SHARDA UNIVERSITY
Beyond Boundaries

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING,
SCHOOL OF ENGINEERING AND TECHNOLOGY,
SHARDA UNIVERSITY, GREATER NOIDA

MOLECULE: Web-Based Open Community Networking Platform

A project submitted
in partial fulfillment of the requirements for the degree of
Bachelor of Technology in Computer Science and Engineering

by
Aditi Garg (2018015438)
Akash Sharma (2018002620)
Anant Sharma (2018023482)
Gaurav Jain (2018007425)
Mehak Bhatia (2018007470)
Nishant Bansal (2018007510)

Supervised by:
Dr. Megha Chhabra, Asst. Prof (CSE)

May 2022

CERTIFICATE

This is to certify that the report entitled "**MOLECULE: Web-Based Open Community Networking Platform**" submitted by "ADITI GARG (2018015438), AKASH SHARMA (2018002620), ANANT SHARMA (2018023482), GAURAV JAIN (2018007425), MEHAK BHATIA (2018007470) & NISHANT BANSAL (2018007510)" to Sharda University, towards the fulfillment of requirements of the degree of Bachelor of Technology is the record of bonafide final year Project work carried out by him/her in the Department of Computer Science and Engineering, School of Engineering and Technology, Sharda University. The results/findings contained in this Project have not been submitted in part or full to any other University/Institute for an award of any other Degree/Diploma.

Signature of Supervisor

Name: Dr. Megha Chhabra

Designation: Asst. Professor (CSE)

Signature of Head of Department

Name: Prof. (Dr.) Nitin Rakesh

(Office seal)

Place:

Date:

Signature of External Examiner

Date:

ACKNOWLEDGEMENT

A major project is a golden opportunity for learning and self-development. We consider ourselves very lucky and honoured to have so many wonderful people lead us through the completion of this project.

First and foremost we would like to thank **Dr. Nitin Rakesh**, HOD, CSE who gave us an opportunity to undertake this project.

My grateful thanks to **Dr. Megha Chhabra** for her guidance in my project work.

Dr. Megha Chhabra, despite being extraordinarily busy with academics, took time out to hear, guide and keep us on the correct path. We do not know where we would have been without his help.

CSE department monitored our progress and arranged all facilities to make life easier. We choose this moment to acknowledge their contribution gratefully.

Name and signature of Students

Aditi Garg (2018015438)

Akash Sharma (2018002620)

Anant Sharma (2018023482)

Gaurav Jain (2018007425)

Mehak Bhatia (2018007470)

Nishant Bansal (2018007510)

ABSTRACT

The world is changing rapidly, and so is the digital world. Everyone is coming online for interacting, socializing, networking etc. With this growing digitalization of the world, there comes a need for digital platforms, allowing people to connect with ease i.e., through a well-structured networking forum. There are many well established social platforms like Facebook, Twitter, Instagram, Quora, Reddit, Tumblr, LinkedIn and many others. All of them serve their purpose exceptionally well, they provide an interface to connect person A to person B, but what if someone wants to connect a group or community of people belonging to similar fields, for that purpose some platforms like Stack Overflow for developers, Deviant Art for artists, Kaggle for data scientists and many others provides exposure to a cluster of communities and allows the user to connect or network to a much bigger group of people with the same area of interest which can help the user in many ways. Though we have all these well-established platforms, that connect millions of people, still there is a need for a new platform, which allows people to ask questions on various domains, make communities or teams, have open discussions, and write open articles on related topics, hire professionals for help, have access to the latest news related to their field of work, etc. And no platform serves all these features as a whole, which is what we propose to offer. We will make this platform on the MERN technology stack, which is a free and open-source JavaScript software stack for building dynamic websites and web applications, and we will provide modules for creating communities/teams, open discussions, Q&A, hiring people, posting short videos over discussions, posting blogs on various topics, and news articles.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Chapter 01: INTRODUCTION

Nowadays, it has become very accessible to create new online communities which are due to the ever-rising increase in usage of the internet and all the new emerging technologies; most the tech companies like StackExchange, Instagram, Stackoverflow etc. have made the complete process very simple, but still, there are a lot of gaps to fill, especially in regards to open communities with members having similar interests. No other platforms offer an integrated platform with all-important networking features like an open community, open discussion within the community, a dedicated module for freelancing purposes, and a dedicated platform to write and share articles, and get trending news; which should be regulated by using ML, DL & NLP models to make the experience of the user much more customized, smoother & user-centric.

There are a lot of platforms present in the market, which offers community-based networking, but almost each of them lacks some other networking module, or the platform is too complex for a new user to understand. With the ever-rising increase in online community networking, there also arises the need to address the mentioned problems, to create a whole networking experience, integrated with all the necessary networking modules, which should prove a wholesome networking experience to the customer.

Along with the website for networking, the functionality of the website needs some backend work to make the website much more functional, and integral for the users. For this website we have decided to work upon multiple machine learning models to integrate into the website, first model is based upon text classification using BERT, which will be integrated with two of our modules, initially, for the purpose of searching for the best result possible for the users.

The second machine learning model is based upon image classification, which will be used for nudity detection in images uploaded in one of our modules, which will help us to keep the website safe for the users. And the last machine learning model is also based upon text

classification, but will be used to verify the links uploaded to our website; it will check the integrity and security of that link when uploaded to the website.

That's why a website/platform with all the necessary networking modules like question and answer, post sharing, article sharing and freelancing integrated into it, and guided by a few ML algorithms can make the website a successful and productive product for the users.

Overview:

To create an open networking environment, completely based upon community networking, we need to integrate multiple networking modules into a single website or platform so that people can network much more easily and efficiently. This will also require some machine learning models to work efficiently for better searching, nudity detection and link verification.

Project Description:

The project is based upon MERN stack development, in which we have created a community-based networking website for users to interact within the community or communities of their choice, which will be integrated with major networking modules like question and answer module, post sharing module, article sharing module and freelancing module. To make the website much more functional and secured we are also using multiple machine learning models which will be used for searching for better results, nudity detection in images and link verification to ensure the uploaded link is secured or not.

1.1 Problem Definition

The project's goal is to provide a well developed and simplistic community-based networking website or platform so that people can build their network in the community of their choice efficiently, and the website has to be regulated by some machine learning modules to keep the website efficient and secured.

The project aims to provide a mix of functionalities to its user, just like a few other existing platforms on the web, for example, QnA modules like Stackoverflow or Quora, Post sharing

modules like LinkedIn, and Article sharing modules like Medium and Freelancing modules like Upwork.

These four modules are the major networking modules present on the web today, and a user has to create individual profiles on each of these platforms, and the actions performed on one website are not recognised by the other, which is not good for the user, as he or she should have been able to reflect their actions performed on one networking website onto another networking website, and that is the problem we are trying to solve with this project.

Also to keep the website efficient and secured, a few machine learning models are also being used for better search results, nudity detection in images and link verification.

1.2 Project Overview / Requirement Specifications

To deal with all the above-mentioned problems, we aspire to create an open community networking platform which will allow users, from any community, interest or profession to join communities dedicated to a single area, in which they can build their network better and can share their thoughts and knowledge with people having similar interests as themselves. We know that fields like IT already have a lot of community-based platforms present on the web, but people belonging to different professions like Commerce, Arts etc, have very less or not-famous platforms to engage in a community discussion or knowledge sharing. This makes our aim clear, to create an application/platform which is open to all, open to choosing communities of choice and to engage in multiple networking modules in a single platform. The website will be integrated with the following features:

- It will allow the user to join the communities of their choice.
- It will be integrated with all major networking modules like question and answer (discussion), article writing and sharing, hiring freelancers, registering self as a freelancer, posts sharing and reading the latest news.

1.2.1 Functional Requirements

1.2.1.1 Introduction

This venture aims to provide a website or platform based upon community networking, to its users, integrated with all the major networking modules like question and answer, post sharing, article sharing and freelancing to provide the user with a whole networking experience. The user will be allowed to choose the community of his/her choice to get the most efficient networking experience. It will also be regulated by multiple ML modules to keep it efficient and secure.

1.2.1.2 Input

The input for the website is just the user, who will be asked to log in with their respective email id, and after that, we will be asking for some essential information about them and also their community of choice, and after that, the user will be free to make the most of the website.

For the machine learning models, the first model is based upon text classification, so it will be needing some text as input to output the best search results. The second machine learning model is based upon image classification, which will be needing some images as input to perform nudity detection and classify it as safe for work or not safe for work and the final model will be needing a link to verify the integrity of that link, whether the link is secured or not.

1.2.1.3 Error Handling

Expectation and Try Catch are used to monitor the handling of a few defined errors. Like the user must input the right value in the forms as per the field, and if the user's input is incorrect, the machine will detect it and shows an error to verify the given input.

1.2.2 Normal Requirements

These are the requirements which are specified considering the requirements of the consumer.

N1: The website should have a secured database to keep the user collected data secured.

N2: The UI of the website should be simple to understand and operate.

N3: The website must be responsive and functional in all its basic functionalities.

N4: The ML models must be trained over a good amount of data for efficient working.

N5: The website must be quick in actions and responses.

1.2.3 Non-Functional Requirements

1.2.3.1 Performance Requirements

The performance is mainly measured on high execution PC, which might involve multiples of the accompanying reaction time for any action performed by the user. This may vary as the project is using multiple technologies based upon different sources, like ML, cloud, database etc., so the performance can vary in every system.

1.2.3.2 Reliability

The project is reliable at its core, as all the data collected from it is stored in a secured database and will not be used in any other activities except within the website. Also, the project does not contain any outdated libraries or modules from which the PC or system might get affected.

1.2.3.3 Security

The project is highly secured as all the packages and libraries used to develop the project are approved by top organizations or are open source which does not allow any security breach from the website. Also, the collected user data is stored in MongoDB, which is one of the most secured databases present on the web, and also the creators of the website do not use the data in any other activities.

1.2.3.4 Maintainability

The project is maintained over time, and all the libraries, packages, and modules used in the website are regularly checked for any security updates or new versions, which are then integrated into the website as per the requirement. Also, the website is based upon the AGILE model, so the website is regularly updated as per the user's and venture's requirements to keep the website up to date.

1.2.3.5 Ability to Learn

It is basic and simple to operate & understand, & it also improves the networking experience of the user.

1.3 Hardware Specifications

Minimum Requirements	Windows
Operating System	Windows 7
RAM	2 GB RAM
Processor	Dual core, Intel i3
Internet Connectivity	Required and should be stable

Table 1: Minimum Hardware Requirements

Recommended Requirements	Windows
Operating System	Windows 10
RAM	2 GB RAM or more
Processor	Dual core, Intel i5 or i7
Internet Connectivity	Required and should be stable and high

Table 2: Recommended Hardware Requirements

1.4 Software Specifications

1.4.1 Any Latest Web Browser (e.g. – Google Chrome)

1.4.2 Any suitable Code Editor (e.g. – Visual Studio Code Editor)

1.4.3 Python Notebook (e.g. – Jupyter Notebook)

1.5 Technical Specifications

MERN Stack	- M: MongoDB as Database - E: Express JS for maintaining the server - R: React JS for frontend - N: Node JS for backend
Machine Learning	- Deep Learning - Natural Language Processing
Languages	- Python - JavaScript - HTML - JSX - CSS - ES6

Table 3:Technical Specifications

Chapter 02: LITERATURE SURVEY

2.1 Existing System

We did our research, by searching and exploring all the existing platforms or applications which try to provide a similar kind of experience to the users, which we are proposing, we did find a few platforms to provide useful services and features to the users, but almost each of them had some shortcomings in them. No one offers the same what we are offering, so have included only those platforms which provide their service at its best and are most relevant to our proposed system.

Name	Features	Shortcomings
Quora	o It allows users to "follow" other people and topics o It allows knowledge-sharing	o No hiring system o Can't create a team o No feature for personal contact o No feature to showcase works done on other websites
Stackoverflow	o It allows an IT-based discussion platform o It allows a large number of developers to create and engage in communities	o Doesn't allow the uploading of blogs or videos o Not a general platform o No hiring system o Corresponds to a single field
LinkedIn	o Open netwokring platform o Allows posts sharing o Allows job posting and hiring o Allows knowledge sharing	o No dedicated feature for discussions o Mainly focused on the professional culture of the society o Not dedicated to the arts community

Medium	o Community-based platform o Allows article sharing	o No feature for posts, discussions and freelancing o Only dedicated to articles
Upwork	o Platform for freelancers o Allow to register as a freelancer and hire a freelancer	o No module for discussion, knowledge sharing and posts sharing o Only dedicated for freelancers o No feature to showcase works done on other websites
StackExchange	o Open Community based Networking Platform o Allows multiple communities to integrate at one place and do discussions, questions and answers.	o Only dedicated to QnA module o No module for posts sharing and freelancing o No feature to showcase works done on other websites
Scoold	o Fully featured Q&A platform o Allows badging system for ranking	o No hiring system o No chat feature is available
Answeree	o Get paid for answering o Can contact communities for answers	o Can't create a team o No hiring system
Mdtalks	o No subscriptions, it is free o Free personalized answers to medical questions	o No hiring system o Can't create a team o Only ask questions
Law Community	o Provides Internship opportunities o Get the latest updates o Have good library support	o No Q&A based module present o Corresponds to a single field only

Table 4: Literature Survey

2.2 Proposed System

We propose a few existing core features and some new features in our platform, and all of them will be implemented in UI in such a way that it will ease and increase the customer experience.

2.2.1 Provided Features:

i. Community Selection: Users are allowed to choose the community of their choice while registering themselves. This will allow them to have a much more customized and efficient networking experience.

ii. Questioning: Users can ask questions in their community.

iii. Answering: Users can reply/answer the questions asked in their community, allowing knowledge sharing.

iv. Posts Sharing: Users can share posts in their community, to which other community members can react by liking or commenting.

v. Freelancing: Our platform provides a feature for users to register themselves as freelancers so that people can hire them.

vi. Hiring: Our platform provides a feature for users to hire registered freelancers using the information provided by the freelancer.

vii. Article Writing: Users can write blogs/articles of their choice, related to any domain and can publish their articles in their community.

viii. Article Reading: Users can read the articles, published in their community.

ix. Searching: Machine Learning-based searching allows users to search for the best results in QnA and Post modules.

x. Nudity Detection: Machine Learning-based nudity detection on images, integrated with the Posts sharing module will make the website safe and secure.

xi. Link verification: Machine Learning-based link verification will detect whether the uploaded link is secured or not.

xii. News: Our platform also offers trending news using API.

2.2.2 Machine Learning Modules:

i. Searching: A BERT based machine learning module will be used to provide the best search results to the user, which will be initially integrated with two of our modules i.e., QnA and Post sharing module. This will allow users to search for the best matching questions, if available on the website or to search for best matching posts if available on the website. This will make the website much more efficient and productive for the user.

ii. Nudity Detection: A CNN based machine learning module will be used to detect nudity in the images uploaded to the website, which will be initially integrated with only one of our modules i.e., the posts sharing module. This will allow us to make the website much safer for workers and secured so that no user can misuse the website in any form.

iii. Link Verification: A machine learning module will be used to detect the integrity of a link, uploaded on the website, and will detect whether the link is secured or not. Initially, it will only be integrated with one of our modules i.e., the article module. This will allow us to make the website safe and secure so that no one can share an anonymous link on the website containing some malware.

2.3 Feasibility Study

We performed a thorough feasibility check of our platform in every possible way.

2.3.1 Operational Feasibility

Our product is majorly designed to serve the following functionalities:
- Ability to choose a community of choice

- Discussion platform (Question and Answer)

- Post sharing platform

- To search for freelancers to hire

- Freelancer registration

- Article publishing and reading platform

- Access to trending news

As every feature in the list uses existing technologies to work and doesn't need any new technologies to be built. It is operationally feasible.

2.3.2 Economic Feasibility

- Hardware – All the members involved in this project have their systems to work, which satisfy all the recommended requirements needed to work on this project efficiently and also a stable internet connection.

- Software – All the mentioned software, which we used during this project are freely available. (like VS-Code, Google Chrome and Jupyter Notebooks)

- Technical – We are using ExpressJS for maintaining the server and MongoDB as our database, which is free to use up to a certain number of entries.

2.3.3 Technical Feasibility

As the web development and machine learning modules are technically possible, are the component of the state of the art and have regular updates on its technological resources, we can safely conclude that the project is technically feasible.

2.3.4 Financial Feasibility

The website requires some financial resources to run, like cloud-based services to host Machine Learning modules, and a server to deploy the website and database to keep the data secured and running; it is financially feasible, as these services are cheap and affordable and the revenue generated from the website can cancel out all the investments.

2.3.5 Project's Market Place

The project will be designed to serve as a mediator between people belonging to different communities, based on their profession, interests, field of work etc.

The target audience for this online platform is the people with internet access, and a computing system, who want to connect to more and more people with the same interest as theirs.

The majority of the target will be the youth, and it will encourage more discussions with professionals, more knowledge sharing, more exposure to real-world problems and will allow more and more people to build their network.

2.3.6 Legality and Ethics

The project is planned to serve online users only, and will not promote any illegal activities like computer-related frauds, cyber defamation, cyber harassment, child predation, identity theft, extortion, travel scam, stock market manipulation etc.

Everything planned for this project will be ethical, as it will only be designed to promote educational exposure, knowledge sharing, and open discussions & debate.

We will also try to make sure that any user found doing any illegal or unethical activities, will be banned from the platform.

DON'Ts for People (Guidelines):
- Write lengthy comments beyond the scope of the original post.
- Make off-topic remarks.
- Degrade others based on gender, race, class, ethnicity, national origin, religion, sexual preference, disability, or other classification.
- Use language that is libellous, defamatory, obscene, threatening, offensive, demeaning, derogatory, disparaging, or abusive, or post links to content that contains any of this language.
- Use profanity.
- Attack other commenters.

2.3.7 Risk List

There can be some risks while working on this project:

i. Not getting a good number of initial users – As it'll be a networking platform, it can only run if users run it, so to make sure we get a good number of users at the start, we will ask our friends and family to register and try to make some contribution to the platform.

ii. Maintenance of Server – As we will be using the free version of the ExpressJS server to run our project, it will only allow a certain number of entries, so we will make sure, we do not cross that limit, as we have to pay for service otherwise.

iii. Not matching deadline – To deal with this, we have designed our work plan such that, we will have a time buffer of around 1 to 2 months.

iv. Unavailability of data to work upon – As it will be a prototype so we have planned to initially work on medical data, to ensure the working of the project.

2.3.8 Limitations

Initially, our product will be having few limitations:

- For the first deployment, our focus will be on providing the proposed key features like the QnA platform, Articles publishing, Hiring platform and community page but all the modules will be very basic in working and will be serving the service but with a little less customer interaction. With our second deployment, we will improvise it according to data collected from the first deployment.

- For the very first deployment, we are planning to develop a QnA database only for the medical field community, while moving forward, we will increase the communities and database for them.

- When deployed it will not be able to handle a big number of users, as we will be using a free service provided by Firebase.

- For the development we will be using localhost, so we will not be able to check its networking feature until it gets deployed on the server.

2.3.9 Future Scope

All the technologies (like ML, Web development etc.), concepts and ideas (networking, online discussions etc.) behind this project, are based upon new emerging technologies, emerging interests, and new trends. So, we can argue the fact that any platform working on these features has a high chance to succeed if executed properly. If our project does well in its prototype stage, we can further take it to next level, by uploading data from all types of fields (like engineering, law etc.) and allowing other people from different interests to participate in it, as for now the prototype is planned to work with medical data only. And a lot of new features can be added and existing features can be optimized to serve the best experience to the customers. With time and an increasing user base, we can increase the server size to handle more number of users, we can add much more features in our modules to improve the user engagement like upvotes and downvotes on articles and questions, rating features on freelancers etc. All of these features can make this project a well establish networking platform.

2.4 Risk Management

2.4.1 Risk Identification

R1. Website Related: The website might hang in the process, due to any code error or server issues, which might result in data loss or customer dissatisfaction.

R2. Customer Related: The customer might face some issues with the functionality of the website or fear the data could be misused.

R3. Data Security Related: The data saved in the database can be breached if some issue came in the code for data handling and data collection.

R4. Machine Learning Related: Machine learning modules might not perform as they were supposed to, so the user might face issues with wrong information.

R5. Server Related: If a large number of users visit the website, the server could go down, due to overloading.

2.4.2 Strategies used to manage Risks

S1. The website code must go through multiple test cases to ensure the code's credibility, and also it should be thoroughly checked when deployed on the server.

S2. All the queries of the customer must be greeted at the earliest and we must ensure that the customer must not face any difficulty while operating the website functionalities.

S3. The code used for data collection and data security must be thoroughly checked and updated from time to time to keep the data secured.

S4. Machine learning modules must be checked for their predictions and must be updated with the latest libraries for better efficiency.

S5. Using analysis and past information, server size must be changed as per the customer's behaviour so that server must not go down.

Chapter 03: SYSTEM ANALYSIS & DESIGN

3.1 Flowchart / Data Flow Diagrams / ER Diagram / Use Case Diagram
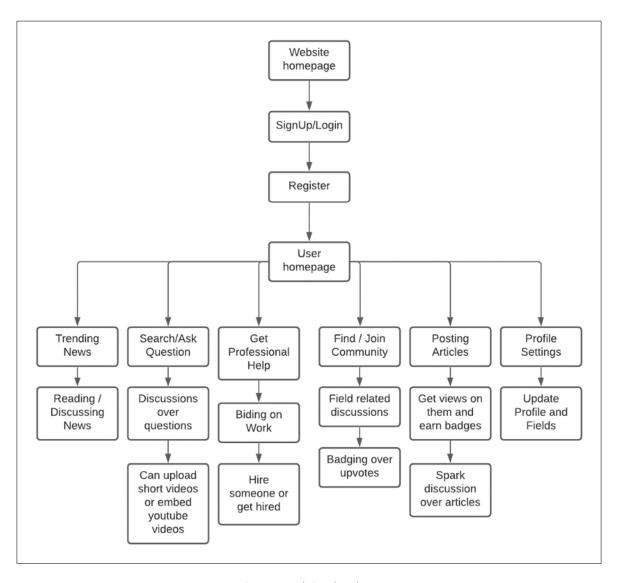
3.1.1 Flowchart for website

Figure 1. Website Flowchart

The above chart shows the flow of our website, and it mainly has the following steps:

- **Website Landing Page:** The user will land on the landing page, where the login button is provided to the user to either register himself or log in.

- **Signup/Login:** Here, if the user is visiting the first time, then he/she will be asked for registration and have to fill 2 forms to register himself, in which they will choose the community of their choice, and if the user is already registered then he/she will be redirected to the homepage of their profile.

- **Website Landing Page:** From the homepage of the website which is the Posts Sharing module, the user can navigate to other modules as well i.e. QnA module, Freelancing Module and Article Sharing Module. Users will also get to read news which will be fetched using API on the homepage.

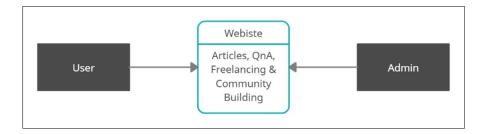### 3.1.2 Data Flow Diagram for the Website

i. Level 0 DFD

Figure 2. Level 0 - DFD

The above Level 0 DFD clearly shows how the website will work on the most basic level; the user will interact with the website and the admin will be regulating the website from the backend.
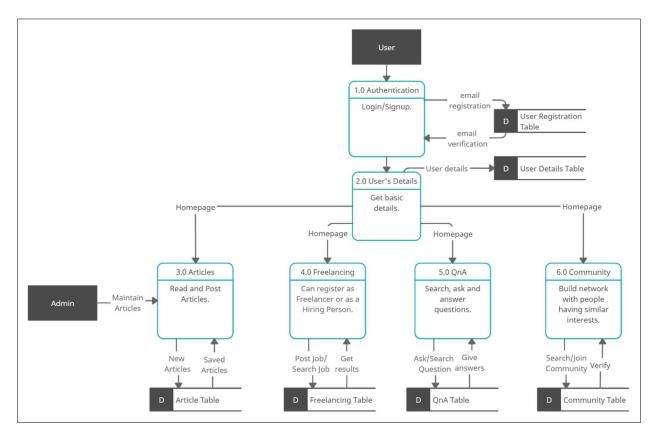
ii. Level 1 DFD

Figure 3. Level 1 - DFD

The above Level 1 DFD clearly shows how the website will work on its secondary level, it clearly shows how the modules will be integrated into the website and how the database will be connected to it. We can see the following steps:

1. User authentication with the website, and how the process will take place, when registered, the user will be redirected to the homepage of the website.
2. How the data collected from registration will be saved into the database.
3. User's interaction with the modules on the basic level.
4. How the data will be collected into the database and is getting used, in each module.

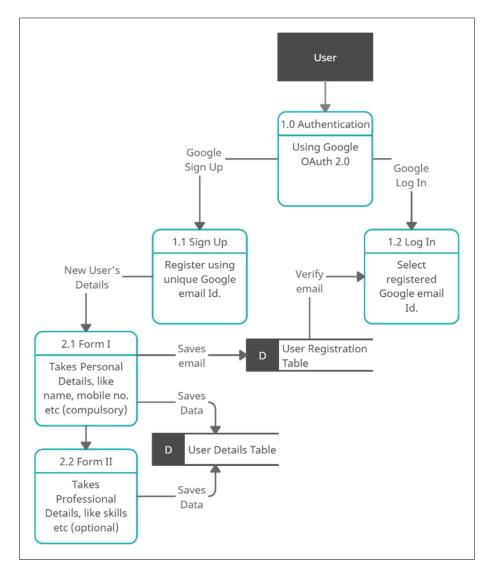### iii. Level 2 (Authentication & Registration)

Figure 4. Level 2 - DFD (Authentication & User's Details)

The above Level 2 DFD for Authentication and Registration clearly shows how these two modules will work at the core of the website. We can see the following steps:

1. User authentication and registration on the website.
2. For first time users, details get saved into the database.
3. For already registered users, email verification.
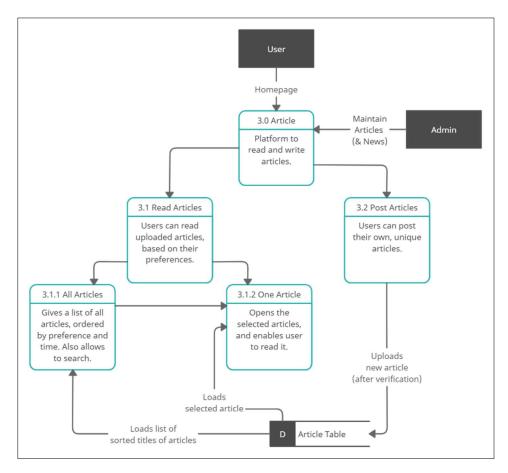
iv. Level 2 (Article)

Figure 5. Level 2 - DFD (Article)

The above Level 2 DFD for Article clearly shows how the module will work at the core of the website. We can see the following steps:

1. User exploring the homepage of the article, where the user can navigate through the published articles.
2. User's access to published articles and ability to read them.
3. User's access to write and publish articles on the application.
4. How the articles are getting saved into the database and how the data is getting used.

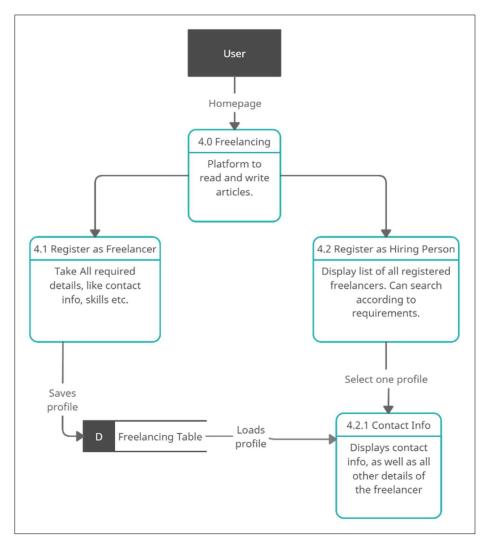## v. Level 2 (Freelancing)

Figure 6. Level 2 - DFD (Freelancing)

The above Level 2 DFD for Freelancing clearly shows how the module will work at the core of the website. We can see the following steps:

1. User exploring the homepage of freelancing, where the user can navigate through the registered freelancers to contact them for hiring.
2. User's access to registered freelancers and ability to contact them.
3. User's access to registering themselves as a freelancer on the application.
4. How the information about the user is getting saved into the database and how the data is getting used.
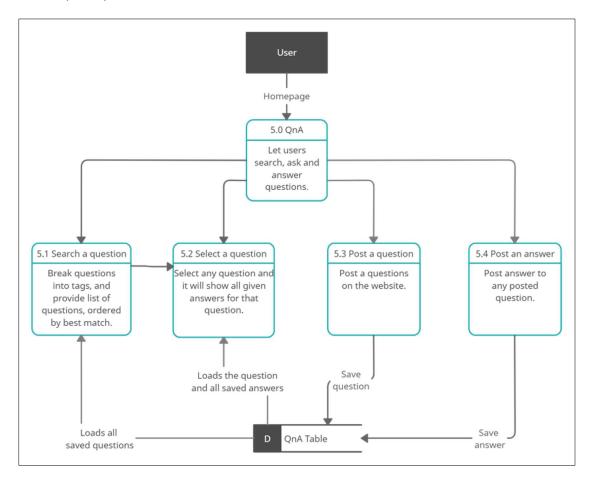
vi. Level 2 (QnA)

Figure 7. Level 2 - DFD (QnA)

The above Level 2 DFD for QnA clearly shows how the module will work at the core of the website. We can see the following steps:

1. User exploring the homepage of QnA, where the user can navigate through all the posted questions in the community.
2. User's ability to search for any question and read all the posted answers for the question.
3. User's ability to post/ask any question.
4. User's ability to put a comment and answer any question.
5. How the questions, answers & comments are getting saved into the database and how the data is getting used.
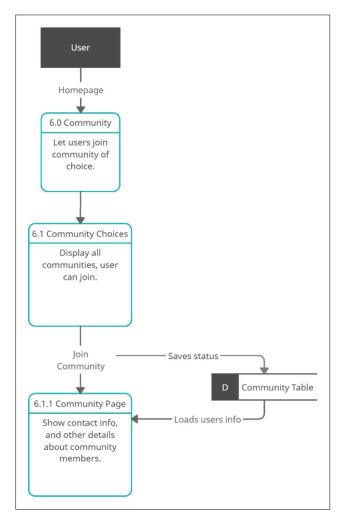
vii. Level 2 (Post Sharing/Community)

Figure 8. Level 2 - DFD (Community)

The above Level 2 DFD for Post sharing clearly shows how the module will work at the core of the website. We can see the following steps:

1. User exploring the homepage of post sharing, where the user can navigate through all the posts on the community wall.
2. User's ability to search for any post on the community wall.
3. User's ability to post on the community wall.
4. Users' ability to like any post on the community wall.
5. User's ability to read trending news which is fetched through API.
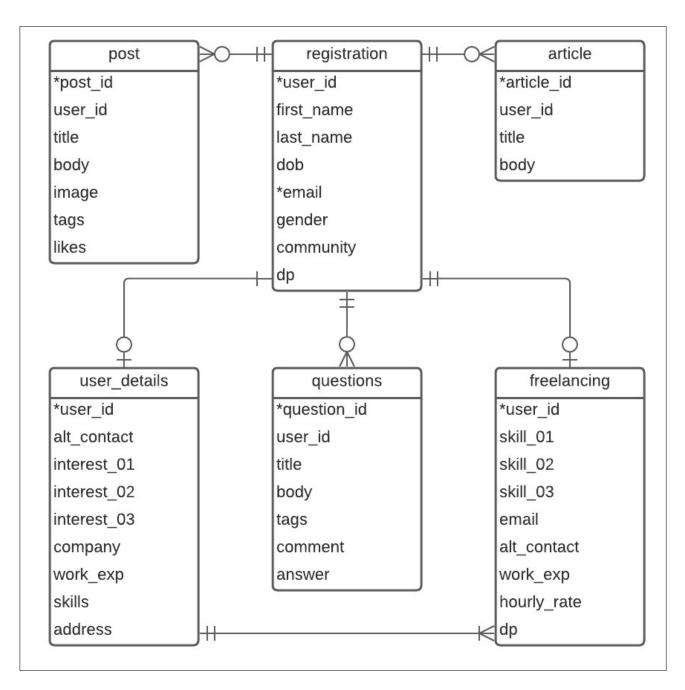
3.1.3 ER Diagram for the Website

Figure 9. ER Diagram

The above Entity-Relationship Diagram clearly shows the data distribution throughout the website. We can see every table used on the website and how they are related to each other. The tables are as follows:

1. **registration:**
   i. *user_id:* Unique id for each user (Primary Key)
   ii. *first_name*: first name of the user
   iii. *last_name*: Last name of the user
   iv. *dob*: Date of Birth of the user
   v. *email*: Email id of the user (from google log in)
   vi. *gender*: Gender of the user
   vii. *community*: Community chosen by the user
   viii. *DP*: Display-Picture of the user (from google log in)

2. **user_details:**
   i. *user_id*: Unique id for each user (from registration table)
   ii. *alt_contact*: Alternative contact info of the user
   iii. *interest_01*: First Interest of the user
   iv. *interest_02*: Second Interest of the user
   v. *interest_03*: Third Interest of the user
   vi. *company*: Current company, the user is working at (if any)
   vii. *work_exp*: Work experience of the user
   viii. *skills*: List of skills of the user
   ix. *address*: Address of the user

3. **freelancing:**
   i. *user_id:* Unique id for each user (from registration table)
   ii. *skills_01:* First skill of the user to show in freelancing profile
   iii. *skill_02*: Second skill of the user to show in freelancing profile
   iv. *skill_03*: Third skill of the user to show in freelancing profile
   v. *email*: Email id of the user (from google log in)
   vi. *alt_contact*: Alternative contact info of the user (from user_details table)
   vii. *work_exp*: Work experience of the user (from user_details table)
   viii. *hourly_rate*: Hourly rate of the user in INR
   ix. *DP*: Display-Picture of the user (from google log in)

4. **questions:**
   i. *question_*id: Unique id for each question (Primary Key)
   ii. *user_id*: Unique id for each user (from registration table)

iii.   *title*: Title of the question
iv.   *body*: Body of the question
v.   *tags*: Tags of the question
vi.   *comment*: Comments on the question
vii.   *answer*: Answers to the question

5. **posts:**
   i.   post_id: Unique id for each post (Primary Key)
   ii.   user_id: Unique id for each user (from registration table)
   iii.   title: Title of the post
   iv.   body: Body of the post
   v.   image: Image uploaded in the post
   vi.   tags: Tags of the post
   vii.   likes: Count of likes on the post

6. **article:**
   i.   article_id: Unique id for each article (Primary Key)
   ii.   user_id: Unique id for each user (from registration table)
   iii.   title: Title of the article
   iv.   body: Body of the article
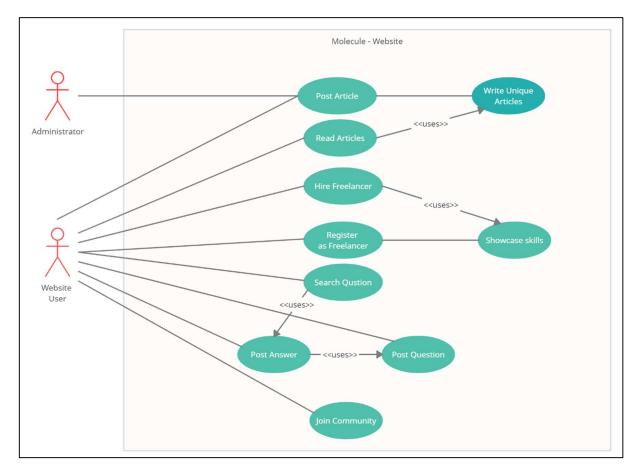
3.1.4 Use Case Diagram for the Website

Figure 10. Use Case Diagram

The above Use Case Diagram clearly shows how the website would work considering the user and admin. We can see the following steps:

1. User 's ability to have access to multiple features on the website like Share, view and react on posts, read news, ask and search questions, comment and answer on the posted questions, hire a freelancer, register self as a freelancer, publish or read articles.
2. Admin's involvement in regulating all the content on the website, regulating user's actions on the website and managing the website and all its module integrity and working.

3.2 Algorithms

3.2.1 Algorithm for BERT Based Machine Learning Model (Searching)

1. Procedure: MYPROCEDURE
   i. Query: ques entered by the user
   ii. Tokens: word tokenization
   iii. Clean_tokens: tokens without stop words
   iv. Base_tokens: tokens obtained after stemming or lemmatization
2. If Base_tokens in Database then do steps 7 and 8
3. Result: output from the database
4. End
5. Model : base_tokens(input)
6. Result: output from the model
7. Store the result in Database for future
8. End

3.2.2 Algorithm for CNN Based Machine Learning Model (Nudity Detection)

1. Image is taken from the web or mobile application.
2. Image is fed to the trained CNN network.
3. CNN network consists of many layers of convolution, and max-pooling, layers.
4. Convolution layers try to find out the features in the image by convolving various kinds of filters.
5. Max pooling layers try to reduce the size of the image in general by pooling those pixels which contain maximum information.
6. Each layer has a relu activation function which activates the neuron only when it has a value greater than 0.
7. Then finally flattened layer is used to make all the tensors flat.
8. These flat tensors are then fed to a dense layer and they are further connected to the output layer which has neurons equal to the number of categories.
9. Then at the output layer softmax is used to get the prediction from the network.
10. Then this array must be converted to a single value by calculating the max value index and accordingly assigning the category.
11. Then after classification is done, a label is provided to the application whether the image is safe for work (non-nude image) or safe for work (nude image).
12. End.

3.2.3 Algorithm for determining Phishing Links

1. URL is taken as input.
2. Check if entered URL is present in the database.
3. If the URL is present in the database and the result is a Phishing URL:
    i. Stored in database
    ii. URL adds to blacklist
4. Else-If URL is not present in the database:
    i. Feature extraction is done and a feature vector created
    ii. Vector passed to various models
    iii. 3 models give output: SVM, LR and Random Forest
    iv. The final result is calculated
    v. If the result is Phishing, stored in the database and gets added to the blacklist
5. End

3.3 Testing

3.3.1 Test Case for Registration

Use Case ID	1
Test Case Name	Check User Registration Process
Test Case Description	The user registration process must be smooth and without bugs
Steps	1. If first time user, select any Gmail id while login
	2. Fill all the necessary details in the form
	3. Press-Register
Expected Results	The user should be registered successfully
Actual Results	As expected

Table 5. Test Case 01 - User Registration

3.3.2 Test Case for Sharing Post

Use Case ID	2
Test Case Name	Check Post Sharing Feature
Test Case Description	Post sharing in post-module should be smooth and working
Steps	1. Click on Add Post button
	2. Provide title, body, image and tags for the post
	3. Press the Post button
Expected Results	The post should be shared successfully
Actual Results	As expected

Table 6. Test Case 02 - Post Sharing

3.3.3 Test Case for Posting Question

Use Case ID	3
Test Case Name	Check Question Posting Feature

Test Case Description	The question posting feature should be working and smooth
Steps	1. Click on Add Question button
	2. Provide title, body, and tags for the question
	3. Press the Add Question button
Expected Results	The question should get posted successfully
Actual Results	As expected

Table 7. Test Case 03 - Posting Question

3.3.4 Test Case for Comment/Answer on Question

Use Case ID	4
Test Case Name	Check Comment & Answer Feature on Question
Test Case Description	Putting comments and answers to questions should be work and smooth
Steps	1. Open any question
	2. Write a Comment and click the add button
	3. Write Answer and click add answer button
Expected Results	Comment/Answer should be added successfully
Actual Results	As expected

Table 8. Test Case 04 - Comment/Answer on Question

3.3.5 Test Case for Freelancer Registration

Use Case ID	5
Test Case Name	Check Freelancer Registration Process
Test Case Description	The freelancer registration process must be smooth, working and without bugs
Steps	1. Fill the registration form with all the necessary details
	2. Click the register button
Expected Results	Freelancer should be registered successfully
Actual Results	As expected

Table 9. Test Case 05 - Freelancer Registration

3.3.6 Test Case for Article Writing

Use Case ID	6
Test Case Name	Check Article Writing Feature
Test Case Description	Article writing feature must be smooth and working
Steps	1. Write the title and body of the article
	2. Click the post article button
Expected Results	The article should be published successfully
Actual Results	As expected

Table 10. Test Case 06 - Article Writing

Chapter 04: RESULTS / OUTPUTS

4.1 Website Frontend
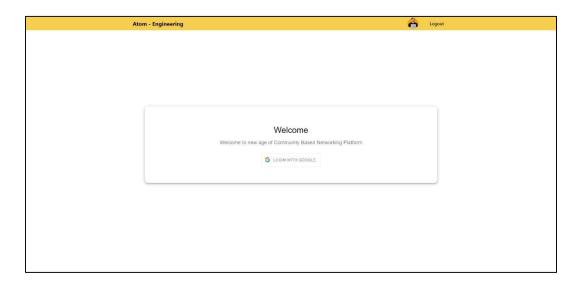
1. Landing page for our website:

Figure 11. Landing Page
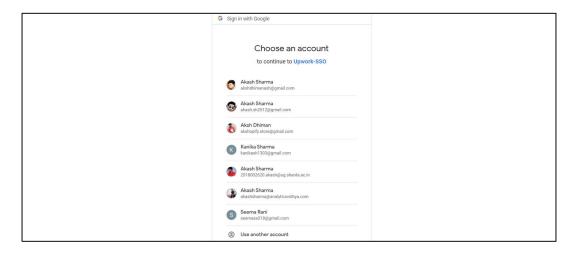
2. Login/Registration using Google Oauth:

Figure 12. Google Login

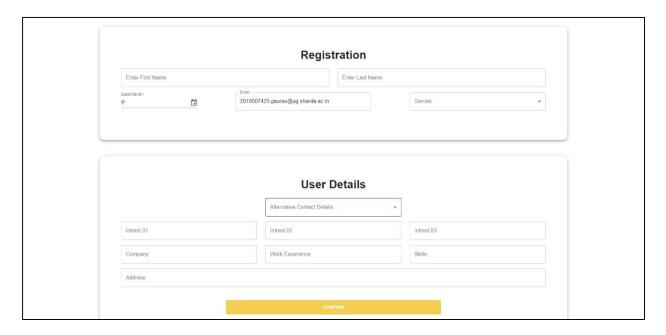## 3. Registration form for first-time users:

Figure 13. User Registration

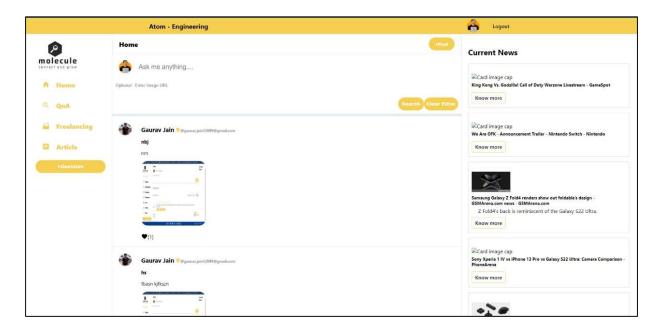## 4. Homepage for the Users or Post Sharing module (integrated with news):

Figure 14. Post Module / Homepage

5. Feature to add a post in the post-module, on the community wall:
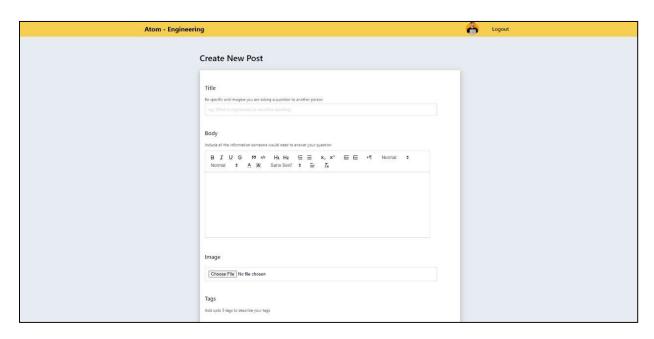

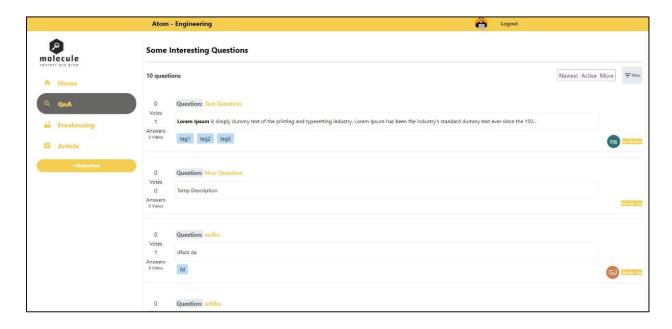*Figure 15. Add Post*

6. QnA module to view all the posted questions:

Figure 16. QnA Homepage

7. Feature in QnA module to read a question and add comment & answer to the question:
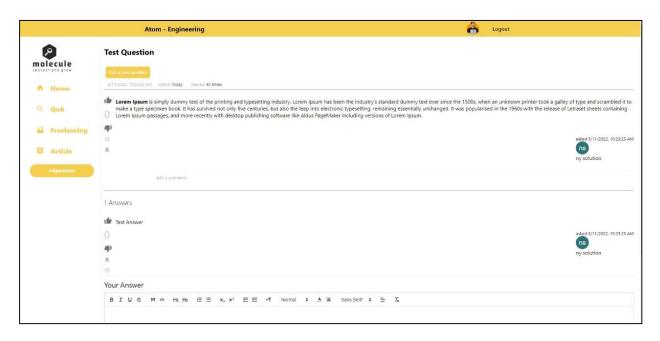

Figure 17. Comment / Answer on Question

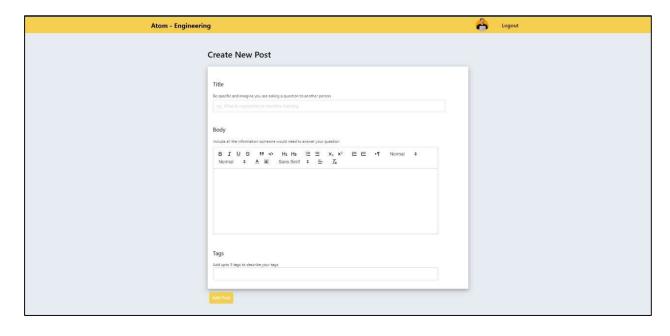## 8. Feature to add a question in the QnA module:

Figure 18. Add Question

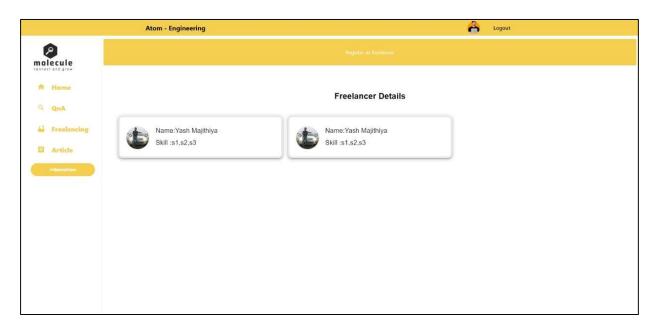9. Freelancing module to view all the registered freelancers:

Figure 19. Freelancing Homepage

10. Feature in freelancing module to view details of the freelancer:
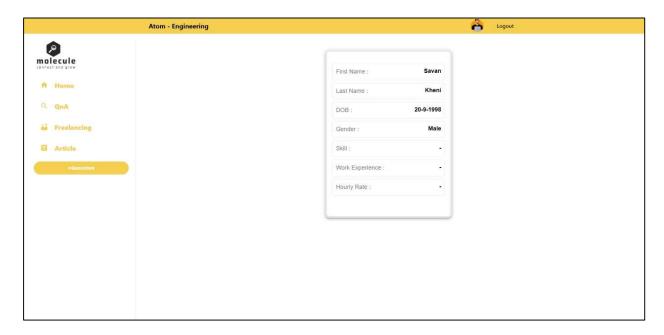

Figure 20. Freelancing Details

11. Feature to register self as a freelancer in the freelancing module:
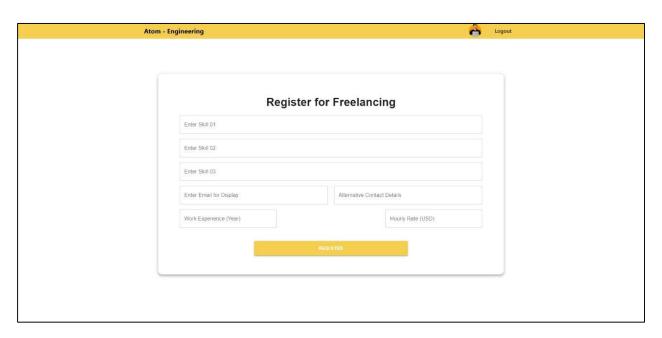

Figure 21. Freelancer Registration

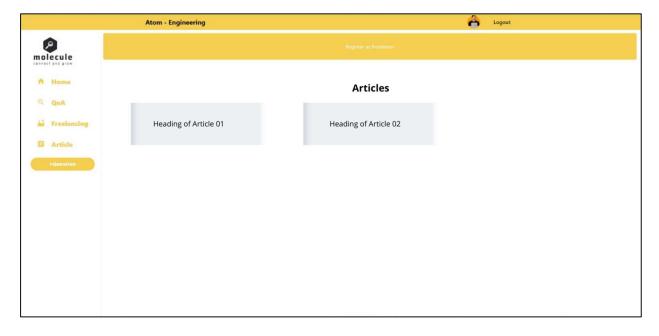12. Article module to view all the published articles:

Figure 22. Article Homepage

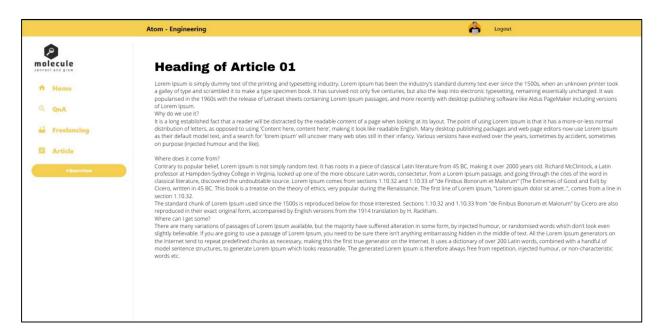## 13. Feature in article module to read any published article:

Figure 23. Article View

14. Feature in article module write an article:

Figure 24. Article Writing

4.2 Machine Learning Models Accuracy

4.2.1 Algorithm for BERT Based Machine Learning Model (Searching)

Labelling:
Classes : ['brainDisorder', 'general', 'inner_body_parts', 'pregnacy_menstrual', 'skin', 'cancer', 'covid19']
Encoded classes : [0 3 4 5 6 1 2]

Encoding Details:
Batch size: 50
Number of batches: 44

Dataset Details:
Total Data Size: 2237
Size of training data: 1737
Size of test data: 500

Model Details:
Optimizer Used: adam
Loss Metric Used : <keras.losses.SparseCategoricalCrossentropy object at 0x000002124ABE29B0>
Model Metrics : ['accuracy']
Default Number of Epochs: 5000
Actual Number of Epochs: 29
Batch Size used: 100

Test Dataset Results:
Test Loss: 0.5564653277397156
Test Accuracy: 0.7720000147819519

Model Results:
Max. Accuracy : 0.85
Min. Loss: 0.39
Max. Validation Accuracy : 0.87
Min. Validation Loss: 0.40

4.2.2 Algorithm for CNN Based Machine Learning Model (Nudity Detection)

Labelling:
NSFW: 1
SFW : 0

Image Details:

Image size used : (50, 50)
Single Image shape : (50, 50, 3)
Image colour channel : 3

Dataset Details:
Total Data Size : 24814
Train-Test Split Ratio : 8:2
Size of training data : 19851
Size of test data : 4963
X_train : 19851
Y_train : 19851
X_test : 4963
Y_test : 4963
X_train shape : (19851, 50, 50, 3)

Model Details:
Optimizer Used: adam
Loss Metric Used : <keras.losses.SparseCategoricalCrossentropy object at 0x000001B4551F3640>
Model Metrics : ['accuracy']
Default Number of Epochs: 5000
Actual Number of Epochs: 18
Batch Size used: 500

Test Dataset Results:
Test Accuracy: 84%

4.2.3 Algorithm for determining Phishing Links

Labelling:
Phishing: 1
Not Phishing : 0

Dataset Details:
Total Data Size : 11,055
Size of training data: 7,718
Size of test data: 3,336
The ratio of Train-Test Split: 70:30

Model Details:
Stacking: Support Vector Machine, Random Forest & Logistic Regression

Test Dataset Results:
Accuracy: 93%

Chapter 05: CONCLUSION

In the project report, we discussed the development of our website which is focused on providing an open community networking experience to its users in which the user can choose the communities of their choice and the website will be integrated with four major networking modules i.e., questions and answer module, posts sharing module, freelancing module and article sharing module. These four networking modules will help the user to build a profile within their selected community and build their network, which will be efficient for the user. For making the website much more secure and regulated, we are also using a few machine learning models, first one is for providing the best search results to the user, which will be integrated with QnA and Posts sharing module, and the second model will be sued for nudity detection in images uploaded in the posts sharing module and the last model will be used to check the credibility of any link uploaded on the website, to check whether the link is safe or not.

5.1 Further Improvement:

The website can be improved in many aspects, some of them are as follows:

- The landing page of the website can be made much more interactive and informative and can be integrated with a contact form and links for the website.
- The registration process can become better by integrating two forms into one.
- The Posts Sharing Module can be improved in terms of reactions, where a dislike function can be added, comments can be checked for vulgarity, share options can be added and it can become more responsive and happening.
- The QnA module can be integrated with upvotes and downvotes, and also a verified answer mark, it can be checked for vulgarity and can be integrated with an option to upload images into the question.
- The freelancing module can become much more functional, by integrating features like bidding, project posting etc, which can make the module much more interactive.
- The article sharing module can become more functional by adding features to check the grammatical errors, to edit texts much better, by adding an ML model to check its credibility etc.
- And finally, much more machine learning modules can be developed to be integrated with the website to check for vulgarity, nudity detection in videos, audio classification etc.

REFERENCES

[1] *Saleem Alhabash and Mengyan Ma, A Tale of Four Platforms: Motivations and Uses of Facebook, Twitter, Instagram, and Snapchat Among College Students? (2017)*

[2] *Peter Groenewegen and Christine Moser, Online communities: challenges and opportunities for social network research. (2014)*

[3] *Rajat Kathuria Mansi Kedia Gangesh Varma Kaushambi Bagchi, and Saumitra Khullar, Future of Work in a Digital Era: The Potential and Challenges for Online Freelancing and Microwork in India. (2017)*

[4] *Yaniv Dover, Guy Kelman, Emergence of online communities: Empirical evidence and theory. (2018)*

[5] *Neha Bhardwaj, Dr Anupama Vohra, Virtual Communities on Social Networks and the Indian Consumer. (2015)*

[6] *Mihai Gheorghe, State of Freelancing in IT and FuturecTrends. (2015)*

[7] *Shaowei Wang, David Lo, and Lingxiao JIANG, An Empirical Study on Developer Interactions in StackOverflow. (2013)*

[8] *Tomayess Issa, and Piet Kommers, Social Networking for Web-Based Communities. (2013)*

[9] *Katarina Stanoevska-Slabeva, and Beat F. Schmid, A Typology of Online Communities and Community Supporting Platforms. (2001)*

[10] *Michael Dunn, Fabian Stephany, Steve Sawyer, Isabel Munoz, Raghav Raheja, Gabrielle Vaccaro, and Vili Lehdonvirta, When Motivation Becomes Desperation: Online Freelancing During the COVID-19 Pandemic. (2020)*

[11] *Tanveer Ahmed, and Abhishek Srivastava, Understanding and evaluating the behaviour of technical users. A study of developer interaction at StackOverflow. (2017)*

[12] *Fabrˊıcio Matheus Gonc¸alves(B), Emanuel Felipe Duarte, Julio Cesar dos Reis, and M. Cecˊılia C. Baranauskas, An Analysis of Online Discussion Platforms for Academic Deliberation Support. (2017)*

[13] *Fei Gao, Tianyi Zhang, and Teresa Franklin, Designing asynchronous online discussion environments: Recent progress and possible future directions. (2013)*

[14] *Fei Gao, Charles Xiaoxue Wang, and Yanling Sun, A New Model of Productive Online Discussion and Its Implications for Research and Instruction. (2009)*

[15] *Alexandru Spatariu, Kendall Hartley, and Lisa D. Bendixen, Defining and Measuring Quality in Online Discussions. (2003)*

[16] *J. Zhang, L. Zhuo, Z. Li and Y. Yang, "An approach of bag-of-words based on visualattention model for pornographic image recognition in the compressed domain," Neurocomputing. (2015)*

[17] *D. I. Oddis, "Combating Child Pornography on the Internet: The Council of Europe's Convention on Cybercrime,". (2017)*

[18] *R. Ap-Apid, "An algorithm for nudity detection", College of Computer Studies, De La Salle University. (2017)*

[19] *M. M. Fleck and D. A. Forsyth, "Identifying nude pictures," Proc. Third IEEE Work. Appl. Comput. Vision. (1996)*

[20] *M. M. Fleck and D. A. Forsyth, "Finding naked people," Proceedings of European Conference on Computer Vision. (1996)*

[21] *D. C. Moreira and J. M. Fechine, "A Machine Learning-based Forensic Discriminator of Pornographic and Bikini Images," 2018 International Joint Conference on Neural Networks (IJCNN) (2018).*

[22] *C. Platzer, M. Stuetz, and M. Lindorfer, "Skin sheriff," Proceedings of the 2nd international workshop on Security and forensics in communication systems - SFCS '14, (2014).*

[23] *Zheng, Q.-F., Zeng, W., Wen, G., Wang, W.-Q.: Shape-based adult image detection. IEEE International Conference on Image and Graphics, CAS, Beijing, China. (2004).*

[24] *Yin, H., Xu, X., Ye, L.: Big skin regions detection for adult image identification. IEEE Workshop on Digital Media and Digital Content Management, Hangzhou, Zhejiang, China. (2011).*

[25] *Fu, Y., Wang, W.: Fast and effectively identify pornographic images. IEEE International Conference on Computer intelligence and security, Beijing, China. (2011).*

[26] *I. P. Herman, Physics of the Human Body, New York: Springer. (2007)*

[27] *R. Balamurali and A. Chandrasekar, "Multiple parameter algorithm approaches for adult image identification," Cluster Computing. (2018)*

[28] *Kolkur, S. & Kalbande, Dhananjay & Shimpi, P. & Bapat, C. & Jatakia, Janvi. Human Skin Detection Using RGB, HSV and YCbCr Color Models (2017).*

[29] *Osman, G., Hitam, M.S., Ismail, M.N.: Enhanced Skin Colour Classifier Using RGB Ratio Model. IJSC. Computer Vision and Pattern Recognition. (2012).*

[30] *J. G. Marks and F. Miller, Lookingbill and marks' principles of dermatology, 4th ed.: Elservier Inc., (2006).*

LIST OF PUBLICATIONS

Published Paper:

A. Garg *et al*., "Machine-Learning based Community Website — A Case Study Research for a Networking Website," *2022 9th International Conference on Computing for Sustainable Global Development (INDIACom)*, 2022, pp. 604-608, doi: 10.23919/INDIACom54597.2022.9763240.

(IEEE Scopus)

Link: https://ieeexplore.ieee.org/document/9763240

Submitted Paper:

"A Nudity Detection Algorithm for Web-based Online Networking Platform," *CCICT2022 (Fifth International Conference on Computational Intelligence and Communication Technologies)*, 2022

(IEEE Scopus)