 8773)	0.33299122848737406
(1, 4579)	0.45974446005706154
(1, 14890)	0.25924653752839827
(2, 13402)	0.5306462715649448
(2, 5475)	0.5534945049279915
(2, 9906)	0.3682960621988969
(2, 5575)	0.30926968920808784
(2, 7308)	0.4251689515882591
(3, 14871)	0.30374603979096404
(3, 13958)	0.3084085205820291
(3, 3842)	0.370764912246796
(3, 497)	0.42268418173172095
(3, 4468)	0.43851069071723386
(3, 14137)	0.34318190370579155
:	:
(27131, 11949)	0.4760420078856128
(27131, 10318)	0.48688421681605326
(27131, 9485)	0.4760420078856128
(27131, 8310)	0.50879143852567
(27131, 13745)	0.13066296506705072
(27131, 15326)	0.13283602915071527
(27131, 9204)	0.12698355749554052
(27132, 6401)	0.41316585607762357
(27132, 8064)	0.4200417943744182
(27132, 638)	0.364161918388557
(27132, 1171)	0.39300492117023966
(27132, 13456)	0.37358742524480626
(27132, 11792)	0.2973087188204927
(27132, 11091)	0.3212154500448458
(27132, 13745)	0.10787112783203183
(27132, 15326)	0.10966513942082355
(27132, 9204)	0.10483352766515326

```
(27133, 1029) 0.4920721764839098
(27133, 4608) 0.4558521893640597
(27133, 9889) 0.4920721764839098
(27133, 9088) 0.3488512370523384
(27133, 6261) 0.3547956388734035
(27133, 11730)          0.24566561980919474
(27134, 341)    0.778786765687214
(27134, 7270)   0.6272887481777817
```

In [38]:

```
print(X[5])
```

```
(0, 14489)      0.6191007694825683
(0, 13557)      0.7853115542420674
```

In [39]:

```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size = 0.2, stratify=Y, random_state=2)
```

In [40]:

```
model = LogisticRegression()
```

In [41]:

```
model.fit(X_train, Y_train)
```

Out[41]:

`LogisticRegression()`

In [42]:

```
X_train_prediction = model.predict(X_train)
training_data_accuracy = accuracy_score(X_train_prediction, Y_train)
```

In [43]:

```
print('Accuracy score of the training data : ', training_data_accuracy)
```

Accuracy score of the training data : 0.845
2183526810393

In [44]:

```
import pickle
```

In [45]:

```
filename = 'finalized_model.sav'
pickle.dump(model, open(filename, 'wb'))
```

In [46]:

```
loaded_model = pickle.load(open(filename, 'rb'))
result = loaded_model.predict(X_test[5])
print(result)
```

[1]

In [47]:

```
print(Y_test[5])
```

1

In [48]:

```
print(X_train[5])
```

(0, 9454)	0.5386709169349737
(0, 4935)	0.5386709169349737
(0, 15296)	0.3910790997551082
(0, 13745)	0.13262580910022567
(0, 4141)	0.3717047287438517
(0, 830)	0.27600353706346353
(0, 15326)	0.13483151736786492
(0, 9204)	0.12889112876497824

TESTING

Functional Test Cases

Test ID (#)	Test Scenario	Test Case	Execution Steps	Expected Outcome	Actual Outcome	Status	Remarks
TC-01	Verify User Registration	Accept Valid Username	<ul style="list-style-type: none"> User click on sign up User enter credentials Click sign up 	User should view a message "Registration successful"	User should view a message "Registration successful"	Pass	Sign up successful
TC-02	Verify User Registration	Don't Accept Username	<ul style="list-style-type: none"> User click on sign up User enter credentials Click sign up 	User should view a message "Registration successful"	User should view a message "Use name already taken"	Fail	Invalid username
TC-03	Verify Login Credentials	Accept login credentials	<ul style="list-style-type: none"> User click login User enter username and password Click login 	User can see his username at top right corner	User can see his username at top right corner	Pass	Login successful
TC-04	Verify login Credentials	Reject login credentials	<ul style="list-style-type: none"> User click login User enter username and password Click login 	User can see his username at top right corner	User sees a message "invalid username or password"	Fail	Invalid Username or Password
TC-05	Verify Relevance of article	Accept input and show result	<ul style="list-style-type: none"> User will either type or paste article 	User can see	User can see	Pass	Model Predicted

			<ul style="list-style-type: none"> User click on predict 	Real/Fake as a result	Real/Fake as a result		successfully
TC-06	Verify relevance of article	Reject Input	<ul style="list-style-type: none"> User will either type or paste article User click on predict 	User can see Real/Fake as a result	User will see Invalid Input	Fail	Input is either blank or just numbers or symbols
TC-07	Verify Search option	Accept Input shows result	<ul style="list-style-type: none"> User will type the title of a news User will click search 	User can see "News is present is database and it is Real/Fake"	User can see "News is present is database and it is Real/Fake"	Pass	Search successful
TC-08	Verify Search option	Accept Input and doesn't show result	<ol style="list-style-type: none"> 3. User will type the tile of a news 4. User will click search 	User can see "News is present is database and it is Real/Fake"	User will see the article is not present in the database	Fail	Search Unsuccessful

Non-Functional Test Cases

Test ID (#)	Test Scenario	Test Case	Execution Steps	Expected Outcome	Actual Outcome	Status	Remarks
TC-9	Load Testing	Request is sent successfully and data is received	6. The browser will send DNS request to the DNS Server 7. Server replies with the hostname to the DNS server 8. The browser will send HTTPS request 9. The server responds 10. The data is send to the user	Website will load under 3 seconds	Website will load under 3 seconds	Success	Website loaded successfully
TC-10	Load Testing	Request failed to sent	3. The browser will send the request to the server 4. The server doesn't respond	Website will load within 3 seconds	The user sees the prompt "Site can't be reached"		

TC-11	Performance Testing	Model predicts results within 5 seconds	2. The User enters the article and press predict	The model will display Real/Fake under 5 seconds	The model will display Real/Fake under 5 seconds	Success	Model predicted successfully
TC-12	Security Testing	Successfully authenticate user based on user credentials	5. The user will enter the login credentials 6. The system will encrypt the data and send it to the server 7. The server decipts the data and authenticates the data 8. The server sends the data back to the user	Login verified and user can access the website	Login verified and user can access the website	Success	System is secure
TC-13	Stress testing	Website should work even for thousand users at a time	4. Access the website from more than thousand different systems 5. get response from	Website should work for more than thousand users	Website should work for more than thousand users	Success	System can handle at least a thousand users

			each user if the website is working from each user				
			6. Verify the results				

Result:

Thus, the details of architectural design/framework/implementation along with the screenshots were provided.

CONCLUSION

By implementing a Logistic Regression model to detect fake news we hope to achieve an accuracy of above 85% at least. This can be compared to other more complex models that have been used to detect fake news and their accuracy. A result that we hope to observe is how accurately does the model detect fake news that we will be included in our dataset that we generated. Along with that we expect to gain more insights on what are the recognizing factors of fake news.